NBER WORKING PAPER SERIES

# PREVENTIVES VERSUS TREATMENTS

Michael Kremer
Christopher Snyder

Working Paper 21012
http://www.nber.org/papers/w21012

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2015

Previous versions of this paper were circulated as “Why Is There No AIDS Vaccine” among other titles. The authors are grateful to Marcella Alsan, Emmanuelle Auriol, Chris Avery, David Blanchflower, Bryan Boulier, Ryan Bubb, Jim Dana, Esther Duflo, Glenn Ellison, Amy Finkelstein, Jerry Green, Jon Hamilton, Jason Hartline, Corinne Langinier, Scott Lee, David Malueg, David McAdams, Sendhil Mullainathan, Scott Pauls, Robert Porter, Michael Schwarz, Andrew Segal, Lars Stole, GlenWeyl, Heidi Williams, Hongkai Zhang, Jon Zinman, and seminar participants at the American Enterprise Institute, Cornell, Dartmouth, Fordham, Harvard, Indiana, MIT, Northeastern, Northwestern, NYU, Princeton, RAND, Stanford, UCLA, 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; and to Lindsey Beckett, Ruben Enikolopov, Cacey Tang, Suzanne Wang, and Dan Wood for excellent research assistance. We are grateful for the medical expertise of Scott Lee, who directed the effort to code risk heterogeneity for diseases in our sample. We are especially indebted to the editors and referees for extensive suggestions which substantially improved the paper. Snyder gratefully acknowledges funding for research assistants provided by the Dartmouth College Presidential Scholars Program. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Preventives Versus Treatments
Michael Kremer and Christopher Snyder
NBER Working Paper No. 21012
March 2015
JEL No. D42,I18,L11,O31

## ABSTRACT

Preventives are sold ex ante, before disease status is realized, while treatments are sold ex post. Even if the mean of the ex ante distribution of consumer values is the same as that ex post, the shape of the distributions may differ, generating a difference between the surplus each product can extract. If, for example, consumers differ only in ex ante disease risk, then a monopolist would have more difficulty extracting surplus with a preventive than with a treatment because treatment consumers, having contracted the disease, no longer differ in disease risk. We show that the ratio of preventive to treatment producer surplus can be arbitrarily small, in particular when the distribution of consumer values has a Zipf shape and the disease is rare. The firm's bias toward treatments can be reversed, for example, if the source of private information is disease severity learned ex post. The difference between the producer surplus earned from the products can result in distorted R&D incentives; the deadweight loss from this distortion can be as large as the entire producer-surplus difference. Calibrations for HIV and heart attacks based on risk factors in the U.S. population suggest that the distribution of disease risk is sufficiently Zipf-similar to generate substantial differences between producer surplus from preventives and treatments. Empirically, we find that proxies for the Zipf-similarity of the disease-risk distribution are associated a significantly lower likelihood of vaccine development but not drug development.

Michael Kremer
Harvard University
Department of Economics
Littauer Center M20
Cambridge, MA 02138
and NBER
mkremer@fas.harvard.edu

Christopher Snyder
Department of Economics
Dartmouth College
301 Rockefeller Hall
Hanover, NH 03755
and NBER
chris.snyder@dartmouth.edu

# 1. Introduction

Many industry observers and public health advocates believe that disease preventives are inherently less lucrative than treatments (see, e.g., Rosenberg 1999). Thomas (2002) and others have argued that private incentives for research and development (R&D) on preventives fall far short of social needs particularly in the case of human immunodeficiency virus (HIV). Consistent with this view, the manufacturer did not secure the approval of Truvada as an HIV preventive until eight years after its initial approval as an HIV treatment, and the revenue generated by its use as a treatment continues to dwarf that from its use as a preventive. Commentators have argued that weak preventive incentives call for government involvement, and indeed governments have undertaken special programs to support preventive R&D for HIV and other diseases, programs including the International AIDS Vaccine Initiative (IAVI), the U.S. Advisory Council on Immunization Policy (ACIP), and the Pneumococcus Advance Market Commitment.

The delay in the development of any particular product such as Truvada could be ascribed to a variety of forces. The analysis in this paper will shed light on one determinant of the gap between private and social R&D incentives. We argue that differences in the distribution of consumer values at the time preventives and treatments are sold can affect the manufacturer's ability to extract surplus with the products. The difference in producer surplus can drive a wedge between private and social incentives to invest in the two products.

To see the logic, consider an example in which a monopoly pharmaceutical manufacturer sells directly to rational, risk-neutral, completely patient consumers. The firm can develop a preventive or a treatment, each of which has its own associated R&D cost. To make the example as simple as possible, assume both the preventive and treatment have no production costs, are perfectively effective, costless to manufacture, and have no side effects. Suppose the harm to a consumer from contracting the disease is \$100. Assume that the firm knows the distribution of disease risk in the population but not an individual's risk (or at least cannot price discriminate based on this risk).

Suppose first there is no heterogeneity in disease risk in the population. In particular, assume there are 100 consumers who have a 19% chance of contracting the disease. The firm would earn \$1,900 from a preventive by selling it to all 100 consumers at a price equal to their expected harm of \$19. The firm would earn \$1,900 in expectation from a treatment by selling it to the 19 consumers expected to contract the disease at a price equal to avoided harm of \$100. The producer surplus is the same from either product, so the firm develops whichever has lower R&D costs. Since the firm captures all of the social surplus created by each product, it would choose to develop the same product as a social planner.

Now incorporate consumer heterogeneity by assuming that they have two types: 90 have a 10% disease risk and 10 have a 100% risk. As before, 19 consumers are expected to contract the disease, so the firm continues to earn \$1,900 in expectation from a treatment. However the firm now earns less from a preventive. It can either sell only to high-risk consumers at \$100 or to all consumers at \$10. Either way, producer surplus from a preventive falls to \$1,000, biasing the firm away from developing a preventive toward a treatment. The bias can result in substantial distortions. To see how large, suppose R&D costs are slightly less than \$1,900 for a treatment and slightly more than \$1,000 for a preventive. Then the firm develops a treatment while a social planner would prefer a preventive because it also eliminates the entire $1,900 disease burden while saving nearly $900 on R&D. This extra R&D spending on the treatment represents the potential deadweight loss from a biased product choice. An even starker illustration of this large potential deadweight loss is provided in an example in which no product is developed. Take the previous example but now suppose R&D costs are slightly more than $1,900 for a treatment. Then no product is developed even though, as before, the social gain from developing a preventive exceeds the social cost by nearly $900, so again society experiences a nearly $900 deadweight loss.

When will this bias arise? What characteristics of the disease-risk distribution leads to large biases? How large can the bias possibly be? To address these questions, we construct a benchmark model in which a monopolist sells a perfectly safe and effective product, which is costless to produce, directly to risk-neutral consumers. We show that producer surplus from a treatment will always be strictly higher than from a preventive as long as there is any heterogeneity among consumers with positive disease risk. The numerical example illustrates the features of a risk distribution contributing to a large bias against preventives: although the high consumer type values the preventive ten times more than the low, it only makes up a tenth of the population, so that the preventive monopolist earns the same (low) producer surplus whichever type it targets. Distributions with the property that higher values have proportionately fewer types with at least that value are called Zipf distributions. We derive the "worst case" distribution, minimizing the ratio of producer surplus from a preventive to that from a treatment holding constant disease prevalence, showing it is a particular type of Zipf distribution, which we call a symmetrically truncated Zipf (abbreviated STRZ, read "stars") distribution. The STRZ distribution leads to a simple demand curve, one that is globally unit elastic over all prices between its truncated endpoints, generating constant producer surplus for all these prices. Indeed, constant producer surplus across prices is a necessary condition for producer-surplus minimization: if one price were more lucrative than others, there would be scope for reducing producer surplus by converting some of the mass from the marginal type into disease risk at other points in the distribution, thus holding mean disease risk constant.

We provide a formula that can be used to decompose the ratio of preventive to treatment producer surplus into just two factors: an index of similarity to the STRZ distribution and disease prevalence. For extremely rare and Zipf-similar disease distributions, the ratio of preventive to treatment producer surplus can approach zero. This limiting result can be illustrated by extending the previous numerical example by including additional consumer types: 900 with 1% disease risk, 9,000 with 0.1% risk, etc. Each additional type adds $900 to the producer surplus from a treatment but nothing to the producer surplus from a preventive because, by construction, the firm continues to earn $1,000 whichever types it targets with the preventive price. If enough types are added, the producer-surplus ratio can thus be driven arbitrarily close to zero. If disease prevalence is bounded away from zero, the producer-surplus ratio is bounded away from zero according to a formula we derive.

Much of the analysis uses a simple model of direct-to-consumer sales, which serves as a benchmark against which other institutions can be compared and policy responses evaluated. In extensions we explore some alternative institutions. We show that third-party purchases, for example by insurers or governments, can mitigate biases but will not eliminate biases if the third party bargains with the firm over prices ex-post, after R&D costs are sunk. In other extensions, we generalize the results to arbitrary production costs, efficacies, and side effects for each of the two products and to market structures beyond monopoly.

Just as ex ante heterogeneity in disease risk can lead to a bias against preventives, ex post heterogeneity in harm from a disease can lead to a bias against treatments. We characterize relative preventive and treatment revenue with heterogeneity in ex ante and ex post private information and heterogeneity in private information that persists across both ex ante and ex post periods (such as information on income or wealth).

After going through the theory, we show how the model can be used to calibrate potential deadweight loss for any disease using demographic information to estimate the distribution of consumer values in the population. We first examine the case of HIV in the United States. As others have noted, many real-world distributions are well-characterized by power laws, and this includes the distribution of sexual partners. In a simple model in which willingness to pay is proportional to risk, calibrated producer surplus from a preventive from an HIV vaccine is only 26% of the producer surplus from a treatment. We estimate that the negative correlation between HIV risk and income raises this to 38%. In contrast, for the much higher prevalence human papilloma virus (HPV), calibrations generate preventive-to-treatment producer-surplus ratios much closer to one. Zipf-like risk distributions are not limited to sexually transmitted diseases. Calibration for heart disease, the leading killer in the United States, suggests that disease-risk heterogeneity substantially reduces producer surplus from preventives.

In the empirical section, we find that an indicator for factors contributing to Zipf-similarity of the distribution of disease risk (e.g., sexual transmission, transmission through other specialized vectors, disease concentration in certain subpopulations or regions) is associated with a significantly lower probability of vaccine development—by as much as 40 percentage points—but not with a lower probability of drug development.

In focusing on the implications of time-varying consumer heterogeneity for incentives to develop preventives and treatments, we do not mean to minimize the role of other factors such as consumer behavioral biases, legal rules regarding product liability, or differences in technological opportunities that could also affect incentives. In the case of infectious disease, another factor has received some attention in the literature is the possibility that by reducing the spread of infectious disease—a positive epidemiological externality—preventives may reduce their own future demand.[1] The factors examined in the present paper are unique in that they apply to non-communicable diseases and apply to neoclassical consumers. Perhaps most importantly, the factor we study is a true bias in the sense of driving a wedge between private and social incentives. A firm may prefer to develop the product that uses "easier" science or that can be tested without harming healthy people, but these are not biases per se if the social planner shares these preferences.

[1] See Brito, Sheshinski, and Intrilligator (1991); Boulier (2006); Francis (1997); Geoffard and Philipson (1997); Gersovitz (2003); Gersovitz and Hammer (2004, 2005). Our own work (Kremer, Snyder, and Williams, 2012) argues that epidemiological externalities will tend to induce bigger distortions for communicable diseases that are rare, reinforcing the results in the present paper on the relationship between prevalence and potential deadweight loss.

Some of the core theoretical results have general implications for product markets beyond pharmaceuticals, which we touch on in this paper but explore more thoroughly in a companion paper (Kremer and Snyder 2015). There we show the fraction of surplus a monopolist is unable to extract from the market is a tight upper bound on potential deadweight loss from all sources—whether distorted product-development decisions or distorted pricing decisions. Any demand curve with a finite price intercept, when suitably rescaled, can be compared to the STRZ distribution of consumer values, which minimizes producer surplus for a given mean value. This allows us to relate potential deadweight loss to the Zipf-similarity of demand for a general class of markets. We derive bounds on static deadweight loss on gains from optimal subsidy policies and losses from banning price discrimination. The present paper focuses on a comparison of preventives versus treatments. This is a unique "laboratory" in which to study rent-extraction effects because the level of demand can be held relatively fixed (both products target the same disease with the same overall burden) while allowing the shape of the demand curve to change depending on how the distribution of consumer values evolves from the ex ante to the ex post period.

Our paper links social welfare generated by different pharmaceutical products to the ability of the monopolist to appropriate surplus selling that product. Perhaps the clearest link between appropriability and efficiency was drawn by Makowski and Ostroy (1995, 2001). They show that the reason price-taking is central to the first welfare theorem of general equilibrium is that it is sufficient for suppliers to be able to appropriate 100% of the surplus they create. The latter condition is what is needed for first-best efficiency. Our analysis complements theirs in that instead of bounding how efficient the equilibrium can be when appropriability is easy, we bound how inefficient it can be what appropriability is difficult, and offer characteristics of products and shapes of demands that lead to particular difficulties.

This and the companion paper together contribute to a microtheory literature linking the shape of the demand curve to producer surplus. Anderson and Renault (2003) bound the ratio of producer to total surplus as a function of a generalized notion of the concavity or convexity of demand in a Cournot market. Some of our special cases (Propositions 9 and 10) can be proved as immediate corollaries of their theorems. Weyl and Fabinger (2013) (see also Fabinger and Weyl 2014) provide bounds on the surplus ratio for arbitrary demand and cost curves and oligopoly models; the bounds are tied to the elasticity of the marginal-producer-surplus function and oligopoly conduct parameters. Johnson and Myatt (2006) construct several orderings on demand curves including clockwise rotations. They provide a rich set of applications in which a firm's strategy—e.g., advertising or product design—is isomorphic to a choice of an ordered demand curve.[2] They show profit is typically quasiconvex in the demand ordering, rationalizing all-or-nothing strategy choices observed in the applications. Our decomposition formula provides a different way to order demand curves based on Zipf similarity and mean surplus.

Demand curves generating equal producer surplus whatever price is charged (alternatively labeled Zipf,

[2] The idea that the firm will choose the most profitable shape for consumer demand has been applied to a diverse set of phenomena in industrial organization. DeGraba (1995) explains buying frenzies as a response to supply limitations, inducing consumers to race to buy before acquiring more information about their true valuations. Biehl (2001) applies the idea to the sell versus lease decision.

unit-elasticity, equal-revenue, or extremal) have been discovered to be a useful tool in proving a wide range of important recent results.[3] Bergemann, Brooks, and Morris (2014) show that any market can be represented as a convex combination of segmented markets with such demands. Because the monopolist is indifferent among prices charged in one of these segmented markets, the modeler can design equilibrium discriminatory prices attaining any division of first-best surplus across the firm, consumers, and deadweight loss (as long as the monopolist earns at least the profit under uniform pricing). Brooks (2013) contemporaneously derived the symmetrically truncated Zipf demand to provide a worst case for his belief survey auction, the optimal mechanism for an uninformed principal when agents are informed about their types and rivals' type spaces. The formula for the revenue ratio in his Proposition 1 is in fact identical to our equation (7). A growing literature in computer science uses the construction to bound worst cases for approximately optimal mechanisms in different settings. Hartline and Roughgarden (2009) provides an early such reference: they use a constant-revenue demand to compare revenue from an optimal auction to that from a Vickery auction with no reserve but with one more bidder (see their Example 4.6). Weyl and Fabinger (2013) anticipate the special nature of Zipf demand: in the case of a monopoly with costless production, their formulae imply that the ratio of producer to total surplus vanishes for Zipf demand. We contribute by showing that the STRZ demand is the unique minimizer of this surplus ratio among demand curves with a given mean and finite support, without imposing continuity or differentiability restrictions.

Our paper contributes to the literature on incentives for innovation in R&D-intensive industries (see, e.g., Newell, Jaffee, and Stavins 1999; Acemoglu and Linn 2004; Finkelstein 2004; and Budish, Roin, and Williams 2013). Most closely related are studies of innovation in healthcare markets by Lakdawalla and Sood (2013) and especially Garber, Jones, and Romer (2006). The latter paper relates static and dynamic deadweight loss to the shape of the demand curve as we do. They focus on a different distortion, that coinsurance can induce overconsumption and excess entry by defraying a fraction of the pharmaceutical price. Although we have an extension to third-party procurement, in our benchmark model of private-market sales deadweight loss is solely due to underconsumption and too little entry, generating novel conditions on the demand shapes generating the biggest distortions.

We also contribute 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 ex ante private information about their disease risk. 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 preventives because, once the preventive 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 to disease risk calls for a specific mapping from ex

[3] An early working-paper version of this paper (Kremer and Snyder 2003) used the equal-revenue property to construct discrete distributions attaining certain lower bounds and attaining a limit on the ratio of preventive to treatment producer surplus of zero. That version did not derive the STRZ distribution, bounds on deadweight loss, or a number of other core results and extensions in this published version. Earlier work by one of us with a coauthor (Malueg and Snyder 2006) considers a sequence of linear demands with the equal-revenue property to bound the ratio of profits from discriminatory to uniform pricing by a function of the number of markets.

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 disease risk rather than the particular class of beta distributions examined by Miravete, to characterize the worst case distribution as the Zipf, and to decompose the producer-surplus ratio into prevalence and Zipf-similarity factors, all of which are new results in the literature. Our results on deadweight loss, and our calibrations and empirical work are new as well.

The remainder of this paper is organized as follows. Section 2 sets up the benchmark model. Section 3 provides a full analysis of this model. We show that if consumers differ only in disease risk, producer surplus from a treatment exceeds that from a preventive and that the gap between these surpluses provides a bound on the potential deadweight loss from the resulting R&D distortions. We decompose the producer-surplus ratio into two contributing factors: low disease prevalence and Zipf-similarity of the distribution of disease risk. Section 4 generalizes the analysis in a number of directions, allowing for imperfect efficacy, side effects, production costs, a broad class of models with competition among suppliers, and third-party purchases. Section 5 explores alternative sources of heterogeneity besides disease risk. Just as ex ante heterogeneity in disease risk leads to a bias against preventives, ex post heterogeneity in harm from disease can lead to a bias against treatments. The section connects the ratio of preventive to treatment producer surplus to the joint distribution of ex ante disease risk, harm from disease realized ex post, and persistent differences in willingness to pay for expected reductions in harm from disease, for example due to differences in income. Section 6 calibrates the model for HIV/AIDS, HPV, and heart attacks. Section 7 provides a first-pass empirical test of the model. Section 8 concludes with a discussion of implications for public policy, noting that our results provide a potential rationale for programs like the U.S. Advisory Committee on Immunization Practices or Advance Market Commitments.

## 2. Benchmark Model

We begin with a benchmark model of a monopoly pharmaceutical manufacturer. The firm can produce either a preventive or a treatment, which it sells directly to consumers. In the next subsection, we will verify that the restriction to a single product is made without loss of generality given the parametric assumptions imposed in the benchmark model. Extensions allowing allowing for general market structures with competing suppliers, allowing for government or other third-party procurement, and allowing for both products to be produced under general values of the parameters are deferred to Section 4, where we show that the key welfare results continue to hold. Let $j$ index products, with $j = p$ for the preventive and $j = t$ for the treatment. Its decision to enter the market for product $j$ is reflected by the indicator variable $E_j$. Entry requires a fixed R&D expenditure $k_j$. After entering, it produces at constant marginal cost $c_j$.

Let $p_j$ be the price it sets for product $j$, $Q_j(p_j)$ be the demand curve, $PS_j(p_j) = (p_j - c_j)Q_j(p_j)$ be producer surplus, $CS_j(p_j) = \int_{p_j}^{\infty} Q(x)dx$ be consumer surplus, and $TS_j(p_j) = PS_j(p_j) + CS_j(p_j)$ be total surplus. Note $PS_j(p_j)$ and $TS_j(p_j)$ are surpluses from an ex post perspective, i.e., treating $k_j$ as a sunk cost and thus ignoring it. Profit from an ex ante perspective—treating $k_j$ as an economic cost—is denoted

$\Pi_j(p_j) = PS_j(p_j) - k_j$. Ex ante social welfare is denoted $W_j(p_j) = TS_j(p_j) - k_j$.

On the demand side, consumers are risk neutral. Before purchasing any product, each consumer learns his or her disease risk, $x \in [0,1]$, i.e., the probability he or she contracts the disease. Assume $x$ is a realization of random variable $X$ with cumulative distribution function (cdf) $F_X(x)$ and complementary cdf $\bar{F}_X(x) = \Pr(X > x) = 1 - F_X(x)$. The proportion of consumers with disease risk at least as great as some value $x$ is denoted $\Phi_X(x) = \Pr(X \geq x) = \bar{F}(x) + \Pr(X = x)$. The mean disease risk—also disease prevalence in the absence of a preventive—is $\mu_X = \int_0^1 x\, dF_X(x)$. Assume the firm knows the distribution of $X$ in the population but cannot price discriminate across consumers based on realized values $x$.[4] If a consumer contracts a disease and has not had the preventive or does not receive the treatment, he or she experiences harm $h \geq 0$ in present discounted value terms. Prior to Section 5, which explores various sources of heterogeneity in willingness to pay, we assume that consumers all would pay the same $h$ to avoid harm.

We will impose a number of simplifying assumptions in the benchmark model. Both products are assumed to be costless to manufacture and administer (i.e., $c_j = 0$), are perfectly effective, and have no side effects. The discount rate is normalized to 0. (Discounting can be accommodated by interpreting dollar values in present-discounted-value terms.) Normalize both $h$ and the mass of consumers to 1.

The products differ in the timing of when they are sold relative to when consumer learn their disease status. To understand how this affects the firm's ability to extract surplus, consider each product in turn. First, suppose the firm develops a preventive. A consumer is willing to purchase the preventive if its price $p_p$ does not exceed the expected value of avoided harm—disease risk $x$ times harm normalized to 1.[5] Given the mass of consumers is normalized to 1, the mass of consumers who purchase the preventive is $\Phi_X(p_p)$.[6] Hence preventive demand is

$$Q_p(p_p) = \Phi_X(p_p). \tag{1}$$

Suppose instead the firm develops a treatment, which the consumer purchases after becoming infected. Infected consumers buy if avoided harm, 1, at least weakly exceeds $p_t$. Hence treatment demand is

$$Q_t(p_t) = \begin{cases} \mu_X & p_t \leq 1 \\ 0 & p_t > 1. \end{cases} \tag{2}$$

[4] Price discrimination can be ruled out if $x$ is private information for consumers (for example, related to their sexual behavior or intravenous drug use, conducted in private) or if $x$ is public information but discrimination is prevented by the difficulty of controlling resale or other administrative, institutional, or legal barriers.

[5] By Theorem 4 of Harris and Raviv (1981), selling at a simple linear price $p_p$ is optimal among the set of potentially complicated mechanisms that could be used to sell the preventive.

[6] For continuous distributions of disease risk, marginal consumers' purchasing decisions are immaterial because they have zero measure. Discrete and mixed distributions may have a probability atom at the marginal consumer's disease risk. To ensure existence of equilibrium in this case, we must assume that indifferent consumers make the same choice as inframarginal consumers with higher disease risks. Mixed distributions merit careful treatment because of their role in the later analysis: the symmetrically truncated Zipf distribution bounding the producer-surplus ratio is mixed. The practical implication of the assumption about the behavior of indifferent consumers is that $\Phi_X(x)$ rather than $\bar{F}(x)$ will be the relevant demand curve in the analysis.

# 3. Benchmark Analysis

## 3.1. Initial Results

To facilitate the derivation of equilibrium and comparison to the first best, we introduce some additional notation. Let stars denote equilibrium values. Thus, for example, $p_j^* = \operatorname{argmax}_{p_j \geq 0} PS_j(p_j)$ is the monopoly price for product $j$, $q_j^* = Q_j(p_j^*)$, $PS_j^* = PS_j(p_j^*)$, and $E_j^*$ indicates whether $j$ is produced in equilibrium. Dropping the subscript, $W^*$ indicates social welfare given the equilibrium product choice, i.e., $W^* = E_p^* W_p^* + E_t^* W_t^*$. Let double stars denote first-best values. Thus, for example, $p_j^{**} = c_j$, $TS_j^{**} = TS_j(c_j)$, $W^{**} = W_j(c_j)$, and $E_j^{**}$ indicates whether product $j$ is produced in the first best. Dropping the subscript, $W^{**}$ denotes first-best welfare given the first-best product choice, i.e., $W^{**} = E_p^{**} W_p^{**} + E_t^{**} W_t^{**}$.

The first result is a lemma with a number of useful implications. Although there are two potential products for the disease, the lemma implies that there is no ambiguity in speaking of "the" first-best surplus on this market. Second, the lemma implies that a social planner has no reason to favor one product over the other in the first best of the benchmark model. Both provide the same total surplus in the first best, so the planner would develop whichever one had the lower development cost $k_j$.

**Lemma 1.** *In the benchmark model, the first-best surplus (from an ex post perspective, conditional on some product being available) is the same whether the firm produces a preventive or a treatment. Letting $TS^{**}$ denote this first-best surplus, we have $TS^{**} = \mu_X = \int_0^1 \Phi_X(x)dx$.*

The proof is provided in the appendix. To gain some intuition for the proof, note that with harm normalized to $h = 1$, $\mu_X$ is the total disease burden. But if both products are costless to produce, perfectly effective, and have no side effects, either can relieve the entire disease burden, generating first-best surplus $\mu_X$.

The firm's equilibrium product choice can be easily characterized in the benchmark model. It develops a preventive if $\Pi_p^* > \max\{\Pi_t^*, 0\}$, a treatment if $\Pi_t^* > \max\{\Pi_p^*, 0\}$, and neither if $\max\{\Pi_p^*, \Pi_t^*\} < 0$. The remaining strategy—developing both products—can be ignored because it is dominated by developing the treatment alone in the benchmark model. To see this, note from (2) that at the optimal treatment price $p_t^* = 1$, the firm sells to all $\mu_X$ consumers who contract the disease, implying

$$PS_t^* = \mu_X = TS^{**}. \tag{3}$$

Able to extract all social surplus with the treatment, the firm has no additional reason to expend the fixed cost of developing the preventive. (Section 4.1 allows for the possibility that both products are developed in an extension with imperfectly safe and effective products.)

We saw from Lemma 1 that the social planner has no reason to favor one product over the other. The next proposition states that the monopolist may have a bias toward the treatment because it is better at extracting surplus from the market.

**Proposition 1.** *In a benchmark pharmaceutical market, the firm never develops a preventive unless it is socially efficient to do so both in equilibrium and in the first best. There exist cases in which the firm develops a treatment but it would have been socially efficient to develop a preventive.*

*Proof.* Suppose the firm develops the preventive in equilibrium. Then $\Pi_p^* \geq \max(\Pi_t^*, 0)$. But $W_p^{**} \geq W_p^* \geq \Pi_p^*$ for any product. By equation (3), $\Pi_t^* = PS_t^* - k_t = TS^{**} - k_t = W_t^* = W_t^{**}$. Substituting into the first inequality yields $W_p^{**} \geq \max(W_t^{**}, 0)$ and $W_p^* \geq \max(W_t^*, 0)$. Thus it is socially efficient to develop a preventive, whether equilibrium or first-best prices are set.

The appendix completes the proof by constructing an example in which a treatment is developed but a preventive would be socially more efficient. *Q.E.D.*

We next turn to quantifying the social loss from the firm's bias toward treatments. To do so, it is useful to first distinguish between two deadweight-loss concepts. Static deadweight loss in the market for product $j$ is $SDWL_j(p_j) = TS_j^{**} - TS_j(p_j)$. This difference between first-best and equilibrium surplus effectively takes the decision to develop product $j$ as given, reflecting just the distortion at the intensive margin of charging some supra-competitive price $p_j \geq c_j$. The equilibrium value of static deadweight loss is $SDWL_j^* = SDWL_j(p_j^*)$. Deadweight loss without the "static" modifier is a more comprehensive concept, capturing distortions at all margins, both the intensive margin (pricing) and the extensive margin (entry). Denote this deadweight-loss concept by $DWL = W^{**} - E^*W^*$. The next proposition states that this second deadweight-loss concept is bounded by the difference between treatment and preventive producer surpluses. If rather than the level of deadweight loss, one considers what Tirole (1988) calls relative deadweight loss, i.e., deadweight loss as a proportion of first-best surplus, the next proposition shows that is bounded by $1 - \rho_X^*$, where $\rho_X^* = PS_p^*/PS_t^*$ denotes the producer-surplus ratio.

**Proposition 2.** *In a benchmark pharmaceutical market, a tight upper bound on potential deadweight loss is given by the difference in producer surpluses, $PS_t^* - PS_p^*$; i.e.,*

$$\sup_{k_p, k_t \geq 0} (DWL) = PS_t^* - PS_p^*. \tag{4}$$

*Expressed as a percentage of disease burden $\mu_X$, this tight upper bound is given by $1 - \rho_X^*$; i.e.,*

$$\sup_{k_p, k_t \geq 0} \left( \frac{DWL}{\mu_X} \right) = 1 - \rho_X^*, \tag{5}$$

Proposition 2 is a corollary of a more general proposition in Section 4.1. There it is shown that equation (5) continues to hold when the parameter space is expanded to allow for arbitrary production costs, efficacies, and side effects for each product. The proof of Proposition 2 will thus be deferred to the later section. Because it is a core result in the paper, however, the proof merits a sketch here.

In any equilibrium in which a preventive is developed, by Proposition 1 the product choice must be efficient. The only source of deadweight loss is static deadweight loss from the intensive margin of price exceeding cost; i.e., $DWL = SDWL_p^*$. In an equilibrium in which a treatment is developed, deadweight loss can include the dynamic effect of the inefficient product choice. First-best welfare is $W^{**} = TS^{**} - k_p$ when a preventive is the first-best product, while equilibrium welfare is $W^* = TS^{**} - k_t$ when a treatment is the equilibrium product (because the treatment extracts all surplus in equilibrium). Deadweight loss then is $DWL = W^{**} - W^* = (TS^{**} - k_p) - (TS^{**} - k_t) = k_t - k_p$. The firm would still be willing to develop the treatment even as the gap in development costs $k_t - k_p$ approaches $PS_t^* - PS_p^*$. One can check that this

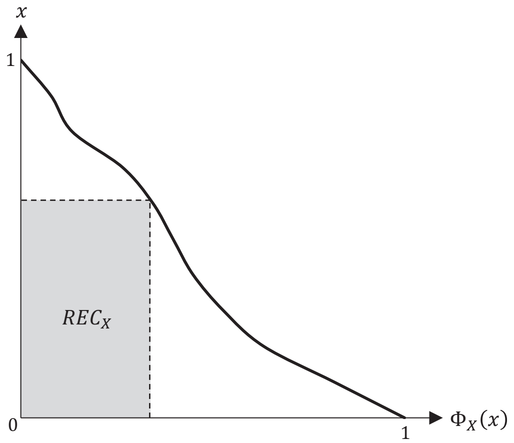

**Figure 1: Producer-surplus ratio.** Producer-surplus ratio $\rho^*_X$ equals the ratio of the area of the shaded rectangle to the area under the curve.

potential deadweight loss from inefficient product choice exceeds the static deadweight loss in a preventive equilibrium: $PS^*_t - PS^*_p = TS^{**} - PS^*_p = CS^*_p + SDWL^*_p \geq SDWL^*_p$, where the first equality holds by (3). This establishes the bound on potential deadweight loss in levels given in (4). To express the bound in relative terms as in (5), divide by $TS^{**}$ and substitute from (3): $(PS^*_t - PS^*_p)/TS^{**} = (TS^{**}/TS^{**}) - (PS^*_p/PS^*_t) = 1 - \rho^*_X$.

The sketch of the proof evinces two important economic principles. First, the extra cost of developing a product just because it is better at extracting surplus can constitute a social loss. Second, this dynamic deadweight loss can potentially swamp any static deadweight loss from super-competitive product prices.

The next lemma provides a simple formula for $\rho^*_X$ that can be read off a graph. Note that $x\Phi_X(x)$ is the area of the rectangle of height $x$ inscribed under $\Phi_X$. Let $REC_X = \max_{x\in[0,1]}[x\Phi_X(x)]$ denote the area of the largest such rectangle, shown in Figure 1 as the shaded region. Lemma 1 showed that $\mu_X$ equals the whole area under $\Phi_X$. The next lemma states that $\rho^*_X = REC_X/\mu_X$, implying that $\rho^*_X$ is the ratio of the area of the shaded rectangle to the area under the whole curve.

**Lemma 2.** *The producer-surplus ratio $\rho^*_X$ satisfies $\rho^*_X = REC_X/\mu_X$.*

*Proof.* In the preventive market, $PS^*_p = \max_{p_p \geq 0}[p_p Q_p(p_p)] = \max_{x\in[0,1]}[x\Phi_X(x)] = REC_X$, where the first equality holds by definition, the second by substituting from (1) and making the change of variables $x = p_p$, and the last by definition. Dividing, $\rho^*_X = PS^*_p/PS^*_t = REC_X/\mu_X$, where the last equality follows from equation (3). *Q.E.D.*

It is obvious from Figure 1 that the area of the shaded rectangle cannot exceed the area under the curve, and thus $\rho^*_X \leq 1$. The economic intuition behind this graphical result is that the treatment monopolist knows more about consumers than the preventive monopolist, in particular the treatment monopolist learns their disease status from their decision to purchase. This additional information cannot harm the monopolist, as we know from Ottaviani and Prat (2001). When is the weak inequality strict, i.e., $\rho^*_X < 1$? According to the next proposition, always, unless all consumers with positive values are homogeneous.

**Proposition 3.** *$\rho_X^* = 1$ if and only if there exists some $x' \in (0,1]$ such that $\Pr(X = x'|X > 0) = 1$. Otherwise $\rho_X^* < 1$.*

The proof is provided in the appendix. Intuitively, if consumers are homogeneous, the monopolist can extract 100% of total surplus with either product, eliminating any wedge between private and social R&D incentives. The first best is obtained in equilibrium, and there is no deadweight loss. Graphically, the demand curve analogous to Figure 1 is itself a rectangle in the case of homogeneous disease risk, so an inscribed rectangle would fill the entire area below the curve. On the other hand, any heterogeneity in positive disease risk will prevent the firm from extracting 100% of surplus with a preventive given the firm cannot price discriminate. Although they may be heterogeneous ex ante, infected consumers are homogeneous ex post, so the firm will still be able to extract 100% of surplus with a treatment. Thus producer surplus is strictly less from a preventive than from a treatment, implying $\rho_X^* < 1$. Graphically, with nontrivial heterogeneity in disease risk, there is no way to capture all the area under the demand curve with an inscribed rectangle.

It would be convenient to use some familiar feature of the distribution of $X$ as a proxy for $\rho_X^*$. One natural candidate is variance. According to Proposition 3, the move from an $X$ with no heterogeneity in positive values to one with some heterogeneity—equivalent to introducing variance in positive values of $X$—reduces $\rho_X^*$ from 1 to some value below 1. One might hope that $\rho_X^*$ is inversely related to the variance of $X$, providing a simple proxy for comparative statics. Unfortunately, we will show this is not the case. We saw in Lemma 2 that $\rho_X^*$ equals the ratio of the area of the largest rectangle inscribed under $\Phi_X$ to the area under $\Phi_X$, which in turn depends on the detailed shape of the whole distribution of $X$. Not only does variance not proxy for $\rho_X^*$ but neither does skewness, kurtosis, or any other higher moment of the distribution of $X$, as the next proposition states.

**Proposition 4.** *Let $M_n(X)$ be the order $n$ moment of random variable $X$. The ranking of $M_n(X)$ does not in general determine the ranking of $\rho_X^*$. Formally, for all $n \geq 2$, we can find random variables $X_1$, $X_2$, $X_3$, $X_4$ with mean $\mu_X$ and support $[0,1]$ such that*

$$[M_n(X_1) - M_n(X_2)][\rho_{X_1}^* - \rho_{X_2}^*] > 0$$
$$[M_n(X_3) - M_n(X_4)][\rho_{X_3}^* - \rho_{X_4}^*] < 0.$$

*This result holds whether $M_n(X)$ is taken to be the raw moment $E(X^n)$, the central moment $E((X-\mu_X)^n)$, or the standardized moment $E((X-\mu_X)^n)/\sigma_X^n$, where $\sigma_X$ is the standard deviation of $X$.*

The proof provided in the appendix is by construction, working with the simplest of distributions, the two-type case, which can be completely characterized by three parameters (probability of the low type and the disease risks of the high and low type). For a given moment, we construct two parameter changes, both of which increase the moment, but one of which increases $\rho_X^*$, the other of which decreases $\rho_X^*$.

Proposition 4 says that a moment cannot determine which distributions are associated with low values of $\rho_X^*$.[7] Are there any other features of a distribution that can be used for this end? The answer provided in

[7] Similar results hold for other ways to capture an increase in heterogeneity such as mean-preserving spreads or increases in Gini mean difference.

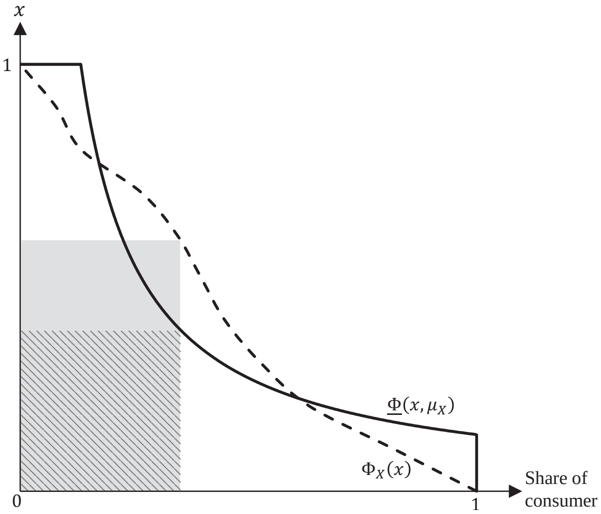

**Figure 2:** Derivation of demand attaining lower bound.

the next subsection is "yes." We show $\rho^*_X$ can be decomposed into two factors: (1) disease prevalence $\mu_X$ and (2) resemblance of the distribution to the worst case, which has the lowest possible $\rho^*_X$ for a given $\mu_X$.

## 3.2. STRZ Distribution

The key step in the decomposition of $\rho^*_X$ for a given demand curve is to find the worst case, i.e., the demand curve solving the problem of minimizing producer surplus subject to having area underneath of $\mu_X$. We will do this with the help of Figure 2. Consider an arbitrary demand $\Phi_X(x)$, drawn as the dotted line. Equilibrium producer surplus is the largest that can be inscribed under it, the shaded rectangle. Imagine transforming $\Phi_X(x)$ by moving some area away from the corner of the shaded rectangle to other parts of the curve, maintaining $\mu_X$ as the area under the curve. This transformation will reduce $REC_X$, implying that $\Phi_X(x)$ could not have been the solution to the minimization problem. For a demand curve to be solve the minimization problem, all inscribed rectangles must have the same area, as is the case with demand curve $\underline{\Phi}(x,\mu_X)$. This argument shows that $\underline{\Phi}(x,\mu_X)$ is the unique minimizer of $REC_X$ among distributions with a given $\mu_X$ and, because $\mu_X$ is constant, the unique minimizer of $\rho^*_X$ given $\mu_X$. Let $\underline{\rho}(\mu_X)$ denote the minimized value of $\rho^*_X$ over distributions with prevalence $\mu_X$.

It remains to determine the functional form of $\underline{\Phi}$. It was constructed so that the area of the inscribed rectangle, $x\underline{\Phi}(x,\mu_X)$, equals some constant $A$ over realizations of $X$. Rearranging, $\underline{\Phi}(x,\mu_X) = A/x$. This is a globally unit-elastic demand curve (as expected from the well-known property that revenue is constant in price where demand is unit elastic). Filling in the remaining details, we need to incorporate the constraint that demand not exceed 1, yielding $\underline{\Phi}(x,\mu_X) = \min\{A/x, 1\}$. To find an expression for $A$, note that the area of the largest inscribed rectangle under $\underline{\Phi}$ is $A$ by construction. Hence $\underline{\rho}(\mu_X) = A/\mu_X$, implying $A = \mu_X\underline{\rho}(\mu_X)$. Although a closed-form expression is not available for $\underline{\rho}(\mu_X)$, we can drive an implicit expression for it. Lemma 1 implies

$$\mu_X = \int_0^1 \underline{\Phi}(x,\mu_X)dx = \int_0^A dx + \int_A^1 \frac{A}{x}dx = A(1-\ln A). \tag{6}$$

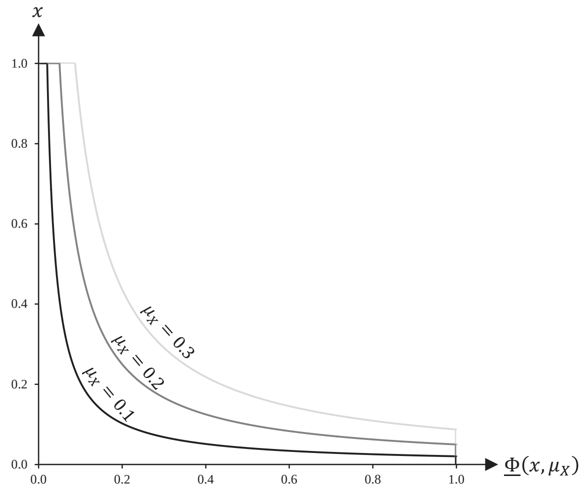

**Figure 3:** Symmetrically truncated Zipf (STRZ) demand for various $\mu_X$.

Substituting $A = \mu_X \underline{\rho}(\mu_X)$ into (6) and simplifying shows that $\underline{\rho}(\mu_X)$ is the implicit solution to

$$\underline{\rho}(\mu_X)[1 - \ln(\mu_X \underline{\rho}(\mu_X))] = 1. \tag{7}$$

The following proposition summarizes this analysis. A rigorous proof following the outlines of the preceding sketch is provided in the appendix.

**Proposition 5.** *The following demand curve is associated with the unique (almost everywhere) distribution minimizing $\rho_X^*$ subject to disease prevalence being at least $\mu_X$:*

$$\underline{\Phi}(x, \mu_X) = \min\left\{ \frac{\mu_X \underline{\rho}(\mu_X)}{x}, 1 \right\}, \tag{8}$$

*where $\underline{\rho}(\mu_X)$ is the lower bound on $\rho_X^*$ attained by the distribution, the implicit solution to* (7).

Demand curve $\underline{\Phi}$ can be connected to the growing literature on power laws. According to the terminology in Gabaix (2009), a distribution over $X$ is said to satisfy a power law if $\bar{F}_X(x) = Ax^{-\zeta}$ for some constants $A, \zeta > 0$ and for an interval of $x$; $\zeta$ is called the power-law exponent. A distribution is said to satisfy Zipf's law if it is a power-law distribution with exponent $\zeta$ near 1. A bit of work shows that the distribution underlying $\underline{\Phi}$ satisfies Zipf's law.[8] It is a special case with support truncated so that its upper and lower ends match (see Figure 2). We will therefore call equation (8) a *symmetrically truncated Zipf (STRZ) demand* and the distribution underlying it a *symmetrically truncated Zipf (STRZ) distribution*.

Figure 3 graphs examples of STRZ demand $\underline{\Phi}(x, \mu_X)$ for various prevalence levels ranging from $\mu_X = 0.1$ to 0.3. Consider $\underline{\Phi}(x, 0.3)$. In this example, as can be inferred from (8), the distribution has no mass for the lowest types ($x < 0.087$). For higher types, $\underline{\Phi}(x, 0.3)$ resembles a rectangular hyperbola. The highest

[8] For all $x$ in the interior of its support, the distribution under lying $\underline{\Phi}$ is continuous. Thus, for this range of $x$, $\underline{\Phi}$ is the complementary cdf as well as being the demand curve. But this complementary cdf is of the power-law form with exponent 1, proving the distribution satisfies Zipf's law.

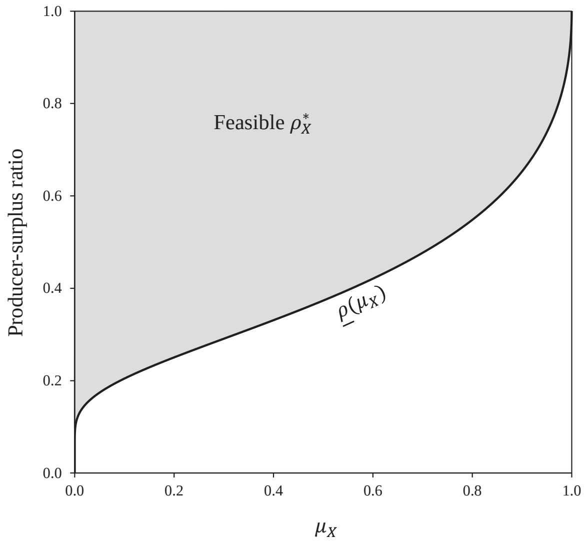

**Figure 4:** Lower bound $\underline{\rho}(\mu_X)$ as a function of prevalence $\mu_X$.

relative net value ($x = 1$) has a point mass with exactly that probability (approximately 0.087) required for the curve's truncated extremes to look identical. It can be shown that the bound on the producer-surplus ratio attained in the example is $\underline{\rho}(0.3) = 0.29$. As $\mu_X$ falls, the associated curves "hug" the axes more closely and are able to generate lower producer-surplus ratios, $\underline{\rho}(0.2) = 0.25$ and $\underline{\rho}(0.1) = 0.21$.

Figure 4 graphs the lowest producer-surplus ratio attainable, $\underline{\rho}(\mu_X)$, as a function of disease prevalence $\mu_X$.[9] The set of feasible $\rho^*_X$ is shown as the shaded region above the curve. An empirical implication of the figure is that for the most common diseases, disease-risk heterogeneity cannot be an important factor in a firm's decision to develop a preventive versus a treatment. For example, the figure shows that if the prevalence of the disease is above 0.74, it is mathematically impossible to generate enough disease-risk heterogeneity to drive $\rho^*_X$ below 1/2. For heterogeneity in disease risk to generate a substantial bias against preventives requires the disease to be sufficiently rare.

Inspection of Figure 4 suggests that $\underline{\rho}(\mu_X)$ is increasing in prevalence, ranging from 0 for the lowest prevalence to 1 for the highest prevalence. Intuitively, if the disease is ubiquitous, most consumers' disease risk must be close to 1, implying disease risk is effectively homogeneous. Lower values of prevalence allow for a substantial bias against preventives. Figure 3 shows that STRZ demands "hug" the axes more tightly the lower $\mu_X$, reducing the area of the largest rectangle that can be inscribed under the curve faster than the area under the curve. For the rarest diseases, i.e., as $\mu_X$ approaches 0, $\underline{\rho}(\mu_X)$ approaches 0. These claims are stated formally in the next proposition, proved in the appendix.

**Proposition 6.** $\underline{\rho}'(\mu_X) > 0$, $\lim_{\mu_X \downarrow 0} \underline{\rho}(\mu_X) = 0$, *and* $\lim_{\mu_X \uparrow 1} \underline{\rho}(\mu_X) = 1$.

An important corollary of the proposition, in particular of the statement $\lim_{\mu_X \downarrow 0} \underline{\rho}(\mu_X) = 0$, is that cases can be constructed such that $\rho^*_X$ is arbitrarily close to 0. But this means, by Proposition 2, that cases can be

[9] Although there is no closed-form expression for $\underline{\rho}(\mu_X)$, there is a closed-form expression for its inverse. Figure 4 graphs this inverse on the horizontal axis.

constructed in which the bias against preventives dissipates 100% of total surplus.

With this analysis in hand, we can return to the decomposition of the bias against preventives promised at start of the subsection. We suggested that one factor in the decomposition is how closely the risk distribution for the disease resembles the worst case, which we just found to be the STRZ distribution. A challenge in deriving an index of similarity is capturing global shape differences with a single number. We will define similarity between a STRZ and another demand curve as the ratio of uncaptured surpluses they entail. Formally, let $Z_X$ denote the Zipf similarity of the distribution of $X$, defined as

$$Z_X = \frac{\mu_X - REC_X}{\mu_X - \underline{REC}_X} = \frac{1-\rho^*_X}{1-\underline{\rho}(\mu_X)}, \tag{9}$$

where the second equality follows from dividing numerator and denominator by $\mu_X$. Since $REC_X \in [\underline{REC}(\mu_X), \mu_X]$, it follows that $Z_X \in [0,1]$, with $Z_X = 0$ for homogeneous consumers and $Z_X = 1$ for a STRZ distribution. Rearranging (9) gives the decomposition provided in the next proposition.

**Proposition 7.** *The producer-surplus ratio for a given disease-risk distribution satisfies*

$$\rho^*_X = 1 - Z_X[1-\underline{\rho}(\mu_X)]. \tag{10}$$

To gain some intuition for this decomposition, if demand is not Zipf-similar at all ($Z_X = 0$), then $\rho^*_X = 1$, implying there is no bias against preventives. As Zipf-similarity increases, $\rho^*_X$ falls. How much $\rho^*_X$ falls depends on $1-\underline{\rho}(\mu_X)$, which can be interpreted as how difficult it is to capture surplus with a fully Zipf demand. The lower prevalence $\mu_X$, the more difficult capturing surplus is. The two factors $Z_X$ and $\mu_X$ completely determine $\rho^*_X$. We will show how to apply the decomposition in practice in the Section 6 calibrations.

## 3.3. Special Cases

The analysis so far has provided general results for unrestricted distributions of consumer values $X$. This subsection derives additional results in several special cases that are of pedagogical and practical interest. Considering these special cases puts more structure on demand, allowing us to derive a more refined set of results for these cases. We first look at markets with a discrete distribution of consumer values and second with a continuous distribution of consumer values having global curvature properties.

**Proposition 8.** *Suppose the distribution of disease risk involves $T$ discrete types. Then $1/T$ is a tight lower bound on $\rho^*_X$.*

The proof in the appendix is by construction. We construct a discrete version of a Zipf distribution that approaches the bound in the limit as the disease prevalence approaches 0.

An implication of the proposition for examples like the one from the Introduction with two consumer types, $\rho^*_X$ can come arbitrarily close to 1/2 but can be no lower. This implies that potential deadweight loss in a market with two consumer types can be nearly 1/2.

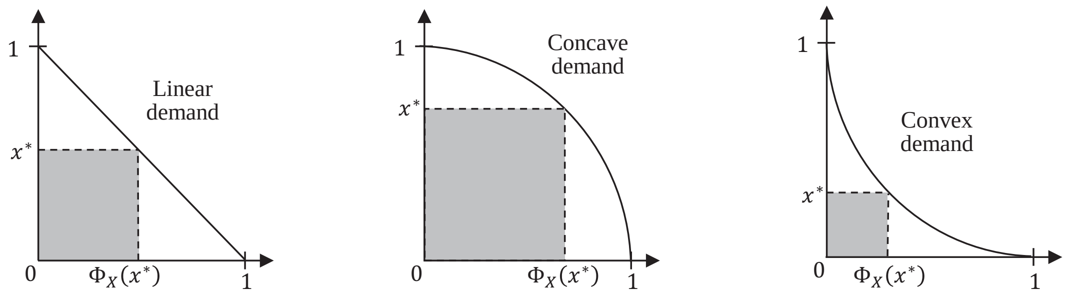

**Figure 5:** Producer-surplus ratio depends on curvature of demand.

The next special case moves from discrete to continuous distributions. Attention is further restricted to twice differentiable demands that are either globally concave, convex, or linear over the full support $(0,1)$. The curvature of $\Phi_X$ constrains $\rho_X^*$, as the next proposition states.

**Proposition 9.** *If $\Phi_X''(x) = 0$ for all $x \in [0,1]$, then $\rho_X^* = 1/2$. If $\Phi_X''(x) \leq 0$ for all $x \in [0,1]$, then $\rho_X^* \geq 1/2$. If $\Phi_X''(x) \geq 0$ for all $x \in [0,1]$, then $\rho_X^* \leq 1/2$.*

The result, proved in the appendix, is a corollary of a quite general proposition in Anderson and Renault (2003) relating $PS_j^*/TS_j^{**}$ to general degrees of concavity or convexity under *n*-firm Cournot competition in product market *j*. Figure 5 provides intuition in our simpler monopoly setting. The first graph illustrates the linear-demand case. Standard results imply that the area of the largest rectangle that can be inscribed under a line is half of the area under the line, so by Lemma 2, which relates $\rho_X^*$ to these areas, $\rho_X^* = 1/2$. The second graph illustrates the case of concave demand. As the figure suggests, the area of the largest rectangle that can be inscribed under the curve is at least half the area under the curve, so $\rho_X^* \geq 1/2$. The third graph shows the case of convex demand. As the figure suggests, the area of the largest rectangle that can be inscribed under the curve is no more than half the area under the curve, so $\rho_X^* \leq 1/2$.

As a further corollary of Anderson and Renault (2003), we have the following results for demands with log-curvature.

**Proposition 10.** *Suppose $\Phi_X(x)$ is twice continuously differentiable. If $\Phi_X(x)$ is log-concave for all $x \in [0,1]$, then $\rho_X^* \geq 1/e$. If $\Phi_X(x)$ is log-convex for all $x \in [0,1]$, then then $\rho_X^* \leq 1/e$.*

For a proof, see the proof of Proposition 9 in the appendix, which in fact provides general results for arbitrary degrees of curvature which nest the last two propositions.

A new feature of Propositions 9 and 10 is that they provide upper bounds on $\rho_X^*$ in some instances, whereas results up to that point provided lower bounds. This feature is worth emphasizing. The set of convex demands covered by Proposition 9 is equivalent to the set of downward-sloping densities, a broad and empirically plausible set of cases. The proposition guarantees that preventives generate no more than 50% of the producer surplus from treatments, implying that potential deadweight loss exceeds 50% of total disease burden. If demand is log-convex, then Proposition 10 guarantees that preventives generate no more than $1/e \approx 37\%$ of the producer surplus from treatments, implying that potential deadweight loss exceeds

63% of total disease burden. Actual deadweight loss will depend on the actual R&D costs realized for preventives and treatments, but the large potential for deadweight loss means that there is some realization of R&D costs for which the social benefit from eliminating disease burden would be mostly dissipated.

# 4. Generalizations

In this section we show that the insights obtained from analysis of the benchmark model are robust to a number of generalizations. A series of subsections explores imperfect efficacy, side effects, production costs, a broad range of models of competition among suppliers, and third-party purchases. We will show that the key welfare results from Section 3 continue to hold in this more general setting.

## 4.1. Expanded Parameter Space

This subsection relaxes the assumption that all products are perfectly safe, effective, and costless to manufacture. Let $c_j \geq 0$ be the present discounted value of the marginal cost of manufacturing product $j \in \{p,t\}$ and administering it to a consumer. Let $e_j \in [0,1]$ be the efficacy of product $j$, i.e., the probability that product $j$ prevents the consumer from experiencing harm from the disease. Let $s_j \geq 0$ be the expected harm of side effects from product $j$, i.e., 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. Variables with stars ($p_j^*$, etc.) are equilibrium values and with double stars ($TS^{**}$, etc.) are first-best values for general parameters $c_j$, $e_j$, $s_j$. Variables with the superscript $o$ ($p_j^o$, etc.) are equilibrium values and with double $oo$ ($TS^{oo}$) are first-best values when the cost, efficacy, and side-effects parameters are returned to their original levels in the benchmark model: i.e., $c_j^o = s_j^o = 0$, $e_j^o = 1$.

Proposition 1 stated that incentives to develop treatments can be socially excessive but not incentives to develop preventives. This result continues to hold under general parameters, as the next proposition states.

**Proposition 11.** *Extend the benchmark pharmaceutical model to allow general values of the parameters $c_j, s_j \in [0,\infty)$ and $e_j \in [0,1]$ for $j = p,t$. The firm never develops a preventive unless it is socially efficient to do so both in equilibrium and in the first best. There exist cases in which the firm develops a treatment but it would have been socially efficient to develop a preventive.*

Furthermore, the upper bound on the welfare distortion found in Proposition 2 continues to hold for more general parameters.

**Proposition 12.** *Extend the benchmark pharmaceutical model to allow general values of the parameters $c_j, s_j \in [0,\infty)$ and $e_j \in [0,1]$ for $j = p,t$. Letting $\rho_X^o = PS_p^o/PS_t^o$, $1-\rho_X^o$ provides a tight upper bound on deadweight loss as a percentage of disease burden; i.e.,*

$$\sup_{\{k_j,c_j,e_j,s_j|j=p,t\}} \left(\frac{DWL}{\mu_X}\right) = 1-\rho_X^o. \tag{11}$$

Notice that the producer surpluses in the proposition do not correspond to some arbitrary configuration of cost, efficacy, and side effects but the original parameter values.

The proofs of Propositions 11 and 12 are complicated by the fact that a third strategy becomes viable for the monopolist with the expanded parameter space, that of producing both a preventive and a treatment. This strategy never emerges in equilibrium with perfectly effective and costless products. In that case, if the monopolist's equilibrium strategy involved producing a treatment, there was no reason for it to also produce a preventive because it could extract 100% of social surplus with the treatment. With an imperfect treatment, the combination of products may be more profitable than a treatment alone. The proofs in the appendix provide expressions for the profit and welfare from the production of both products, which are used to verify that the results in Propositions 11 and 12 are robust to allowing for this additional strategy.

## 4.2. Alternative Market Structures

The analysis so far assumed a monopoly market structure. More realistically, a number of firms could engage in one of a variety of different forms of competition (Bertrand, Cournot, perfect or imperfect cartel, etc.), both at the R&D and product-market stages. The monopoly benchmark is valuable because it is perhaps the simplest setting in which to examine the ideas about surplus extraction developed in this paper. In this subsection we provide another virtue of the monopoly benchmark. It provides a conservative bound on the potential deadweight loss; letting $C$ be the model of competition under consideration, with few restrictions on $C$ these models will be able to generate at least as much deadweight loss for any number of competing firms.

Consider the following model of competition, nesting various alternatives. To streamline the notation, return to the original parametric assumptions of costless manufacturing, no side effects, and perfect efficacy (i.e., $c_j = c_j^o = 0$, $s_j = s_j^o = 1$, and $e_j = e_j^o = 1$ for $j = p,t$.) This builds in symmetry, allowing us to keep track of just the number of firms rather than the vector of firm characteristics. At the start of the game, each of $N$ potential entrants obtain draws $k_{ji} \geq 0$ of fixed costs for the development of the two products $j = p,t$, where $i = 1,\ldots,N$ indexes firms. Firms observe the vector of draws and decide whether or not to enter each market by sinking the relevant fixed cost.

Let $PS_j^*(n_j,n_\ell,n_b)$ be the most any single firm earns in equilibrium in market $j$ given $n_j$ firms enter the market for that product alone, $n_\ell$ firms enter the market for the other product alone, and $n_b$ firms enter both markets, where $n_j,n_\ell,n_b \in \mathbb{N}$. To allow for general forms of competition, we will put few constraints on this function. Assume

$$PS_j^*(n_j,n_\ell,n_b) \leq PS_j^*(1,0,0), \tag{12}$$

capturing the idea that competition destroys industry profits, so a monopoly generates weakly more producer surplus than any other market structure. Further, assume

$$PS_j^*(1,0,0) = PS_j^*. \tag{13}$$

Condition (13) implies that the existence of potential entrants who do not materialize as actual entrants does not constrain the profit that a monopoly can earn.[10] Finally, assume

$$k_{ji} \leq PS_j^*(n_j^*, n_\ell^*, n_b^*) \tag{14}$$

for each $i$ of the $n_j^*$ firms entering the market for product $j = p, t$ alone and

$$k_{pi} + k_{ti} \leq PS_p^*(n_p^*, n_t^*, n_b^*) + PS_t^*(n_t^*, n_p^*, n_b^*) \tag{15}$$

for each $i$ of the $n_b^*$ firms entering both markets. Conditions (14) and (15) are minimal assumptions on the rationality of the entry decision: if either is violated for some $i$, that firm would have gained by staying out of the markets. We have the following proposition, proved in the appendix.

**Proposition 13.** *Consider any model of competition C satisfying conditions* (12)–(15) *and any number of potential entrants $N \geq 1$. The upper bound on equilibrium deadweight loss is weakly higher than under monopoly:*

$$\sup_{\{k_{ji} \geq 0 | j = p,t; i = 1,\ldots,N\}} \left[\frac{DWL(C,N)}{\mu_X}\right] \geq 1 - \rho_X^*, \tag{16}$$

*where $DWL(C,N)$ is the deadweight loss in model C with N firms.*

The proof is fairly simple. With $N$ firms, the entry costs may be sufficiently high for all but one firm that the only feasible outcome involves monopoly. Thus the monopoly distortion is always a possibility with any of the models under consideration. The $N-1$ additional entrants and entry costs just add "degrees of freedom" that can create even greater distortions.

## 4.3. Third-Party Purchases

In our benchmark model, consumers purchase pharmaceuticals directly from the manufacturer. Real-world pharmaceutical markets involve a broad array of different purchasing arrangements and institutions including nonlinear pricing, insurance, and government provision. A detailed analysis of all of these arrangements is beyond the scope of the present paper. Here we will provide a brief analysis of one, the case in which a third party such as the government or health maintenance organization purchases pharmaceuticals from the manufacturer on behalf of its constituents. We will show that the insights from the benchmark model carry over with this alternative purchasing arrangement.

Assume the firm and the third party 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. If bargaining breaks down, the firm resorts to the option of selling directly to consumers on the private market. Assume the third party's objective is to maximize consumer surplus. Its threat point is thus the consumer surplus from private sales.

The firm's Nash-bargaining surplus conditional on its having developed product $j$ is

$$NB_j = \frac{1}{2}[TS^{**} + PS_j^* - CS_j^*], \tag{17}$$

[10] This assumption rules out some forms of contestability along the lines of Baumol, Panzar, and Willig (1982).

a combination of the first-best "pie" toward which parties bargain, $TS^{**}$, plus the firm's threat-point surplus from selling product $j$ on the private market, $PS_j^*$, minus the third-party purchaser's surplus in this threat point, $CS_j^*$. Substituting $TS^{**} = PS_j^* + CS_j^* + SDWL_j^*$ and subtracting $k_j$ to convert (17) into an objective function used to decide which product the firm develops yields $\Pi_j^* + SDWL_j^*/2$. Comparing the firm's objective function with third-party purchasing to the objective function with direct-to-consumer sales, $\Pi_j^*$, we see that they differ only by the term $SDWL_j^*/2$, reflecting the firm's share of the static deadweight loss avoided with third-party purchasing.

This second term mitigates—but does not eliminate—the potential deadweight loss from the firm's bias against preventives. This 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 third-party purchaser. Recognizing that it does not appropriate all the surplus in bargaining, the firm may distort its decision in order to appropriate more surplus. Note that this will be the case even for a third party representing all potential consumers (like a hypothetical consortium of national governments) as long as bargaining takes place after products are developed.

# 5. Other Sources of Heterogeneity

The paper has so far restricted attention to one source of consumer heterogeneity: disease risk, $X$. This source gives consumers private information only in the ex ante period; ex post, the act of seeking treatment reveals the consumer's disease status. In this section we examine alternative sources of heterogeneity with different timing structures. We begin with a general model of the arrival time of consumer private information in the next subsection; further subsections fill in the details for the new cases entailed by the general model.

To streamline the analysis, the same notation used for random variable $X$ will also be used for the marginal distributions for the other random variables we will introduce. Specifically, let $\Theta \in [0, \bar{\theta}]$ denote any positive random variable. Then $\theta$ willl denote a realization of $\Theta$, $F_\Theta(\theta)$ the marginal cumulative distribution function, $\bar{F}_\Theta(\theta) = 1 - F_\Theta(\theta)$ the complementary distribution, $\Phi_\Theta(\theta) = \bar{F}_\Theta(\theta) + \Pr(\Theta = \theta)$ the "demand" function, $\mu_\Theta = \int_0^{\bar{\theta}} \theta dF_\Theta(\theta)$ the mean—also the area under the "demand" function as can be shown using arguments from the proof of Lemma 1, and $REC_\Theta = \max_{\theta \in [0,\bar{\theta}]} [\theta \Phi_\Theta(\theta)]$ the area of the largest inscribed rectangle under the "demand" function. Let $F$ (without a subscript) denote the joint distribution function for all the random variables under consideration.

## 5.1. Timing of Private Information

Consider a generalization of the model in which the consumer's ex ante willingness to pay for a preventive is $xv_0$ and ex post willingness to pay for a treatment is

$$\begin{cases} v_1 & \text{with probability } x \\ 0 & \text{with probability } 1-x, \end{cases}$$

where $v_\tau \geq 0$ are realizations of random variables $V_\tau$ and where $\tau$ indexes periods, with $\tau = 0$ representing the ex ante period when the preventive is sold and $\tau = 1$ representing the ex post period when the treatment is sold. The dividing line between periods comes when the consumer realizes whether he or she has contracted the disease. The $V_\tau$ embody signals that the consumer receives each period of the amount he or she will end up valuing a perfectly effective cure if he or she contracts the disease ex post. To start, we will place few restrictions on $V_\tau$ other than the following consistency requirement:

$$E(V_1|V_0 = v_0, X = x) = v_0. \tag{18}$$

Equation (18) means that the consumer's current signal is the best guess of his or her signal next period given all available private information, required because $v_0$ and $v_1$ are signals of the same ultimate value.[11]

We can express the model in an equivalent form that helps identify the different sources of private information.

**Proposition 14.** *The model introduced in this section can be written equivalently as $v_0 = y$, $v_1 = hy$, where $y$ is a realization of random variable $Y \geq 0$ and $h$ is a realization of random variable $H \geq 0$ that has unit mean and is mean independent of $X$ and $Y$: $E(H|Y = y) = E(H|X = x) = \mu_H = 1$.*

The proposition is a consequence of (18) together with the law of iterated expectations. The proof is provided in the appendix.

The new way of expressing the model points to three sources of private information embodied in three random variables: $X$ embodies private information existing ex ante that disappears upon the realization of disease status ex post, $Y$ private information that exists ex ante that persists ex post, and $H$ private information that arises only ex post upon realization of disease status. For example, $X$ could represent number of sexual partners, affecting the probability of contracting a disease but not necessarily the severity of the disease conditional on contracting it; $Y$ could represent income, wealth, or some other proxy for willingness to pay that is the same before and after disease status is realized; $H$ could represent the severity of harm learned only after the disease is contracted.

The next proposition generalizes Propositions 2 and 12, bounding potential deadweight loss when there are multiple sources of heterogeneity. The statement of the proposition requires some new notation. In the model introduced in this section, the burden of the disease becomes $E(XV_0)$. This can be rewritten $E(XV_0) = E(XY) = \mu_U$, where the first equality follows from $V_0 = Y$ by Proposition 14 and the second holds defining $U = XY$. Further, define $PS^*_{\max} = \max(PS^*_p, PS^*_t)$ to be the maximum producer surplus available from the two products and $PS^*_{\min} = \min(PS^*_p, PS^*_t)$ be the minimum. These are equilibrium producer surpluses of the expanded parameter space: $c_j, s_j \geq 0$, $e_j \in [0,1]$, $j = p,t$. Define the analogous expressions $PS^o_{\max} = \max(PS^o_p, PS^o_t)$ and $PS^o_{\min} = \min(PS^o_p, PS^o_t)$ for producer surpluses given the original parameter values $c_j = s_j = 0$, $e_j = 1$, $j = p,t$.

[11] An additional restriction on $V_\tau$ is that it is common across products $j = p,t$. This means that any variables that are allowed to differ across products (such as $c_j$, $e_j$, $s_j$) cannot also differ across consumers.

**Proposition 15.** *Consider a pharmaceutical market with multiple sources of heterogeneity. An upper bound on deadweight loss as a percentage of disease burden $\mu_U$ is*

$$\sup_{\{k_j,c_j,e_j,s_j|j=p,t\}} \left(\frac{DWL}{\mu_U}\right) \geq \frac{PS^o_{\max}}{\mu_U} - \frac{PS^o_{\min}}{\mu_U}. \tag{19}$$

A few remarks about the proposition are in order. Note first that with multiple sources of heterogeneity, the bound is no longer guaranteed to be tight as in the earlier propositions.[12] Note second that treatments are no longer guaranteed to be more lucrative than preventives. The notation in Proposition 15 allows for the reverse possibility, for which we will derive conditions below. Note third that Proposition 15 nests the earlier results, as can be verified: with heterogeneity in just $X$ and with $Y$ and $H$ normalized to 1, we have that $PS^o_{\max} = PS^o_t = \mu_X$ and that $\mu_U = \mu_X$; thus, $(PS^o_{\max}/\mu_U) - (PS^o_{\min}/\mu_U) = 1 - PS^o_p/PS^o_t$.

Because the benchmark parameter values are central to Proposition 15, we will maintain these ($c_j = s_j = 0$, $e_j = 1$ for $j = p,t$) for the remainder of the section.

Proposition 15 bounds deadweight loss under quite general time-varying heterogeneity in consumer values. Because of the generality of the case, the bound is necessarily abstract. In The theoretical bounds can be refined if one is willing to restrict the number of sources of private information to one or at most two. The remainder of this section undertakes this theoretical analysis. The case in which $X$ is the sole source of private information has been examined exhaustively already, so the analysis will focus on the other random variables $Y$ and $H$ by themselves and in combinations with others.

In any case in which there is private information in a random variable, in the remainder of the section we assume that the firm cannot condition price on the realization of that random variable, either because the firm does not observe it or, if it does, because it is prevented by legal rules or arbitrage constraints from price discriminating on the basis of it. If the firm can price discriminate on the variable, then its pricing problem simplifies to one in which it conditions on the known value of the variable for each of its possible realizations.

## 5.2. Heterogeneity in $Y$

Random variable $Y$ embodies any consumer characteristics that are private information both ex ante and ex post. Thus $Y$ may be income, wealth, functions of these, or any other demographic factor that affects willingness or ability to pay.[13] We will first analyze the simple case in which there is no other source of heterogeneity than $Y$. It is immediate that consumer heterogeneity in $Y$ alone reduces the producer surplus from either product, but does not result in a bias because the firm faces the same private information ex ante when preventives are sold as ex post when treatments are sold. Producer surplus is the same for both products.

[12] Intuitively, the bound in (19) reflects dynamic deadweight loss from the inefficient product choice. With heterogeneity in just disease risk, a treatment can extract all social surplus, limiting the importance of static deadweight loss from super-competitive pricing. With multiple sources of heterogeneity, no product is guaranteed to extract all surplus. Thus static deadweight loss may become more significant than dynamic deadweight loss. Thus (19) is a lower bound on potential deadweight loss from all sources.

[13] Kessing and Nuscheler (2006) also study monopoly vaccine pricing when income is the sole source of consumer heterogeneity. Their dynamic model generates a feedback effect whereby leaving the poor susceptible increases the willingness to pay of the rich.

Next consider combined heterogeneity in $X$ and $Y$. Continue to suppose $H$ takes on one value: its mean, which Proposition 14 shows is equal to 1. We have been unable to obtain meaningful results for arbitrary covariance between $X$ and $Y$. Indeed, the logic of Proposition 4 suggests that no single joint moment like covariance can adequately capture the pattern of association between random variables over their whole joint distribution. Instead, we provide results for three special cases that span the set of possibilities: $X$ and $Y$ are independent; $Y$ is an increasing deterministic function of $X$; and $Y$ is inversely proportional to $X$. While the analysis covers just these three special cases here, it is in fact possible to compare the producer surpluses from preventives and treatments given any specific joint distribution of $X$ and $Y$. We illustrate how to do this in the calibrations in Section 6 using U.S. data on the distributions of disease and willingness to pay proxied by a function of income.

Assume that $X$ and $Y$ are independent. Consider the preventive producer's profit-maximization problem. Recalling the definition $U = XY$ and letting $u$ be a realization of $U$, consumers buy the preventive if $xy = u \geq p_p$. Hence preventive demand is $\Phi_U(p_p)$, and producer surplus is

$$PS_p^o = \max_{p_p \in [0,\bar{u}]} [p_p \Phi_U(p_p)] = REC_U. \tag{20}$$

Next consider the treatment producer's profit maximization problem. Conditional on contracting the disease, a consumer would be willing to buy the treatment as long as his or her willingness to pay $y$ exceeds $p_t$. Because $X$ is independent of $Y$, the fraction of consumers with income $y$ who contract the disease is the mean $\mu_X$. Hence demand for the treatment is $\mu_X \Phi_Y(p_t)$, implying

$$PS_t^o = \max_{p_t \in [0,\infty)} [\mu_X p_t \Phi_Y(p_t)] = \mu_X REC_Y. \tag{21}$$

The producer surpluses in (20) and (21) can be ranked. One of the sources of private information integrates out of (21) and becomes the constant $\mu_X$; (20) retains both independent sources of private information, translating into lower producer surplus. We have the following proposition. (The proof of this and the remaining propositions in this subsection have been omitted from the published paper for space considerations, instead provided in online Appendix B.)

**Proposition 16.** *Assume there is heterogeneity in positive values of $X$ and $Y$ but not $H$. If $X$ and $Y$ are independent, then $PS_p^o/PS_t^o < 1$.*

The proposition says that, starting with heterogeneity in $X$ alone, adding independently distributed heterogeneity in $Y$ cannot reverse the result from Proposition 3 that treatments are more lucrative than preventives. Although adding independently distributed heterogeneity in $Y$ cannot change the sign of the gap between the producer surplus from treatments and preventives, it will reduce the gap as the next proposition shows.

**Proposition 17.** *Adding heterogeneity in $Y$ that is distributed independently from the heterogeneity in $X$ causes $PS_p^o/PS_t^o$ to rise at least weakly (strictly for continuous distributions).*

Next, consider the extreme case of positive correlation, letting $Y$ be a deterministic function of $X$ that is increasing. Ex ante, the two sources of private information compound each other; ex post one of them disappears. The reduction in private information ex post leads treatments to be more lucrative than preventives. Formally, we have the following proposition.

**Proposition 18.** *Assume there is heterogeneity in positive values of $X$ and $Y$ but not $H$. If $Y$ is an increasing, deterministic function of $X$, then $PS_p^o/PS_t^o < 1$.*

Thus far we have not uncovered a case in which preventives are more lucrative than treatments. Such cases do arise as can be seen by considering the extreme case of negative association in which $X$ and $Y$ are inversely proportional: $Y = u/X$ for some constant $u$. In this case the maximum willingness to pay for a preventive would be the same $u$ across consumers, allowing a preventive monopolist to extract all social welfare—the entire disease burden $\mu_U$. A treatment monopolist, on the other hand, cannot fully extract $\mu_U$ if there is nontrivial heterogeneity in $Y$. This leads all the results from Section 3 to flip. Preventives now deliver the first best. As in Proposition 3, the firm is guaranteed to have a bias, only now against treatments. This bias can be quantified and bounded as in Proposition 2, decomposed as in Proposition 7, and shown to depend on the curvature of $\Phi_Y$ as in Proposition 9, and so forth.

## 5.3. Heterogeneity in $H$

Random variable $H$ embodies any private information that arises only ex post. A natural candidate is the harm suffered from the disease, which in some cases is only learned after contracting it. Polio provides an example of a disease for which victims show an extremely wide range of harms, roughly following a power-law distribution. Only around 5% of polio infections result in any symptoms. Of the infections resulting in symptoms, most result in a mild, flu-like illness. Only around 10% of the symptomatic infections (0.5% of total infections) result in severe nerve damage such as afflicted U.S. President Franklin Roosevelt, whose legs were paralyzed by polio (Mueller, Wimmer, and Cello 2005).[14]

We will begin the analysis with the simple case in which there is no other source of heterogeneity than $H$. It is immediate that switching the source of private information from $X$ ex ante to $H$ ex post flips the results from Section 3, just as the results were flipped in the case studied in the previous subsection in which

[14] Although we can name examples of diseases exhibiting significant heterogeneity in harm, this does not necessarily correspond to heterogeneity in $H$, which embodies only the sort of harm that the consumer cannot predict until contracting the disease. Harm that varies with patient age, weight, genetic information, or other characteristics the patient knows ex ante are embodied in $Y$. For example, a positive result from a genetic test for the BRCA1 mutations not only increases the risk of breast cancer but also increases the chance it is the triple-negative form that has a poorer prognosis than others (National Cancer Institute 2009).

An additional reason why $H$ will be narrower than the practical range of harm heterogeneity is that patients often must be treated before the presentation of severe symptoms to avoid the harm from these symptoms. For example, syphilis eventually leads to blindness in about 15% of untreated cases; however, blindness cannot be reversed by antibiotic treatments for syphilis (Euerle and Chandrasekar 2012). This sort of heterogeneity would not be a source of private information for consumers in either the market for preventives or treatments and thus would not generate a bias toward either product.

Further, producers may be better able to discriminate if the heterogeneity is in ex post harm rather than ex ante risk. The producer could offer different versions of the drug, targeting serious cases with a high-priced version with either a high dosage or in a presentation that is suited to be administered in hospitals. The price differentials can be huge: the hospital studied by Lau et al. (2011) paid 35 to 240 times more for the intravenous than the pill form, depending on the drug. Such price discrimination would eliminate $H$ as a source of private information.

$X$ and $Y$ are inversely proportional.[15]

Next, consider combining heterogeneity in $H$ with other sources of heterogeneity. Begin by assuming consumers are heterogeneous in $X$ and $H$ but not $Y$, which takes on the single value $\mu_Y$ for all $i$. By Proposition 14, $X$ and $H$ are mean independent. To derive useful results, we will make the stronger assumption that $X$ and $H$ are stochastically independent. Under these assumptions, the firm's ability to extract surplus with a preventive ex ante depends solely on the shape of $\Phi_X$. This variable already has the correct rescaling to apply decomposition results from Section 3.2. The firm's ability to extract surplus with a treatment ex post depends solely on the shape of $\Phi_{\tilde{H}}$, where $\tilde{H} = H/h^{\max}$ has been appropriately rescaled, dividing by the maximum harm conceivable $h^{\max}$, to apply the decomposition results from Section 3.2. The next proposition spells out the conditions under which one or the other product extracts more surplus.

**Proposition 19.** *Suppose $X$ and $H$ are the only sources of private information, and these are distributed independently. Let $\tilde{H} = H/h^{\max}$. The firm earns more producer surplus from a treatment than preventive if and only if $Z_X[1-\underline{\rho}(\mu_X)] > Z_{\tilde{H}}[1-\underline{\rho}(\mu_{\tilde{H}})]$ and a preventive than treatment if and only if the reverse inequality holds.*

The proposition follows almost immediately from Proposition 7. The proof, provided in online Appendix B, fills in the details.

The proposition says that when the two sources of heterogeneity $X$ and $H$ are independent, the firm's bias can be determined by looking at properties of the demand curves $\Phi_X$ and $\Phi_{\tilde{H}}$ in isolation. If the disease is rare and the distribution of risk has a high Zipf similarity, then the firm will tend to be biased toward treatments. On the other hand if severe harms are rare and the distribution of harms is highly Zipf-similar, then the firm will tend to be biased toward preventives. The proposition implies that the bias could go either way in theory.

Moving to the remaining combination of sources of heterogeneity to be analyzed, suppose consumers are heterogeneous in $H$ and $Y$ but homogeneous in disease risk $X$. Again, to derive useful results, the mean independence between $Y$ and $H$ guaranteed by Proposition 14 will be strengthened to stochastic independence. In this case we have results analogous to Proposition 16 and 17, but with the inequalities flipped because the variable combined with $Y$ involves ex post rather than ex ante heterogeneity. Thus we have that adding independently distributed heterogeneity in $Y$ cannot reverse the firm's bias against treatments found with heterogeneity in $H$ alone but will reduce the bias. For reference, Table 1 summarizes these and the preceding results from this section.

[15] These results may have empirical relevance for polio. Assuming that polio epidemics were widespread, generally independent of income and other demographic factors embodied in $Y$, the results from this paragraph suggest that a firm would have stronger R&D incentives for a polio vaccine than treatment. In fact, a preventive was developed for polio (the Salk vaccine, followed by the Sabin vaccine), but as yet no good pharmaceutical treatments exist for the disease (Howard 2005). Of course, these outcomes could have been driven by the underlying technological possibility set rather than differences in commercial incentives.

**Table 1:** Summary of results for alternative sources of heterogeneity

| | |
|---|---|
| **Firm's bias toward treatment** | **Firm's bias toward preventive** |
| Heterogeneity in $X$ alone | $X$ and $Y$ inversely proportional |
| Independent variation in $X$ and $Y$ | Heterogeneity in $H$ alone |
| $Y$ an increasing function of $X$ | Independent variation in $Y$ and $H$ |
| **Ambiguous bias** | **No bias** |
| Independent variation in $X$ and $H$ | Heterogeneity in $Y$ alone |

Notes: $X$ represents sources of ex ante consumer heterogeneity such as disease risk. $Y$ represents sources of persistent variation such as income or wealth. $H$ represents sources of ex post variation, such as realized disease severity.

# 6. Calibrations

In this section we show how the theory can be used to calibrate the producer-surplus ratio and potential deadweight loss in particular empirical applications. The calibration method does not require any of the assumptions invoked to derive the propositions in the previous section; in principle, the producer-surplus ratio and potential deadweight loss can be calibrated for any market for which the researcher has sufficient information about demand. The information requirement at first seems daunting, requiring knowledge of the shape of the whole distribution of disease risk in the market rather than just the mean or some other moment. We show how to estimate this distribution for a variety of different diseases.

Overall, the calibrations suggest that the biases identified by the theory can be quantitatively important. For example, one of the HIV calibrations presented in Section  generates producer-surplus ratio $\rho_X^* = 0.214$, indicating that potential deadweight loss due to the bias against preventives could potentially dissipate almost 80% of total surplus. To show that the results are not a special feature of sexually transmitted infections, in Section 6.3 we provide calibrations for the disease that is the leading cause of death in the United States, heart disease. We find higher values of $\rho_X^*$ than for HIV but the values are still consistent with potential deadweight loss of nearly 50%.

## 6.1. NHANES Data

The calibrations focus on the U.S. pharmaceutical market because is the world's largest and is widely seen as the driver of firms' R&D decisions. The National Health and Nutrition Examination Survey (NHANES) obtains rich demographic and disease-risk information from a combination of a survey, physical exam, and blood tests. Table 2 provides descriptive statistics for selected variables from the most recent year of the NHANES, 2010, which we use for the calibrations.[16] The sample is half male, 68% non-Hispanic white, 14% Hispanic, and 11% black. The average age is 42.3. The average family income is about three times the

[16] The means and standard deviations are computed using the same sampling weights as we will use in the calibrations to make the results nationally representative. The descriptive statistics and calibrations are similar if the unweighted data is used.

**Table 2:** Descriptive statistics for 2010 NHANES sample

| Variable | Obs. | Mean | Std. dev. | Min. | Max. |
|---|---|---|---|---|---|
| **General demographic variables** | | | | | |
| Male indicator | 6,527 | 0.50 | 0.50 | 0 | 1 |
| White indicator | 6,527 | 0.68 | 0.47 | 0 | 1 |
| Hispanic indicator | 6,527 | 0.14 | 0.35 | 0 | 1 |
| Black indicator | 6,527 | 0.11 | 0.32 | 0 | 1 |
| Age | 6,527 | 42.3 | 14.3 | 18 | 80 |
| Income (percentage of poverty level) | 5,869 | 3.05 | 1.67 | 0 | 5 |
| **STI factors**[a] | | | | | |
| Lifetime sexual partners | 4,479 | 12.7 | 39.1 | 0 | 1,000 |
| Men who have sex with men (MSM) indicator[b] | 2,242 | 0.03 | 0.18 | 0 | 1 |
| **Heart-attack factors**[c] | | | | | |
| Diabetes indicator | 3,938 | 0.09 | 0.29 | 0 | 1 |
| Smoking indicator | 3,938 | 0.20 | 0.40 | 0 | 1 |
| Total cholesterol | 3,938 | 201.8 | 41.5 | 92 | 528 |
| HDL cholesterol | 3,938 | 53.3 | 17.1 | 15 | 179 |
| Systolic blood pressure | 3,938 | 121.1 | 16.8 | 78 | 228 |
| Diastolic blood pressure | 3,938 | 71.3 | 11.8 | 11 | 132 |

Notes: Means and standard deviations computed using survey weights. [a]Non-response rate higher for survey questions related to sexual behavior. [b]Statistics for male subsample. [b]Statistics for age 30–75 subsample, age range to which Wilson *et al.* model applies.

poverty line.[17]

The next set of factors in the table are important in the calibrations for sexually transmitted infections. Presumably because of the sensitive nature of the questions, the response rate is lower, reflected in the lower number of observations. The average lifetime number of sexual partners is 12.7. The high standard deviation, 39.1, is indicative of substantial heterogeneity in the risk of sexually transmitted infections.[18] About 3% of male respondents report having at least one male sexual partner, our criterion for the MSM variable, an indicator set to 1 for men having sex with men.

The next set of factors are important in the calibrations for heart attacks. We present descriptive statistics for the smaller sample of 30–75 year olds for which the model of heart-attack risk we use applies. About 9% of respondents have diabetes and 20% smoke. The rest of the variables are blood-chemistry and pressure measures from exams and blood tests.

[17] We use income relative to the poverty line for income because it is continuous whereas the family income measure is binned into large intervals. Further, the measure we use accounts for family size. All income measures in the NHANES are top-coded. The measure we use is top-coded at five times the poverty line. We use this top code directly for income. Calibration results are similar if we follow Blanchflower and Oswald's (2004) approach of using 1.25 times the top income code for top-coded observations.

[18] The standard deviation is high in part because of a few reports of partners numbering in the hundreds, a handful even as high as 1,000. The calibrations take these reports as true; results are similar if we censor lifetime partners at 100.

**Table 3:** Calibrations of producer-surplus ratio and potential deadweight loss

| | (1) | | (2) | | (3) | |
|---|---|---|---|---|---|---|
| Ages in sample: | All | | 40–49 | | All | |
| Income elasticity: | None | | None | | 0.4 | |
| | $\rho_X^*$ | $\sup_{k_p,k_t}\left(\frac{DWL}{TS^{**}}\right)$ | $\rho_X^*$ | $\sup_{k_p,k_t}\left(\frac{DWL}{TS^{**}}\right)$ | $\rho_X^*$ | $\sup_{k_p,k_t}\left(\frac{DWL}{TS^{**}}\right)$ |
| **HIV calibrations** | | | | | | |
| HIV1: Linear model | 0.263 | 0.737 | 0.296 | 0.704 | 0.403 | 0.358 |
| HIV2: Kaplan model, $\beta = 0.033\%$ | 0.268 | 0.732 | 0.298 | 0.702 | 0.412 | 0.352 |
| HIV3: Kaplan model, $\beta$ varies by demographics | 0.266 | 0.734 | 0.214 | 0.786 | 0.377 | 0.383 |
| Observations | 4,479 | 4,479 | 923 | 923 | 4,095 | 4,095 |
| **HPV calibration** | | | | | | |
| Kaplan model, $\beta = 13.5\%$ | 0.548 | 0.452 | 0.552 | 0.448 | 0.798 | 0.121 |
| Observations | 4,479 | 4,479 | 923 | 923 | 4,095 | 4,095 |
| **Heart-attack calibration** | | | | | | |
| Wilson *et al.* model | 0.427 | 0.573 | 0.450 | 0.550 | 0.623 | 0.229 |
| Observations | 3,938 | 3,938 | 994 | 994 | 3,565 | 3,565 |

## 6.2. Sexually Transmitted Infections

Our first set of calibrations leverage the NHANES question on lifetime sexual partners to form estimates of the distribution of the risk of sexually transmitted diseases in the population, which we then use to calibrate the producer-surplus ratio $\rho_X^* = PS_p^*/PS_t^*$ and potential relative deadweight loss $\sup_{k_p,k_t}(DWL/TS^{**})$ for these diseases. We focus on the case of HIV for several reasons. First, it is an important disease. Second, we have reasonable proxies for the joint distribution of HIV disease risk and income. Third, until the advent of antiretrovirals, HIV virtually always led to AIDS and ultimately death, thus arguably exhibiting less harm heterogeneity than some other diseases, allowing us to focus on disease risk and income heterogeneity for which we have better data.

The set of columns under (1) of Table 3 provide results from calibrations accounting for disease-risk heterogeneity but not income heterogeneity. The calibration in the row labeled HIV1 involves a simple linear mapping from lifetime sexual partners to infection risk with a constant probability of transmission per partner. Figure 6 graphs the resulting demand curve for this calibration. Recall $PS_p^*$ is given by the area of the largest rectangle that can be inscribed under the curve (the shaded rectangle in the figure) and $PS_t^*$ by the area under the curve. It is apparent that $PS_p^*$ is much less than $PS_t^*$; to be precise, $\rho_X^* = PS_p^*/PS_t^* = 0.263$. As shown in the figure, the firm's optimal strategy in this calibration turns out to be to sell the preventive at a price at which 17% of consumers purchase. The producer-surplus ratio can be translated into a bound on potential deadweight loss, according to the formula in Proposition 2, by subtracting it from 1. This gives the value 0.737 reported in column (1) for the HIV1 calibration, suggesting that the bias against preventives

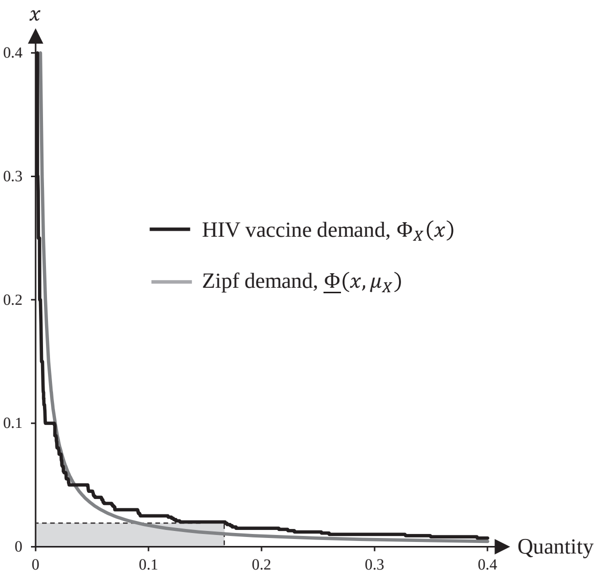

**Figure 6:** Inverse demand curve for calibration in which probability of infection assumed linear in lifetime number of sexual partners. (The demand curves extend beyond the range shown. The axes have been truncated from 1 to 0.4 to aid visualization.)

could dissipate as much as 73.7% of the benefit from a perfectly effective HIV vaccine.

The low producer-surplus ratio in the calibration is due to two factors. First, the calibrated demand curve in Figure 6 looks quite similar to the Zipf demand $\underline{\Phi}_X(x,\mu_X)$ having the same prevalence but attaining the lower bound on potential deadweight loss. Indeed were the axes not truncated to magnify the points of interest in the figure, the calibrated demand curve would be hard to distinguish from the Zipf demand. Using our formal measure of Zipf-similarity, $Z_X = 0.853$ in this calibration. The risk distribution inherits its shape from the Zipfian distribution of lifetime sexual partners owing to the linear mapping between the two.[19] The second factor behind the low calibrated producer-surplus ratio is that, because of the low prevalence of HIV calibrated in the population ($\mu_X = 1.3\%$), the Zipf demand curve has a low producer-surplus ratio, 0.136. Substituting these numbers into (9) returns the ratio $0.263 = 1-0.853(1-0.136)$ reported in Table 3.

The row of calibrations labeled HIV2 replaces 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 $\beta = 0.033\%$, calculated from prevalence rates from Purcell *et al.* (2012) and the per-partner transmission rate from Rockstroh (1995). Details on the calculation of $\beta$ for this and subsequent calibrations of the Kaplan model are provided in online Appendix C. The figure for $\rho_X^*$, 0.268, is quite similar to that from the linear calibration. Indeed, the firm ends up using the same pricing strategy as in the linear calibration. The associated potential deadweight loss is 0.732.

In the row of calibrations labeled HIV3, we allow the $\beta$ in the Kaplan model to vary by sexual orien-

[19] Several previous studies have documented the power-law distribution of the number of sexual partners (e.g., Liljeros *et al.* 2001).

tation, race, and gender. These parameters are calculated from estimates of HIV prevalence for MSM by race from Purcell *et al.* (2012) and HIV transmission rates by gender from Royce *et al.* (1997), combined with population data from the U.S. Census Bureau (2013) and overall HIV rates by race from Centers for Disease Control (2012), which can be used to compute rates for non-MSM by race. These are important sources of disease-risk heterogeneity in the population: estimates in Purcell *et al.* (2012) suggest HIV is over 40 times more prevalent among MSM than non-MSM males and another four times more prevalent among black than white MSM males. This concentration of disease risk in a smaller population leads to an even more Zipf-similar demand curve than in Figure 6, resulting in a slight fall in the producer-surplus ratio to 0.266 and rise in potential deadweight loss to 0.734 times the burden of disease. Interestingly, these barely perceptible changes mask a wholesale change in the firm's pricing strategy: it now sells at a much higher price to a tiny group of the highest-risk consumers. Such swings in strategy are to be expected when demand is Zipf-similar: when all pricing strategies generate similar producer surplus, a small change in the distribution of consumer values can lead to a large change in prices.

Column (2) provides a robustness check, repeating the calibrations from column (1) for a single age cohort, 40–49 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 potential deadweight losses are similar to those in column (1), slightly higher in some cases and slightly lower in others.[20]

Column (3) returns to the full sample, repeating the calibrations from column (1) but now allowing for heterogeneity in willingness to pay $Y$ along with infection risk $X$, as modeled in Section 5.2. We maintain the assumption from that section that the firm cannot price discriminate based on $Y$. To get a measure of $Y$ from the data, we take it to be a function solely of income, in particular taking the elasticity of healthcare expenditure with respect to income to be 0.4 based on empirical estimates.[21] An individual's demand for a preventive equals his or her disease risk $x$ multiplied by $y$. Producer surplus from a preventive is calculated as the rectangle of maximum area under this inverse demand curve. The demand curve for a treatment is constructed by ordering consumers by $y$ and then stepping off the expected drug quantity $x$ each consumer would buy at this reservation price.

In all three HIV calibrations, $X$ and $Y$ are negatively correlated, with a correlations ranging between $-2\%$ and $-5\%$. The analysis from Section 5.2 suggests that adding $Y$ negatively correlated with $X$ to a market can mitigate the bias against preventives. This suggestion is borne out comparing column (3) to (1): we see that accounting for heterogeneity in income increases the calibrated producer-surplus ratio by between 10 and 15

[20] We conducted other robustness checks, not reported in Table 3. We found similar results as in column (1) from calibrations run on 2004 NHANES data. We also found similar results from calibrations run on 1989–2004 data from the General Social Survey.

[21] Getzen (2000) surveys empirical studies of the income elasticity of health expenditures. For purposes of the table, we are interested in the U.S. income elasticity of out-of-pocket expenditures. This is provided by the handful of studies using U.S. micro data from an historical period when most of the population was uninsured. The 0.4 figure, estimated by Anderson, Collette, and Feldman (1960) using 1953 data, is in the middle of the $[0.2, 0.7]$ range from these studies. Micro studies using data from the modern era with more insured consumers find income elasticities near zero. Using such an income elasticity would generate the same results in column (1).

percentage points. Though the bias against preventives is reduced, the most detailed calibrations in column (3), HIV3, still suggest that the producer surplus from a preventive is only 37.7% that from a treatment. Because disease risk is no longer the sole source of consumer heterogeneity, the generalized formula for potential deadweight loss from Proposition 15 must be used in place of Proposition 2. Comparing the results in column (3) to (1), we see that accounting for heterogeneity in willingness to pay cuts potential deadweight almost in half. Though smaller, the potential deadweight loss can still be substantial, in the most detailed calibration 38.3% of total disease burden.

As a counterpoint to the calibrations for HIV, Table 3 adds a set of calibrations for a much more common sexually transmitted infection, HPV. These calibrations are directly comparable to the HIV2 calibrations—both are Kaplan models with fixed values of $\beta$—but $\beta$ is increased from 0.033% to 13.5%. This value of $\beta$ is calculated by combining estimates of the HPV prevalence rate from Dunne *et al.* (2007) with data on the HPV transmission rate from Hernandez *et al.* (2008). The potential deadweight loss measure in the HPV calibrations is about half that for HIV2 on average across all columns. With a disease as prevalent as HPV, the disease risk must be fairly homogeneous, bounding the bias against preventives as Figure 4 shows for large $\mu_X$. The difference between HPV and HIV2 is mainly the prevalence of the two diseases. The Zipf index for the HPV calibration ($Z_X = 0.75$) is quite close to that for the HIV2 calibration ($Z_X = 0.83$).

While many factors outside those we model may be at play, it is worth noting several facts consistent with the calibrations. Only recently has any preventive for HIV become available: Truvada. Truvada was initially developed to be a treatment, approved for that use by the U.S. Food and Drug Administration (FDA) in 2006, not approved for use as a preventive until eight years later (GEN News Highlights 2012). Truvada has a only a niche market as a preventive. The 2014 guidelines issued by U.S. Centers for Disease Control (CDC) recommended its use as a preventive only for such high-risk individuals as men who have unprotected sex with men, drug injectors, and their sexual partners, estimated to be less than 0.2% of the U.S. population (McNeil 2014). The case for HPV is quite different. An preventive for HPV was developed more quickly than for HIV: the HPV vaccine Gardasil was approved by the FDA in 2006. Unlike Truvada, it is a new product, not a re-purposed treatment, and is a recommended by the CDC for all U.S. boys and girls.

## 6.3. Heart Disease

The last row in Table 3 provides calibrations for heart disease, the leading killer in the United States. We derive an estimate of the distribution of disease risk from the influential Framingham Heart Study, reported in Wilson *et al.* (1998). These estimates are available to individuals through the use of a risk calculator widely available on the Internet. Some medicines such as beta blockers are used primarily as preventives for heart attacks, not treatments. Others such as ACE inhibitors are used as treatments, not preventives. Still others such as cholesterol-lowering medications can be used in both capacities. Thus the question of which category firms decide to invest in is interesting for this condition.

The specific condition examined in the calibration is the risk of a heart attack over a ten-year horizon. Wilson *et al.* (1998) estimate the risk of this condition as a function of gender, age, the other risk factors in the bottom rows of Table 2, and their interactions. Calibrating the risk distribution for the subsample of 30–75 year olds, to which the Wilson *et al.* model applies, and then computing the firm's optimal pricing strategies, in column (1) we find a producer-surplus ratio of 0.427 and a potential deadweight loss of 0.573.

Age is an important risk factor in heart attacks. This may lead to medical guidelines specifying that preventives be administered to people of a certain age, effectively neutralizing age as a source of heterogeneity among consumers of the preventive. The calibration in column (2), which restricts the sample to the 40–49 age range, helps account for this possibility. The producer-surplus ratio rises slightly and potential deadweight loss falls slightly relative to column (1). The calibration in column (3) considers a potential case with heterogeneity in willingness to pay for healthcare expenditures given by the same function of income used for the other diseases. Because diabetes, smoking, and other risk factors are negatively correlated with income, this ends up increasing the producer-surplus ratio and reduce potential deadweight loss.

## 7. Empirical Tests

In this section we present a first-pass empirical test of the theory. We will see whether the factors predicted to influence $\rho_X^*$ show up as measurable differences in the types of pharmaceuticals developed. According to the decomposition in equation (10), the most important such factor is Zipf similarity $Z_X$ of the risk distribution as well as prevalence $\mu_X$. We will test whether $Z_X$ and $\mu_X$ affect the probability that a vaccine (the preventive we study) has been developed relative to the probability that a drug (the treatment we study) has been developed over the last century for a sample of microorganisms causing infectious diseases.

Direct computation of $Z_X$ would require the sort of detailed information used in the calibrations. This level of detail on the current distribution of disease risk is not systematically available for a cross-section of diseases, let alone the distribution in the state of nature before any products were developed. Thus we take a different approach in this section, looking for any factor that might lead the risk distribution to be Zipf-similar and combining all such factors into a single indicator $IZ_m$, where $m$ indexes markets (equivalent here to a disease). We will test whether, as implied by the decomposition equation, an increase in $IZ_m$ reduces the probability of vaccine relative to drug development.[22]

This difference-in-differences approach provides power against general alternatives. Factors in $IZ_m$ correlated with Zipf-similarity may also correlate with demand levels. Thus $IZ_m$ may proxy for both the shape as well as the level of demand. It would not be surprising to find fewer vaccines in low-demand markets. To have power against such general alternatives, we will not merely show that an increase in $IZ_m$ decreases vaccine development but that it decreases vaccine development more than it does drug development. Comparing

[22] The left-hand-side variable in the decomposition formula (10) is a ratio of producer surpluses, not entry probabilities. While it is intuitive that a change in the producer-surplus ratio should translate into an analogous change in relative entry probabilities, we provide formal details for this result in online Appendix D. We construct a simple entry model and show that an increase in $IZ_m$ directly reduces the probability of vaccine entry but has no direct effect on the probability of drug entry in market $m$. Indirect effects arise in the model because entry probabilities are strategic substitutes (à la Bulow, Geanakoplos, and Klemperer 1985). A reduction in the probability of vaccine entry makes drug entry more attractive, feeding back to further decreases in the probability of vaccine entry and increases in the probability of drug entry. These indirect effects only reinforce the differential effect of $IZ_m$ on the probability of vaccine versus drug entry.

**Table 4:** Descriptive statistics for sample of CDC-notifiable diseases

| Variable | Mean | Std. dev. | Min. | Max. |
|---|---|---|---|---|
| Vaccine developed | 0.37 | 0.49 | 0 | 1 |
| Drug developed | 0.76 | 0.43 | 0 | 1 |
| Indicator of Zipf similarity ($IZ_m$) | 0.45 | 0.50 | 0 | 1 |
| Childhood onset | 0.14 | 0.35 | 0 | 1 |
| Bacterial | 0.55 | 0.50 | 0 | 1 |
| Viral | 0.33 | 0.47 | 0 | 1 |
| Parasitic | 0.10 | 0.31 | 0 | 1 |
| Fungal | 0.02 | 0.13 | 0 | 1 |
| Prevalence[a] | 0.26 | 0.84 | 0 | 4.74 |
| Producer-surplus-ratio bound[b] | 2.62 | 2.73 | 0 | 7.21 |

Notes: Sample has 58 disease-observations. All variables except those noted *a* or *b* are indicators. [a]Measured as 1,000 cases in 1944. [b]Computed by dividing prevalence by U.S. population to express prevalence as percentage, $\mu_X$, mapping $\mu_X$ into $\underline{\rho}(\mu_X)$ as in Figure 4, then expressing as percentage.
Sources: All variables except those noted *a* or *b* from Harpavat and Nissim (2001), a widely used teaching reference, supplemented by the microbiology reference Mandell, Bennett, and Dolin (2009). [a]*Morbidity and Mortality Weekly Report* (various dates, spanning 1944–2007). [b]Authors' calculations described in the notes.

vaccine to drug development effectively allows us to control for the level of demand in market $m$.

## 7.1. Data

Table 4 provides descriptive statistics for our dataset, comprising the 58 diseases classified by the CDC as notifiable. CDC-notifiable diseases are important to public health but exclude ubiquitous ones such as the common cold and flu. The listed variables were collected from the sources indicated in the notes by a team of research assistants including a senior medical student.

The indicator for Zipf-similarity of the risk distribution, $IZ_m$, deserves special comment because it is the regressor of central interest.[23,24] We set $IZ_m = 1$ if a discrete high-risk group could readily be defined from a review of the disease's epidemiology and transmission patterns. Specifically, $IZ_m = 1$ if the disease satisfies at least one of the following conditions: (a) sexually transmitted; (b) transmitted by animal contact; (c) chiefly affects a concentrated population of either hospitalized patients, immuno-compromised individuals, intravenous-drug users, or soldiers; (d) organism has restricted ecological habitat (e.g., tropics for malaria). This is the comprehensive list of factors identified by our research team, corroborated by conversations with other physicians.

[23] To the extent that the factors included in $IZ_m$ are imperfect measures or other important factors have been left out, the power of our tests will be reduced. $IZ_m$ may be imperfect, for example, if it includes factors on which the vaccine manufacturer can price discriminate. Such factors would then not contribute to a bias against vaccines. The only enumerated factors that may suffer from this problem are the last two, (c) and (d). Our results are robust if we omit diseases exhibiting those factors from the regression (see footnote 26).

[24] Our cross-sectional data will not allow us to obtain market-specific estimates of an increase in $IZ_m$ but just an average across markets. Because $IZ_m$ is a crude indicator rather than a continuous measure of Zipf similarity of the risk distribution, the average effect will average across large and small changes in Zipf similarity.

Theory suggests that prevalence is an important factor in both absolute and relative incentives to develop vaccines and drugs. The listed variable is our attempt to measure prevalence as close to the counterfactual state before any product was developed as our sources allow, here cases reported in 1944, the earliest year available in our data sources. Because CDC-notifiable diseases exclude the most ubiquitous ones, the sample diseases are fairly rare. The last variable, the producer-surplus-ratio bound converts the prevalence variable into the bound $\underline{\rho}(\mu_X)$ on the ratio of vaccine to drug producer surplus from Figure 4, expressed as a percentage. The diseases are rare enough that even the most prevalent of them does not much constrain the range of feasible producer-surplus ratios.

### 7.2. Linear-Probability Model

Table 5 reports the results from a linear probability model, regressing an indicator for product (vaccine or drug) availability on $IZ_m$ and other controls using ordinary least squares. Consider the results for the spare specification reported in the set of columns (1). The –0.408 coefficient in column (1a) indicates that vaccines are 40.8 percentage points less likely to have been developed for Zipf-similar diseases, significant at the 1% level. The analogous coefficient in column (1b) indicates that Zipf similarity has no statistically significant effect on drug development. The difference between the vaccine and drug coefficients in column (1c) indicates that Zipf similarity reduces vaccine development 35.8 percentage points more than it does drug development, a difference significant at the 10% level.

The difference between the constant terms in column (1c) indicates that vaccines are less common than drugs, the average disease being 21.9 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 Zipf similarity that may make vaccines harder to market than drugs, such as tendencies for people to invest less in prevention or the greater epidemiological externalities from vaccines.

One concern with results is that $IZ_m$ 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 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 can be partially addressed in (1) 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 $IZ_m$ were proxying for low overall disease burden, one would expect to find a significantly negative effect on drug development as well, but the coefficient on $IZ_m$ in column (1b) is close to 0. The result in column (1c), 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 richer specification in (2), which adds an explicit prevalence measure as well as other variables mainly intended to control for development costs. The fixed effects for type of organism causing the disease (bacterium, virus, parasite, fungus) control for the possibility that

**Table 5:** Impact of infection-risk heterogeneity on product development

| | (1) | | | (2) | | | (3) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Vaccine developed | Drug developed | Difference | Vaccine developed | Drug developed | Difference | Vaccine developed | Drug developed | Difference |
| Variable | (1a) | (1b) | (1c) = (1a) − (1b) | (2a) | (2b) | (2c) = (2a) − (2b) | (3a) | (3b) | (3c) = (3a) − (3b) |
| $IZ_m$ | −0.408$^{***}$ | −0.050 | −0.358$^{*}$ | −0.368$^{***}$ | −0.038 | −0.330$^{**}$ | −0.394$^{***}$ | −0.031 | −0.380$^{**}$ |
| | (0.115) | (0.116) | (0.184) | (0.131) | (0.083) | (0.138) | (0.132) | (0.082) | (0.138) |
| Viral | | | | 0.196 | −0.696$^{***}$ | 0.892$^{***}$ | 0.226$^{*}$ | −0.727$^{***}$ | 0.953$^{***}$ |
| | | | | (0.122) | (0.117) | (0.145) | (0.128) | (0.104) | (0.142) |
| Parasitic | | | | −0.409$^{***}$ | −0.034 | −0.374$^{***}$ | −0.389$^{***}$ | −0.053 | −0.336$^{***}$ |
| | | | | (0.123) | (0.031) | (0.120) | (0.117) | (0.046) | (0.111) |
| Fungal | | | | −0.103 | −0.003 | −0.100 | −0.023 | −0.087 | 0.064 |
| | | | | (0.076) | (0.045) | (0.082) | (0.119) | (0.062) | (0.135) |
| Childhood onset | | | | 0.420$^{***}$ | −0.222$^{*}$ | 0.642$^{***}$ | 0.382$^{***}$ | −0.200$^{*}$ | 0.582$^{***}$ |
| | | | | (0.129) | (0.114) | (0.130) | (0.131) | (0.110) | (0.133) |
| Prevalence | | | | −0.020 | −0.013 | −0.006 | | | |
| | | | | (0.026) | (0.013) | (0.043) | | | |
| Producer-surplus-ratio bound | | | | | | | 0.020 | −0.024$^{**}$ | 0.044$^{*}$ |
| | | | | | | | (0.020) | (0.011) | (0.022) |
| Constant | 0.563$^{***}$ | 0.781$^{***}$ | −0.219 | 0.471$^{***}$ | 1.041$^{***}$ | −0.570$^{***}$ | 0.417$^{***}$ | 1.101$^{***}$ | −0.684$^{***}$ |
| | (0.089) | (0.074) | (0.134) | (0.120) | (0.043) | (0.121) | (0.128) | (0.052) | (0.124) |
| $R^2$ | 0.175 | 0.003 | 0.232 | 0.434 | 0.681 | 0.609 | 0.445 | 0.702 | 0.622 |

Notes: Ordinary least squares regressions in which dependent variable is an indicator for development of product. Bacterial is omitted organism category in the restricted-sample regressions. Coefficients in column (1c) can be obtained equivalently by either differencing (1a) and (1b) or estimating a regression in which the observations from (1a) and (1b) are stacked, including interactions of all variables with a drug-development indicator, and reporting those interactions. The reported standard errors and $R^2$ in (1c) are from this stacked regression. Entries in (2c) and (3c) obtained similarly. 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.

certain technologies are well-suited to certain organisms and not others; for example, it is believed to be technologically easier to develop vaccines than drugs for viral disease. Greater subsidies may be available for vaccines against childhood diseases because they are easily integrated into childhood immunization programs.

Of course, many other factors are important determinants of product development, including ease of the science involved, other cost factors, government subsidies, and as discussed in Section 5.3 particular forms of harm heterogeneity. Lacking data on these factors, they are included in the error term. We have no particular reason to expect these factors to be systematically correlated with $IZ_m$, but they could be, so this remains a possible threat to identification.

To address this threat, one robustness check is to see if the coefficient on $IZ_m$ changes as we introduce the organism fixed effects and other factors for which we do have data. The coefficients on $IZ_m$ in (2) are in fact quite similar to those in (1), with Zipf similarity decreasing the probability of vaccine development by a statistically significant 36.8 percentage points, but having essentially no effect on drug development, resulting in a differential effect on vaccines versus drugs reported in column (2c) of 33.0 percentage points, now significant at the 5% level.[25,26]

The additional controls in (2) are of some independent interest. Compared to the omitted bacterial category, drugs are significantly less likely to be developed for viral diseases and vaccines less likely to be developed for parasitic diseases. These results are consistent with widespread scientific views about the technological difficulties involved in treating viruses with drugs and in vaccinating against parasites. Vaccines are significantly more likely to be developed for diseases that disproportionately affect children and drugs significantly less likely. This is consistent with the lower cost of delivery of vaccines that can be integrated into childhood immunization programs. Prevalence does not show up as important in the regressions. One explanation is that prevalence does not vary much among the CDC-notifiable diseases in our sample. To the extent prevalence varies, if the CDC determines notifiability on the basis of aggregate health burden, this could generate a negative correlation between prevalence and harm, which would bias the coefficient on prevalence toward zero since the harm is omitted from the regressions.

The regressions in the set of columns (3) substitute the producer-surplus-ratio bound $\underline{\rho}(\mu_X)$ for prevalence $\mu_X$. The two variables are monotonically related, but $\underline{\rho}(\mu_X)$ is in a form connected by theory to relative incentives for product development. We now see some evidence that an increase in the bound increases the

[25] As discussed in Oster (2013), the fact that adding controls substantially boosts the regressions' $R^2$ while leaving the difference-in-differences effect of $IZ_m$ unchanged is further reassurance that the extent of omitted-variable bias is limited.

[26] We performed a variety of other robustness checks not reported here for space considerations. Marginal effects from probits or logits are very similar to the ordinary least squares coefficients in Table 5. To see if one of the four factors behind the $IZ_m$ indicator drives the results, we ran the regressions on subsamples excluding diseases exhibiting each factor in sequence. The results showed the same broad pattern as in column (1), although the difference in development probability was no longer significant when STIs were excluded. We ran the regressions on subsamples excluding different types of organisms. We excluded fungal and parasitic diseases because no vaccines have been developed for such organisms in our sample; this exclusion turned out to strengthen the results. We excluded bacterial diseases, which can be treated with antibiotics. Because each antibiotic typically combats many bacterial diseases, the decision to develop it may have been based on the a portfolio of diseases rather on the characteristics of any one. The exclusion cut the sample in half and reduced the size and significance of the results, but the results have the same basic pattern as in column (1). Results from these robustness checks reported in online Appendix E.

probability of vaccine development more than drug development, but the difference is only marginally statistically significant and is as much due to a reduction in drug development as to an increase in vaccine development.[27]

## 8. Conclusion

R&D incentives depend on innovators' capacity to capture the social value of their innovations by exercising market power. This in turn depends on the shape of the distribution of consumer values for their innovations. We have argued that holding the sum constant, changes in the distribution of values across consumers between the time disease preventives are sold and the time disease treatments are sold will create wedges between the ability of preventive and treatment manufacturers to extract consumer surplus, potentially distorting R&D incentives. In the benchmark model in which consumers differ only in disease risk, Zipf-similar risk distributions generate the largest gap between the producer surplus from a preventive and treatment, which can be translated dollar for dollar into a bound on potential deadweight loss. Measured relative to total surplus, this potential deadweight loss is particularly large for rare diseases. We develop tools to quantify the extent of potential distortions for particular distributions, and use these tools to calibrate potential distortions of R&D incentives in the cases of HIV and heart attacks, finding large gaps between the producer surplus associated with preventives and treatments. Finally, we show that vaccines, but not drugs, are less likely to be developed for diseases with substantial risk heterogeneity.

Considerable scope for further work remains. While this paper focuses on the case of pharmaceuticals, much of the analysis about how the shape of the distribution of valuation influences the ability of firms to extract surplus applies to general product markets. A companion paper, Kremer and Snyder (2015), explores these general implications, tying Zipf-similarity of a suitably rescaled demand curve (to have unit domain and range) to static and dynamic deadweight loss, R&D incentives, gains from optimal subsidies, and losses from banning price discrimination. A considerable range of products may approach the upper bounds we derive for deadweight loss: many economic variables, from income to city size, have been found to follow power laws (at least in the upper tail; see Gabaix 2009); the Zipf-similarity we found for distributions of HIV and heart-attack risk may extend to these other domains as well. The companion paper provides an initial set of calibrations of demand on a global market, assuming consumer unit values for a product are unit elastic in income and assuming income follows Pinkovskiy and Sala-i-Martin's (2009) estimates of the world distribution of income. Calibrated demand for the most recent year of estimates is strikingly similar to STRZ demand, with a Zipf similarity of 83%, indicating a disturbingly high potential for deadweight loss on the global market.

While we have provided bounds on relative deadweight loss applicable to arbitrary market structures, it would be useful to tighten the bounds for particular oligopoly models or distributions of development costs

[27] The model in online Appendix D provides an explanation of the reduction in the probability of drug development—as a strategic response to an increase in the probability of vaccine development, corresponding to the movement along the drug manufacturer's downward-sloping best-response function (where firms' best-response functions determine their entry decisions).

across firms. It would also be useful to allow for nonlinear cost functions and to allow costs, efficacy, side effects, and other parameters that vary across products to also have a distribution across consumers.

While a formal analysis of policies to counteract the identified distortions along the lines of Weyl and Tirole (2012) is beyond the scope of this paper, we can at least draw some qualitative conclusions for policy from the analysis. To the extent that policymakers can identify markets where the shape of the distribution of consumer values makes it difficult for producers to extract consumer surplus, they may wish to target R&D subsidies to these markets. The market for an HIV preventive is a case in point. Industry observers such as Thomas (2002) claim that the profit potential in this market is low relative to the potential social benefit. Our analysis provides one explanation for the low profit potential: firms may have difficulty capturing social surplus because of the low prevalence of the disease and the Zipf-similarity of the disease risk distribution. Given the seriousness of the disease, total surplus, and hence the potential absolute magnitude of the distortions, are likely high. This may provide one potential rationale for programs such as the International AIDS Vaccine Initiative.

To the extent that policymakers can identify markets where the shape of the distribution of consumer values makes it difficult for producers to extract consumer surplus, they may wish to target R&D subsidies to these markets. Thus, for example, the analysis in this paper suggests that, consistent with the claims of industry observers, such as Thomas (2002), the share of potential social value captured by developers of HIV preventives may be low due to the low prevalence of the disease and the Zipf-similarity of the disease risk distribution. Because the disease is so serious, total surplus, and hence the potential absolute magnitude of the distortions, are likely high. This may provide one potential rationale for programs such as the International AIDS Vaccine Initiative.

Zipf-similar distributions not only generate large dynamic distortions, but as explored more fully in Kremer and Snyder (2015), the companion paper, they also generate static distortions. At least in theory, price controls could greatly increase consumer welfare with only a small impact on R&D incentives when consumer demand is Zipf similar. Ordinarily, large forced price reductions could be expected to ruin a firm's R&D incentives. With a Zipf distribution of values, the firm may be largely indifferent between the existing high price and a low control price, allowing the control to be implemented without much effect on producer surplus or R&D incentives. One policy that could potentially improve both static pricing efficiency and dynamic R&D incentives would be for governments to subsidize consumers' purchases of preventives, if possible, targeting the subsidy to consumers with low (net of production costs) valuations. This would robustly increase dynamic R&D incentives. As noted in Section 4.3, if governments bargain with pharmaceutical producers over price after R&D costs have been sunk, with the threat point in the case of a breakdown in negotiations being direct sales to consumers, dynamic distortions may be attenuated but will not be eliminated. A hold-up problem (à la Klein, Crawford, and Alchian, 1978) will remain. However, if the government is able to commit in advance to purchase at an appropriate price or to provide an appropriate subsidy, the first-best dynamic incentives can be achieved. If the distribution of consumer values is close to being Zipf, the consumer subsidy has the potential for dramatic improvements in static (pricing) efficiency as well. Even a modest subsidy would be enough to make a monopolist that is nearly indifferent among a range of prices strictly prefer the lowest of these prices.

An example of ex ante bargaining is provided by advance market commitment programs for vaccines (see Kremer and Glennerster, 2004). A pilot program of this type was implemented for pneumococcal vaccine by a group of donors which committed \$1.5 billion to help finance purchase of a vaccine covering strains of the disease common in developing countries at a price targeted to be between unit production cost and the vaccine's social value (see Snyder, Begor, and Berndt 2011 for description and analysis of the Pneumococcus Advance Market Commitment). The analysis in this paper suggests that advance market commitments address may be particularly well suited for disease preventives when the distribution of disease risk is Zipf-similar, as in the case of HIV.

The policies of the U.S. Advisory Committee on Immunization Practices (ACIP) may de facto act in a way similar to advance market commitments. The ACIP analyzes the cost effectiveness of new vaccines, recommending that a vaccine be added to the schedule of immunizations eligible for government subsidies if its price is below a cost-effectiveness threshold. While the ACIP's recommendations are not legally binding, they are almost always followed in practice. Firms respond by pricing at this threshold. This policy effectively commits the government to subsidizing vaccine purchases that would be socially efficient in a way that both addresses static monopoly pricing distortions and generates R&D incentives for vaccines corresponding to their estimated social value.[28]

[28] See Barder, Kremer, and Levine 2004, chapter 2, for a discussion of these examples.

## Appendix A: Proofs of Propositions

**Proof of Lemma 1:** First-best preventive surplus is $TS_p^{**} = TS_p(0) = PS_p(0) + CS_p(0) = CS_p(0) = \int_0^\infty Q_p(x)dx = \int_0^1 \Phi_X(x)dx$. The first and third equalities follow from the assumption $c_p = 0$ in the benchmark model, and the last from (1). Similarly, first-best surplus for a treatment is $TS_t^{**} = \int_0^\infty Q_t(x)dx = \int_0^1 \mu_X dx = \mu_X$, where the second equality follows from (2).

The proof is completed by showing $TS_p^{**} = TS_t^{**}$. We have $\mu_X = \int_0^1 x\, dF_X(x) = [1 - \int_0^1 F_X(x)dx] = \int_0^1 \bar{F}_X(x)dx = \int_0^1 \Phi_X(x)dx$. The first equality follows the definition of $\mu_X$, the second from integration by parts, the third from the definition of $\bar{F}_X(x)$, and the last from the fact that $\bar{F}_X(x)$ only differs from $\Phi_X(x)$ for at most a countable set of $x$, so their Reimann integrals are equal. *Q.E.D.*

**Proof of Proposition 1:** It remains to construct a case in which $\Pi_t^* > \Pi_p^*$ but $W_p^* > W_t^*$ and $W_p^{**} > W_t^{**}$. Take $X$ to be uniformly distributed on $[0,1]$, $k_p = 0$, and $k_t = 1/5$. One can show $\Pi_p^* = 0.25$, $\Pi_t^* = W_t^* = W_t^{**} = 0.3$, $W_p^* = 0.375$, and $W_p^{**} = 0.5$. *Q.E.D.*

**Proof of Proposition 3:** ($\Longrightarrow$) We will prove the contrapositive. To this end, assume there exists no $x' \in (0,1]$ such that $\Pr(X = x' | X > 0) = 1$. Then $\Pr(X = x^*) < 1$ for $x^* = \text{argmax}_{x \in [0,1]} x\Phi_X(x)$. Thus at least one of the sets $(0, x^*)$ or $(x^*, 1]$ has positive measure.

By Lemma 1,

$$\mu_X - REC_X = \int_0^1 \Phi_X(x)dx - x^*\Phi_X(x^*) \tag{A1}$$

$$= \int_0^{x^*} [\Phi_X(x) - \Phi_X(x^*)]dx + \int_{x^*}^1 \Phi_X(x)dx. \tag{A2}$$

The first term in (A2) is nonnegative because $\Phi_X(x)$ is nonincreasing, implying $\Phi_X(x) \geq \Phi_X(x^*)$ for all $x \leq x^*$. The second term in (A2) is nonnegative because $\Phi_X(x) \geq 0$.

We will show at least one of the terms in (A2) is not just nonnegative but positive, implying (A2) is positive. Suppose $(0, x^*)$ has positive measure. Then $\Phi_X(x) > \Phi_X(x^*)$ for a positive measure of $(0, x^*)$, implying the first term on the right-hand side of (A2) is positive. Suppose $(x^*, 1]$ has positive measure. Then $\Phi_X(x) > 0$ for a positive measure of $(x^*, 1]$, implying the second term on the right-hand side of (A2) is positive. We have shown (A2) is positive in either case, implying $\mu_X > REC_X$, in turn implying $\rho_X^* = REC_X / \mu_X < 1$, where the equality follows from Lemma 2.

($\Longleftarrow$) Assume $\Pr(X = x' | X > 0) = 1$ for some $x' \in (0,1]$. Then $\Phi_X(x) = 1$ for all $x \in [0, x']$ and $\Phi_X(x) = 0$ for all $x \in (x', 1]$. Obviously $x^* = x'$. By Lemma 2,

$$\rho_X^* = \frac{REC_X}{\mu_X} = \frac{x^*\Phi_X(x^*)}{\int_0^1 \Phi_X(x)dx} = \frac{x' \cdot 1}{\int_0^{x'} 1\, dx} = \frac{x'}{x'} = 1.$$

*Q.E.D.*

**Proof of Proposition 4:** Suppose $X$ has a two-type distribution. Its distribution can be characterized by three parameters: risk for the low type $x_1 \in (0,1)$, risk for the high type $x_2 \in (x_1, 1]$, and the probability of the low type $\pi_1 \in (0,1)$. The mean is $\mu_X = \pi_1 x_1 + (1-\pi_1)x_2$, implying

$$x_2 = \frac{\mu_X - \pi_1 x_1}{1 - \pi_1}. \tag{A3}$$

Therefore, the distribution can equivalently be characterized by the three parameters $x_1 \in (0,1)$, $\pi_1 \in (0,1)$, and $\mu_X \in (x_1, 1)$.

Let $x^* = \text{argmax}_{x \in [0,1]} x\Phi_X(x)$. Then

$$x^*\Phi_X(x^*) = \max\{x_1, (1-\pi_1)x_2\} = \max\{x_1, \mu_X - \pi_1 x_1\} \tag{A4}$$

substituting from (A3). By Lemma 2,

$$\rho_X^* = \frac{x^*\Phi_X(x^*)}{\mu_X} = \max\left\{\frac{x_1}{\mu_X}, 1 - \frac{\pi_1 x_1}{\mu_X}\right\}, \tag{A5}$$

substituting from (A4). The $n$th raw moment is

$$E(X^n) = \pi_1 x_1^n + (1-\pi_1)x_2^n = \pi_1 x_1^n + \frac{(\mu_X - \pi_1 x_1)^n}{(1-\pi_1)^{n-1}},$$

substituting from (A3). Differentiating,

$$\frac{\partial E(X^n)}{\partial x_1} = \frac{n\pi_1}{(1-\pi_1)^{n-1}}\left[(x_1 - \pi_1 x_1)^{n-1} - (\mu_X - \pi_1 x_1)^{n-1}\right].$$

This is negative for $n \geq 2$ because $\mu_X > x_1$. Thus $E(X^n)$ is decreasing in $x_1$. On the other hand, (A5) can decrease or increase in $x_1$ holding $\mu_X$ constant. In particular, (A5) is increasing in $x_1$ if $\mu_X < (1+\pi_1)x_1$ and decreasing in $x_1$ if $\mu_X > (1+\pi_1)x_1$.

This shows that an increase in the raw moment of order $n \geq 2$ can be accompanied by a change in either direction of $\rho_X^*$. The proof for the central and standardized moments is similar and thus omitted. *Q.E.D.*

**Proof of Proposition 5:** Consider a risk distribution embodied in $X$. By Lemma 2, $\rho_X^* = REC_X / \mu_X$. Let $\underline{\Phi}(x, \mu_X)$ be the demand defined in (8), where $\underline{\rho}(\mu_X)$ is the implicit function defined in (7). The producer-surplus ratio associated with $\underline{\Phi}(x, \mu_X)$ is $\underline{REC}(\mu_X)/\mu_X$, where

$$\underline{REC}(\mu_X) = \max_{x \in [0,1]} [x\underline{\Phi}(x, \mu_X)] \tag{A6}$$

$$= \max_{x \in [0,1]} [\min\{\mu_X \underline{\rho}(\mu_X), x\}] \tag{A7}$$

$$= \mu_X \underline{\rho}(\mu_X). \tag{A8}$$

Equation (A6) follows from the definition of $\underline{REC}(\mu_X)$, (A7) from (8), and (A8) from substituting the highest value of $x$ and noting that $\mu_X \in (0,1)$ and that $\underline{\rho}(\mu_X) \in (0,1)$ for $\mu_X \in (0,1)$. We defer proving $\underline{\rho}(\mu_X) \in (0,1)$ to the proof of Proposition 6. Rearranging, (A8) implies $\underline{\rho}(\mu_X) = \underline{REC}(\mu_X)/\mu_X$, implying that $\underline{\rho}(\mu_X)$ is the producer-surplus ratio associated with $\underline{\Phi}(x, \mu_X)$.

We will show $\underline{\rho}(\mu_X) \leq \rho_X^*$, with strict inequality unless $\Phi_X(x) = \underline{\Phi}(x, \mu_X)$ almost everywhere (a.e.). If $\Phi_X(x) = \underline{\Phi}(x, \mu_X)$ a.e., then $\rho = \underline{\rho}(\mu_X)$, and we are done. For the remainder of the proof, assume $\Phi_X(x) \neq \underline{\Phi}(x, \mu_X)$ for a positive measure of $x$. We have

$$\int_0^1 \Phi_X(x)dx = \mu_X = \int_0^1 \underline{\Phi}(x, \mu_X)dx,$$

where the first equation follows from Lemma 1 and the second from (6). Given that the integrals of the demands are equal but the demands themselves are not, it must be that $\Phi_X(x) < \underline{\Phi}(x, \mu_X)$ for all $x$ in some subset $S_1$ of positive measure and $\Phi_X(x) > \underline{\Phi}(x, \mu_X)$ for all $x$ in another subset $S_2$ of positive measure. For $x \in S_2$, $x\Phi_X(x) > x\underline{\Phi}(x, \mu_X) = \min\{\mu_X \underline{\rho}(\mu_X), x\}$, implying either $x\Phi_X(x) > \mu_X \underline{\rho}(\mu_X)$ or $x\Phi_X(x) > x$. The latter inequality implies $\Phi_X(x) > 1$, a contradiction to $\Phi_X(x)$ being a proper demand curve bounded above by 1 in a market with mass 1 of consumers having unit demand. This proves that for all $x \in S_2$,

$$x\Phi_X(x) > \mu_X \underline{\rho}(\mu_X). \tag{A9}$$

It follows that, for $x \in S_2$,

$$\underline{\rho}(\mu_X) < \frac{x\Phi_X(x)}{\mu_X} \leq \frac{REC_X}{\mu_X} = \rho^*_X.$$

The first step holds by (A9), the next because $REC_X$ is the maximized value of $x\Phi_X(x)$ over $x \in [0,1]$, and the last by Lemma 2. Since $\underline{\rho}(\mu_X) < \rho^*_X$ for some $x \in [0,1]$, the inequality must hold for all $x \in [0,1]$ because $\underline{\rho}(\mu_X)$ and $\rho^*_X$ do not vary with $x$. *Q.E.D.*

**Proof of Proposition 6:** Rather than working with $\underline{\rho}(\mu_X)$, we will initially work with its inverse, denoted $\mu_X(\underline{\rho})$, which is the solution of (7) for $\mu_X$:

$$\mu_X(\underline{\rho}) = \frac{\exp(1)/\underline{\rho}}{\exp(1/\underline{\rho})}. \tag{A10}$$

Differentiating,

$$\mu'_X(\underline{\rho}) = \frac{(1-\underline{\rho})\exp(1)}{\underline{\rho}^3 \exp(1/\underline{\rho})},$$

implying that $\mu_X(\underline{\rho})$ is continuously differentiable, with $\mu'_X(\underline{\rho}) > 0$, for all $\underline{\rho} \in (0,1)$. By the Inverse Function Theorem, its inverse, $\underline{\rho}(\mu_X)$, exists.

Applying l'Hôpital's Rule to (A10),

$$\lim_{\underline{\rho}\downarrow 0} \mu_X(\underline{\rho}) = \lim_{\underline{\rho}\downarrow 0} \left[\frac{-\exp(1)\underline{\rho}^{-2}}{-\exp(1/\underline{\rho})\underline{\rho}^{-2}}\right] = \lim_{\underline{\rho}\downarrow 0} \left[\frac{\exp(1)}{\exp(1/\underline{\rho})}\right] = 0.$$

Furthermore, $\mu_X(1) = 1$. Thus $\mu_X(\underline{\rho}) \in (0,1)$ for all $\underline{\rho} \in (0,1)$, implying $\mu_X(\underline{\rho})$ is a bijection on $(0,1)$, implying its inverse $\underline{\rho}(\mu_X)$ is also a bijection on $(0,1)$. The fact that $\lim_{\underline{\rho}\downarrow 0} \mu_X(\underline{\rho}) = 0$ implies $\lim_{\mu_X \downarrow 0} \underline{\rho}(\mu_X) = 0$, and the fact that $\mu_X(1) = 1$ implies $\underline{\rho}(1) = 1$. By the Inverse Function Theorem, $\underline{\rho}'(\mu_X) = 1/\mu'_X(\underline{\rho}(\mu_X))$, which is positive for all $\mu_X \in (0,1)$ because $\mu'_X(\underline{\rho}) > 0$ for all $\underline{\rho} \in (0,1)$.

The proof of Proposition 5 relied on the claim $\underline{\rho}(\mu_X) \in (0,1)$ for all $\mu_X \in (0,1)$. This claim follows from the fact just established that $\underline{\rho}(\mu_X)$ is a bijection on $(0,1)$. *Q.E.D.*

**Proof of Proposition 8:** A disease-risk distribution with $T$ discrete types can be fully characterized by $2T$ parameters $\{m_\tau\}_{\tau=1}^T$ and $\{x_\tau\}_{\tau=1}^T$ satisfying the following feasibility conditions:

$$m_\tau \in (0,1) \text{ for all } \tau = 1,\ldots,T, \tag{A11}$$

$$\sum_{\tau=1}^{T} m_\tau = 1, \tag{A12}$$

$$0 \leq x_1 \leq \cdots \leq x_T \leq 1. \tag{A13}$$

We will choose these $2T$ parameters so that the distribution is a discrete Zipf distribution. This will allow us to generate a $\rho^*_X$ arbitrarily close to $1/T$. To this end, define type masses

$$m_\tau = \begin{cases} \theta^{\tau-1} & \text{if } \tau > 1 \\ 1 - \sum_{\tau=1}^{T-1} \theta^\tau & \text{if } \tau = 1. \end{cases} \tag{A14}$$

for some $\theta \in (0, 1/2)$. It can be shown that this geometrically declining sequence respects constraints (A11) and (A12). Define the disease risks recursively as follows: set $x_T = 1$, and set

$$x_\tau \sum_{i=\tau}^{T} m_i = x_{\tau+1} \sum_{i=\tau+1}^{T} m_i. \tag{A15}$$

for $\tau = 1,\ldots,T-1$. The left-hand side of (A15) is the profit from charging a price $x_\tau$ and selling to type $\tau$ and higher. The right-hand side is the profit from charging a price $x_{\tau+1}$ and selling to types $\tau+1$ and higher. It is easy to see that the disease risks respect constraint (A13). By definition, $\mu_X = \sum_{\tau=1}^T m_\tau x_\tau$. By construction implicit in (A15), we have $REC_X = x_1$; that is, it is weakly most profitable to charge $x_1$ for the preventive and sell to all consumers. Thus,

$$\frac{\mu_X}{REC_X} = \frac{\sum_{\tau=1}^T m_\tau x_\tau}{x_1} \tag{A16}$$

$$= m_1 + \sum_{\tau=2}^{T} \frac{m_\tau}{m_\tau + \cdots + m_T} \tag{A17}$$

$$= 1 - \sum_{\tau=1}^{T-1} \theta^\tau + \sum_{\tau=2}^{T} \frac{\theta^{\tau-1}}{\theta^{\tau-1} + \cdots + \theta^{T-1}}. \tag{A18}$$

Equation (A16) follows from previous arguments. Equation (A17) holds since it is equally profitable to sell the preventive to all consumers at price $x_1$ or to consumers of types $\tau$ and above at price $x_\tau$, so that $x_1 = x_\tau(m_\tau + \cdots + m_T)$, implying $x_\tau = x_1/(m_\tau + \cdots + m_T)$. Equation (A18) holds by substituting for $\{m_\tau\}_{\tau=1}^T$ from Equation (A14). Taking limits, $\lim_{\theta\to 0}(\mu_X/REC_X) = 1 - 0 + \sum_{\tau=2}^T 1 = T$, or, equivalently, $\lim_{\theta\to 0}(REC_X/\mu_X) = 1/T$. This shows that for any $\epsilon > 0$, and for the definitions of the parameters in (A14) and (A15), we can find $\theta > 0$ such that $REC_X/\mu_X < 1/T + \epsilon$. By Lemma 2, $\rho^*_X = REC_X/\mu_X$. Hence, $\rho^*_X < 1/T + \epsilon$.

To prove $\rho^*_X \geq 1/T$ for all distributions with $T$ discrete types,

$$\begin{aligned} T \cdot REC_X &= T \max_{\tau \in \{1,\ldots,T\}} \left[ x_\tau \left( 1 - \sum_{i=1}^{\tau-1} m_i \right) \right] \\ &\geq T \max_{\tau \in \{1,\ldots,T\}} \{x_\tau m_\tau\} \\ &\geq \sum_{\tau=1}^{T} x_\tau m_\tau \\ &= \mu_X. \end{aligned}$$

Hence $\rho^*_X = REC_X/\mu_X \geq 1/T$. *Q.E.D.*

**Proof of Proposition 9:** The result is a corollary of Anderson and Renault (2003). Assume demand $\Phi_X$ is twice continuously differentiable and $c$-concave. Their Proposition 1 shows $c$-concavity is equivalent to

$$\frac{\Phi''_X \Phi_X}{(\Phi'_X)^2} \leq 1 - c. \tag{A19}$$

Substituting $n = 1$ (representing the monopoly market structure) into their Proposition 5 and taking reciprocals,

$$\frac{PS^*_j}{TS^{**}_j} \geq \left(\frac{1}{1+c}\right)^{1/c} \tag{A20}$$

for all $c > -1$ such that $c \neq 0$. Observe that letting $c = 1$ in (A19) gives the definition of ordinary concavity. Substituting $c = 1$ into (A20) yields $\rho^*_X = PS^*_j/TS^{**}_j \geq 1/2$, the result for concave demand in Proposition 9. If $c = 0$, corresponding to log-concavity, then Proposition 5 from Anderson and Renault (2003) implies $PS^*_j/TS^{**}_j \geq 1/e$, giving the result for log-concave demand in Proposition 10.

The results for convex and log-convex demands in Proposition 9 and 10 can be proved by reversing the previous inequalities. The result for linear demand in Proposition 9 can be proved by noting linear demand is both concave and convex, so $1/2 \le \rho_X^* \le 1/2$.

More generally, the preceding arguments can be used to establish general bounds on $\rho_X^*$ for $c$-concave or $c$-convex demands for any $c > -1$. *Q.E.D.*

**Proof of Proposition 11:** Suppose there exists an equilibrium in which a preventive is developed, either alone or together with a treatment. Letting subscript $b$ denote the associated variable when both products are developed, we have

$$\max(\Pi_p^*, \Pi_b^*) \ge \max(\Pi_t^*, 0). \tag{A21}$$

Because $W_j^* = \Pi_j^* + CS_j^* \ge \Pi_j^*$, we have

$$\begin{aligned} \max(W_p^*, W_b^*) &\ge \max(\Pi_p^*, \Pi_b^*) && \text{(A22)} \\ &\ge \max(\Pi_t^*, 0) && \text{(A23)} \\ &= \max(W_t^*, 0), && \text{(A24)} \end{aligned}$$

where (A23) follows from (A21) and (A24) follows from $\Pi_t^* = W_t^*$, which holds because a treatment extracts all surplus from the ex post homogeneous consumers. Thus there exists a socially efficient outcome in which a preventive is developed. One can similarly show $\max(W_p^{**}, W_B^{**}) \ge \max(W_t^{**}, 0)$, implying that there exists a first-best outcome in which a preventive is developed.

The proof of Proposition 1 provided an example in which a treatment is developed when it is socially inefficient to do so, an example which serves the purposes of this proof as well. *Q.E.D.*

**Proof of Proposition 12:** We will employ a "sandwiching" argument, showing that potential deadweight loss is first weakly greater than, and second weakly less than, $1-\rho^o$. The proof that

$$\sup_{\{k_j,c_j,e_j,s_j \ge 0 | j=p,t; e_j \le 1\}} \left( \frac{DWL}{\mu_X} \right) \ge 1-\rho^o, \tag{A25}$$

draws on a result proved below for a more general context, one allowing for sources of heterogeneity beyond disease risk. In that context, Proposition 15 states

$$\sup_{\{k_j,c_j,e_j,s_j \ge 0 | j=p,t; e_j \le 1\}} \left( \frac{DWL}{\mu_X} \right) \ge \left( \frac{PS^o_{\max} - PS^o_{\min}}{\mu_X} \right), \tag{A26}$$

where $PS^o_{\max} = \max_{j \in \{p,t\}}(PS^o_j)$ and $PS^o_{\min} = \min_{j \in \{p,t\}}(PS^o_j)$. In the special case relevant to the present proposition, with heterogeneity only in disease risk, $PS^o_{\max} = PS^o_t = \mu_X$ and $PS^o_{\min} = PS^o_p$, implying $(PS^o_{\max} - PS^o_{\min})/\mu_X = 1 - PS^o_p/PS^o_t = 1-\rho^o$. Thus, the right-hand side of (A26) equals $1-\rho^o$, establishing (A25).

Turning to the other side of the "sandwich," we have

$$\begin{aligned} &\sup_{\{k_j,c_j,e_j,s_j \ge 0 | j=p,t; e_j \le 1\}} (DWL) \\ &\le \max_{\ell,m \in \{p,t,b,n\}} \begin{cases} \displaystyle\sup_{\{k_j,c_j,e_j,s_j \ge 0 | j=p,t; e_j \le 1\}} (W_\ell^{**} - W_m^*) \\ \text{subject to } \Pi_m^* = \max(\Pi_p^*, \Pi_t^*, \Pi_b^*, 0). \end{cases} \end{aligned} \tag{A27}$$

Implicit in (A27) is that the monopolist's generic strategies, indexed by $\ell$ and $m$, can include the possibility of producing both products (denoted $b$) or neither (denoted $n$) in addition to producing a preventive ($p$) or treatment ($t$) alone. The generic strategies are chosen to maximize the wedge $W_\ell^{**} - W_m^*$ subject to the constraint that $m$ is an equilibrium strategy for the monopolist, which requires $\Pi_m^*$ to be the highest of the profits. For generic strategy $m$,

$$W_m^* = CS_m^* + \Pi_m^* = CS_m^* + \max(\Pi_p^*, \Pi_t^*, \Pi_b^*, 0), \tag{A28}$$

where the first equality follows from the definition of $W_m^*$ and the second from substitution of the constraint from (A27). For the generic strategy $\ell$, by definition of $W_\ell^{**}$,

$$W_\ell^{**} = CS_\ell^* + SDWL_\ell^* + \Pi_\ell^*. \tag{A29}$$

Combining (A28) and (A29) and rearranging,

$$\begin{aligned} &W_\ell^{**} - W_m^* \\ &= CS_\ell^* + SDWL_\ell^* - CS_m^* + [\Pi_\ell^* - \max(\Pi_p^*, \Pi_t^*, \Pi_b^*, 0)] && \text{(A30)} \\ &\le CS_\ell^* + SDWL_\ell^*. && \text{(A31)} \end{aligned}$$

The second equality follows from the facts that $CS_m^* \ge 0$ and that the term in square brackets is non-positive. Substituting (A31) into the right-hand side of (A27),

$$\begin{aligned} &\sup_{\{k_j,c_j,e_j,s_j \ge 0 | j=p,t; e_j \le 1\}} (DWL) \\ &\le \max_{\ell,m \in \{p,t,b,n\}} \left[ \sup_{\{k_j,c_j,e_j,s_j \ge 0 | j=p,t; e_j \le 1\}} (CS_\ell^* + SDWL_\ell^*) \right] && \text{(A32)} \\ &\le \max_{\ell \in \{p,b\}} \left[ \sup_{\{c_j,e_j,s_j \ge 0 | j=p,t; e_j \le 1\}} (TS_\ell^{**} - PS_\ell^*) \right]. && \text{(A33)} \end{aligned}$$

Equation (A33) follows from substituting from the definition $TS_\ell^{**} = CS_\ell^* + PS_\ell^* + SDWL_\ell^*$. It also reflects several other simplifications. The ex post surplus terms on the right-hand side are not functions of $k_j$, so development costs can be removed from the set of parameters over which the supremum is taken. Furthermore, $TS_t^{**} = PS_t^*$ because a treatment extracts all surplus. Also $TS_n^{**} = PS_n^* = 0$ because producing nothing generates no surplus. Hence $TS_\ell^{**} - PS_\ell^* = 0$ for $\ell = t, n$, so generic strategies $\ell = t, n$ can be ignored in the maximization problem.

The proof is completed by finding a new expression for the term in square brackets in (A33), which holds for both $\ell = p$ and $\ell = b$:

$$\sup_{\{c_j,e_j,s_j \ge 0 | j=p,t; e_j \le 1\}} (TS_\ell^{**} - PS_\ell^*) = \mu_X - PS_p^o \tag{A34}$$

Intuitively, (A34) says that the same parameters that maximize the total-surplus "pie" also maximize the part left over after the producer takes its slice. The arguments needed to prove (A34) are fairly involved and thus relegated to online Appendix B.

Substituting (A34) into (A33) and dividing by $\mu_X$,

$$\sup_{\{k_j,c_j,e_j,s_j \ge 0 | j=p,t; e_j \le 1\}} \left( \frac{DWL}{\mu_X} \right) \le \frac{\mu_X - PS_p^o}{\mu_X} = 1-\rho^o. \tag{A35}$$

Combining (A25) and (A35) completes the "sandwiching" argument, showing potential deadweight loss equals $1-\rho^o$. *Q.E.D.*

**Proof of Proposition 13:** Consider a model of competition $C$ satisfying (12)–(15). Most of the proof is concerned with a preliminary analysis of the set of fixed-cost configurations

$$K_1 = \left\{ (k_{ji}) \;\middle|\; \begin{array}{l} k_{j1} \geq 0 \text{ for } j = p,t; \\ k_{ji} > \mu_X \text{ for } i = 2,\ldots,N, j = p,t \end{array} \right\}. \tag{A36}$$

$K_1$ is the set of fixed-cost vectors, with a component for each product-firm combination, such that the fixed costs for firm 1 can be any non-negative number but for the rest of the firms are higher than disease burden $\mu_X$. Suppose some element of $K_1$ is drawn for firms' fixed costs. For all $i = 2,\ldots,N$, $k_{ji} > \mu_X \geq PS^*_j = PS^*_j(1,0,0) \geq PS^*_j(n_j,n_\ell,n_b)$. The first step follows from construction of $K_1$ in (A36), the second step from the fact that consumers do not expend more than the total disease burden, the third step from (13), and the fourth step from (12). But $k_{ji} > PS^*_j(n_j,n_\ell,n_b)$ implies no firm $i > 1$ enters just one of the two product markets by (14). Similar analysis shows no firm $i > 1$ enters both markets either. Thus for any element of $K_1$, only firm 1 possibly enters any market, implying that equilibrium welfare is monopoly welfare, $W^*$.

Continue to suppose some element of $K_1$ is drawn for firms' fixed costs. In the first best in which the planner can choose prices and which firms enter which markets, no firms but 1 produce. If some firm $i > 1$ produces some product $j$, social welfare is at most $\mu_X - k_{ji}$. This is because gross social welfare cannot exceed the entire disease burden $\mu_X$, and if $i$ produces product $j$, total industry fixed costs that have to be netted out amount to at least $k_{ji}$. But by construction of $K_1$, $k_{ji} > \mu_X$ for $i > 1$, implying social welfare is negative if any firm $i > 1$ enters. Thus the planner would only choose to have firm 1 enter the market. Hence for any element of $K_1$, first-best welfare is the same as with a monopoly, $W^{**}$. Putting this result together with the result from the previous paragraph, for any element of $K_1$, deadweight loss is

$$DWL(C,N) = W^{**} - W^*. \tag{A37}$$

With the preliminary analysis of $K_1$ in hand, we can readily prove the proposition:

$$\begin{aligned} & \sup_{\{k_{ji} \geq 0 | j=p,t; i=1,\ldots,N\}} \left[ \frac{DWL(C,N)}{\mu_X} \right] \\ & \geq \sup_{K_1} \left[ \frac{DWL(C,N)}{\mu_X} \right] && \text{(A38)} \\ & = \sup_{K_1} \left( \frac{W^{**} - W^*}{\mu_X} \right) && \text{(A39)} \\ & = \sup_{\{k_{p1},k_{t1} \geq 0\}} \left( \frac{W^{**} - W^*}{\mu_X} \right). && \text{(A40)} \end{aligned}$$

Condition (A38) holds because $K_1$ is a subset of the set over which the supremum on the left-hand side is taken, (A39) holds by (A37), and (A40) holds because values of entry costs for firms $i > 1$ are irrelevant if only firm 1 enters. By Proposition 2, (A40) equals $1 - \rho^*_X$. *Q.E.D.*

**Proof of Proposition 14:** The initial equalities in the statement of the proposition—$v_0 = y$ and $v_1 = hy$—are a simple relabeling that hold without loss of generality. To see this, define $Y = V_0$ and $H = V_1/V_0$. Then $v_0 = y$ by definition, and $v_1 = (v_1/v_0)v_0 = hy$. The content of the proposition are the statements involving expectations. To prove those,

$$\begin{aligned} y &= v_0 \\ &= E(V_1 | V_0 = v_0, X = x) \\ &= E(V_1 | Y = y, X = x) \\ &= yE(H | Y = y, X = x), \end{aligned}$$

where the first step holds by definition $Y = V_0$, the second by (18), the third again by defintion $Y = V_0$, and the last by definition $H = V_1/V_0$ and by a property of conditional expectations. These equalities together imply

$$E(H | Y = y, X = x) = 1. \tag{A41}$$

Thus

$$E(H) = E(E(H | Y = y, X = x)) = E(1),$$

where the first equality holds by the law of iterated expectations and the second by (A41). This shows $E(H) = 1$. Similar arguments establish $E(H | X = x) = 1$ and $E(H | Y = y) = 1$. *Q.E.D.*

**Proof of Proposition 15:** Suppose $PS^o_p = PS^o_t$. Then $PS^o_{\text{max}} = PS^o_{\text{min}}$, implying that the right-hand side of (19) equals 0. But then the proposition holds trivially because $DWL \geq 0$.

Suppose for the remainder of the proof that $PS^o_p \neq PS^o_t$, a necessary and sufficient condition for $PS^o_{\text{max}} > PS^o_{\text{min}}$. More specifically, we will suppose $PS^o_t > PS^o_p$. Analysis of the alternative $PS^o_p > PS^o_t$ is similar and omitted for brevity. We have

$$\sup_{\{k_j,c_j,e_j,s_j \geq 0 | j=p,t; e_j \leq 1\}} (DWL) \geq \sup_{\substack{k_p \in (PS^o_p, PS^o_t) \\ k_t \in (k_p, PS^o_t)}} (DWL^o). \tag{A42}$$

Equation (A42) follows because the supremum on the right-hand side is taken over a more restrictive set than the left: parameters in $DWL^o$ are implicitly set to their "original" values $c_j = s_j = 0$ and $e_j = 1$, and the set of development costs is smaller.

Some manipulations will help us analyze the right-hand side of (A42). For $k_p \in (PS^o_p, PS^o_t)$, $\Pi^o_p = PS^o_p - k_p < 0$. For $k_t \in (k_p, PS^o_t)$, $\Pi^o_t = PS^o_t - k_t > 0$. Letting $b$ index the strategy of developing both products, a revealed-preference argument can be used to show $PS^o_b \leq PS^o_p + PS^o_t$; i.e., the firm can earn more from a preventive and treatment sold at optimized prices on replicated markets than sold together on one market. Hence $\Pi^o_b = PS^o_b - k_p - k_t \leq PS^o_p + PS^o_t - k_p - k_t = \Pi^o_p + \Pi^o_t < \Pi^o_t$, where the last step follows from $\Pi^o_p < 0$. Hence $\Pi^o_t > \max(\Pi^o_p, \Pi^o_b, 0)$, implying the firm's equilibrium strategy is to develop a treatment alone. Thus equilibrium welfare is $W^o = W^o_t = \mu_U - SDWL^o_t - k_t$. First-best welfare is $W^{oo} = \mu_U - k_p$. To see this, note that, given the original value of the parameters, either product can generate first-best surplus if offered at no charge. Whichever product has the lower development cost, in this case $k_p$, generates first-best welfare.

Using the expressions just derived, $DWL^o = W^{oo} - W^o = \mu_U - k_p - (\mu_U - SDWL^o_t - k_t) = k_t - k_p + SDWL^o_t$. Substituting into (**??**), this expression equals

$$\begin{aligned} \sup_{\substack{k_p \in (PS^o_p, PS^o_t) \\ k_t \in (k_p, PS^o_t)}} (k_t - k_p + SDWL^o_t) \quad & \geq \quad \sup_{\substack{k_p \in (PS^o_p, PS^o_t) \\ k_t \in (k_p, PS^o_t)}} (k_t - k_p) && \text{(A43)} \\ & = \quad PS^o_t - PS^o_p && \text{(A44)} \\ & = \quad PS^o_{\text{max}} - PS^o_{\text{min}}. && \text{(A45)} \end{aligned}$$

Equation (A43) follows from $SDWL^o_t \geq 0$, (A44) from substituting the upper bound on $k_t$ and the lower bound on $k_p$ in the constraint set, and (A45) from maintained assumption $PS^o_t > PS^o_p$, which implies $PS^o_{\text{max}} = PS^o_t$ and $PS^o_{\text{min}} = PS^o_p$. The proof is completed by dividing equations (**??**)–(A45) through by $\mu_U$. *Q.E.D.*

# References

Acemoglu, D. and J. Linn (2004). "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," *Quarterly Journal of Economics* 119: 1049–1090.

Anderson, S. P. and R. Renault. (2003) "Efficiency and Surplus Bounds in Cournot Competition," *Journal of Economic Theory* 113: 253–264.

Anderson, O. W., P. Collette, and J. J. Feldman. (1960) *Expenditure Patterns for Personal Health Services, 1953 and 1958: Nationwide Survey*. New York: Health Information Foundation.

Baumol, W. J., J. C. Panzar, and R. D. Willig. (1982) *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jovanovich.

Bergemann, D., B. Brooks, and S. Morris. (2014) "The Limits of Price Discrimination," Cowles Foundation working paper no. 1896RR.

Biehl, A. R. (2001) "Durable-Goods Monopoly with Stochastic Values," *Rand Journal of Economics* 32: 565–577.

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

Boulier, B. (2006) "A Shot in the Dark: Uncertainty and Vaccine Demand and Supply," George Washington University working paper.

Brito, D. L., E. Sheshinski, and M. D. Intrilligator. (1991) "Externalities and Compulsory Vaccination," *Journal of Public Economics* 45: 69–90.

Brooks, B. A. "Surveying and Selling: Belief and Surplus Extraction in Auctions," Princeton University working paper.

Budish, E., B. N. Roin, and H. Williams. (2013) "Do Fixed Patent Terms Distort Innovation? Evidence from Cancer Clinical Trials," National Bureau of Economic Research working paper no. 19430.

Bulow, J. I., J. D. Geanakoplos, and P. D. Klemperer. (1985) "Multimarket Oligopoly: Strategic Substitutes and Complements," *Journal of Political Economy* 93: 488–511.

Centers for Disease Control. (2012) "Monitoring Selected National HIV Prevention and Care Objectives by Using HIV Surveillance Data, United States and 6 Dependent Areas, 2010," *HIV Surveillance Supplemental Report* vol. 17, no. 3.

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

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

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

De Graba, P. (1995) "Buying Frenzies and Seller-Induced Excess Demand," *Rand Journal of Economics* 26: 331-342.

Dunne, E. F., *et al.* (2007) "Prevalence of HPV Infection Among Females in the United States," *Journal of the American Medical Association* 297: 813–819.

Euerle, B. and P. H. Chandrasekar. (2012) "Syphilis," in B. A. Cunha, ed., *Medscape Reference*. Accessed August 27, 2012 from emedicine.medscape.com/article/229461.

Fabinger, M. and E. G. Weyl. (2014) "A Tractable Approach to Pass-Through Patterns with Applications to International Trade," SSRN working paper, available at http://ssrn.com/abstract=2194855.

Francis, P. J. (1997) "Dynamic Epidemiology and the Market for Vaccinations," *Journal of Public Economics* 63: 383-406.

Finkelstein, A. (2004). "Static and Dynamic Effect of Health Policy: Evidence from the Vaccine Industry," *Quarterly Journal of Economics* 119: 527–564.

Gabaix, X. (2009) "Power Laws in Economics and Finance," *Annual Review of Economics* 1: 255–293.

Garber, A. M., C. I. Jones, and P. Romer. (2006) "Insurance and Incentives for Medical Innovation," *Forum for Health Economics & Policy* 9: 1–27.

GEN News Highlights. (2012) "FDA: HIV Numbers Drove Truvada Decision," July 17, article no. 81247053.

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

Gersovitz, M. (2003) "Births, Recoveries, Vaccinations, and Externalities," in R. Arnott, ed., *Economics for an Imperfect World: Essays in Honor of Joseph E. Stiglitz*, 469–483.

Gersovitz, M. and J. S. Hammer. (2004) "The Economical Control of Infectious Diseases," *Economic Journal* 114: 1–27.

Gersovitz, M. and J. S. Hammer. (2005) "Tax/Subsidy Policy Toward Vector-Borne Infectious Diseases," *Journal of Public Economics* 89: 647–674.

Getzen, T. E. (2000) "Health care is an Individual Necessity and a National Luxury: Applying Multilevel Decision Models to the Analysis of Health Care Expenditures," *Journal of Health Economics* 19: 259–270.

Harpavat, S. and S. Nissim. (2001) *MicroCards: Review Cards for Medical Students*. Philadelphia: Lippincott Williams & Wilkins.

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

Hartline, J. D. and T. Roughgarden. (2009) "Simple Versus Optimal Mechanisms," *Proceedings of the 10th ACM Conference on Electronic Commerce* 225–234.

Hernandez, B. Y., *et al.* (2008) "Transmission of Human Papillomavirus in Heterosexual Couples," *Emerging Infectious Diseases* 14: 888–894.

Howard, Robin S. (2005) "Poliomyelitis and the Postpolio Syndrome," *British Medical Journal* 330: 1314–1318.

Johnson, J. P. and D. P. Myatt. (2006) "On the Simple Economics of Advertising, Marketing, and Product Design," *American Economic Review* 96: 756–784.

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

Kessing, S. G. and R. Nuscheler. (2006) “Monopoly Pricing with Negative Network Effects: The Case of Vaccines,” *European Economic Review* 50: 1061–1069.

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

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

Kremer, M. and C. M. Snyder. (2003) “Why Are Drugs More Profitable Than Vaccines?” National Bureau of Economic Research working paper no. 9833.

Kremer, M. and C. M. Snyder. (2015) “Worst-Case Bounds on R&D and Pricing Distortions: Theory and Disturbing Conclusions if Consumer Values Follow the World Income Distribution,” mimeo, Harvard University.

Kremer, M., C. M. Snyder, and H. Williams. (2012) “Vaccines: Integrated Economic and Epidemiological Models,” mimeo, Harvard University.

Lakdawalla, D. and N. Sood. (2013) “Health Insurance as a Two-Part Pricing Contract,” *Journal of Public Economics* 102: 1–12.

Lau, Brandyn D., Brian L. Pinto, David R. Thiemann, and Christoph U. Lehmann. (2011) “Budget Impact Analysis of Conversion from Intravenous to Oral Medication When Clinically Eligible for Oral Intake,” *Clinical Therapeutics* 33: 1792–1796.

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

Liljeros, F., C. R. Edling, L. A. Nunes Amaral, H. E. Stanley, and Y. Åberg. (2001) “The Web of Human Sexual Contacts,” *Nature* 411: 907–908.

Makowski, L. and J. M. Ostroy. (1995) “Appropriation and Efficiency: A Revision of the First Theorem of Welfare Economics,” *American Economic Review* 85: 808–827.

Makowski, L. and J. M. Ostroy. (2001) “Perfect Competition and the Creativity of the Market,” *Journal of Economic Literature* 39: 479–535.

Maleug, D. A. and C. M. Snyder (2006) “Bounding the Relative Profitability of Price Discrimination,” *International Journal of Industrial Organization* 24: 995-1011.

Mandell, G. L., J. E. Bennett, and R. Dolin. (2009) *Principles and Practice of Infectious Diseases* seventh edition. Philadelphia: Elsevier Churchill Livingstone.

McNeil, D. G. (2014) “Advocating Pill, U.S. Signals Shift to Prevent AIDS,” *New York Times* May 15, A1.

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

Morbidity and Mortality Weekly Report. (various years) “Summary of Notifiable Diseases, United States,” Centers for Disease Control and Prevention, downloaded December 20, 2009 from www.cdc.gov/-mmwr/mmwr_nd/index.html

Mueller, Steffen, Eckard Wimmer, and Jeronimo Cello. (2005) "Poliovirus and Poliomyelitis: A Tale of Guts, Brains, and an Accidental Event," *Virus Research* 111: 175–193.

National Cancer Institute. (2009) "BRCA1 and BRCA2: Cancer Risk and Genetic Testing," *National Cancer Institute Fact Sheet.* Retrieved August 16, 2012, from www.cancer.gov/cancertopics/factsheet/-Risk/BRCA.

Newell, R., A. Jaffee, and R. N. Stavins. (1999) "The Induced Innovation Hypothesis and Energy-Saving Technological Change," *Quarterly Journal of Economics* 114: 907–940.

Oster, E. (2013) "Unobservable Selection and Coefficient Stability: Theory and Evidence," NBER working paper no. 19054.

Ottaviani, M. and A. Prat. (2001) "The Value of Public Information in Monopoly," *Econometrica* 69: 1673–1683.

Purcell, D. W., et al. (2012) "Estimating the Population Size of Men Who Have Sex with Men in the United States to Obtain HIV and Syphillis Rates," *Open AIDS Journal* 6: 98–107.

Rockstroh, J. K., *et al.* (1995) "Male-to-Female Transmission of HIV in a Cohort of Hemophiliacs—Frequency, Risk Factors and Effect of Sexual Counseling," *Infection* 23: 29–32.

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

Royce, R. A., *et al.* (1997) "Sexual Transmission of HIV," *New England Journal of Medicine* 336: 1072–1078.

Snyder, C. M., W. Begor, and E. R. Berndt. (2011) "Economic Perspectives on the Advance Market Commitment for Pneumococcal Vaccines," *Health Affairs* 30: 1508–1517.

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

Tirole, J. (1988) *The Theory of Industrial Organization*. Cambridge, Massachusetts: MIT Press.

U.S. Census Bureau. (2013) "Vintage 2007: National Tables: National Characteristics: National Sex, Race, and Hispanic Origin," downloaded September 16, 2013 from www.census.gov/popest/data/-historical/2000s/vintage_2007/index.html.

Weyl, E. G. and M. Fabinger. (2013) "Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition," *Journal of Political Economy* 121: 528–583.

Weyl, E. G. and J. Tirole. (2012) "Market Power Screens Willingness-to-Pay," *Quarterly Journal of Economics* 127: 1971–2003.

Wilson, P. W. F., *et al.* (1998) "Prediction of Coronary Heart Disease Using Risk Factor Categories," *Circulation* 97: 1837–1847.

# Online Appendices for “Preventives Versus Treatments”

Michael Kremer

*Department of Economics, Harvard University, Littauer Center 207, Cambridge MA 02138; email: mkremer@fas.harvard.edu.*

Christopher M. Snyder

*Department of Economics, Dartmouth College, 301 Rockefeller Hall, Hanover NH 03755; email: chris.snyder@dartmouth.edu.*

## Description

This document contains four appendices not included in the published paper. The appendix included in the published paper, Appendix A, contains proofs of propositions. Appendix B included here provides proofs in part or whole omitted from the published paper for space considerations. Appendix C provides details on the computation of $\beta$ in calibrations of the Kaplan model reported in Section of the published paper. Appendix D introduces a simple entry model showing how results on producer surplus (such as those reported in Section 3 of the published paper) can be translated into testable hypotheses regarding entry probabilities (as tested in Section 7.2 of the published paper). Appendix E provides a table of robustness checks on the linear probability model.

## Online Appendix B: Additional Proofs

This appendix contains additional proofs omitted for space considerations from Appendix A of the published paper.

**Completing the Proof of Proposition 12:** The remaining step in the proof of Proposition 12 is to establish equation (A34) for $\ell = p$ and $\ell = b$. For reference, (A34) is repeated and renumbered here:

$$\sup_{\{c_j,e_j,s_j \geq 0 | j=p,t; e_j \leq 1\}} (TS_\ell^{**} - PS_\ell^*) = \mu_X - PS_p^o \tag{B1}$$

First, consider the firm’s strategy of selling the preventive alone ($\ell = p$). A consumer with disease risk $x$ buys the preventive if his expected net benefit from the preventive, $e_p x - s_p$, exceeds the price $p_p$, or upon rearranging,

$$x \geq \frac{p_p + s_p}{e_p} = x_p, \tag{B2}$$

where $x_p$ is the marginal consumer type when a preventive alone is sold. Producer surplus is

$$PS_p^* = \max_{p_p \in [0,\infty)} \int_{(p_p+s_p)/e_p}^1 (p_p - c_p) dF_X(x) \tag{B3}$$

$$= e_p \int_{x_p^*}^1 (x_p^* - \tilde{c}_p) dF_X(x). \tag{B4}$$

The second equality holds by making the change of variables in (B2), taking $x_p^*$ to be the maximizing such value of $x_p$, and substituting $\tilde{c}_p$, where

$$\tilde{c}_j = \frac{c_j + s_j}{e_j} \tag{B5}$$

can be interpreted as the combined firm and consumer cost for product $j = p, t$, expressed per unit of expected health benefit.

The first-best policy regarding a preventive allocates it to any consumer whose expected net benefit $e_p x - s_p$ exceeds marginal cost $c_p$, or upon rearranging, $x \geq (c_p + s_p)/e_p = \tilde{c}_p$. First-best surplus from a preventive is thus

$$TS_p^{**} = \int_{\tilde{c}_p}^{1} (e_p x - s_p - c_p) dF_X(x) = e_p \int_{\tilde{c}_p}^{1} (x - \tilde{c}_p) dF_X(x). \tag{B6}$$

Combining (B4) and (B6),

$$\sup_{\{c_p, s_p, e_p \geq 0 | e_p \leq 1\}} (TS_p^{**} - PS_p^*) = \sup_{\{\tilde{c}_p, e_p \geq 0 | e_p \geq 1\}} \left\{ e_p \left[ \int_{\tilde{c}_p}^{1} (x - \tilde{c}_p) dF_X(x) - \int_{x_p^*}^{1} (x_p^* - \tilde{c}_p) dF_X(x) \right] \right\}. \tag{B7}$$

Parameters $c_p$ and $s_p$ have been replaced by $\tilde{c}_p$ in the set over which the supremum is taken because they do not appear aside from $\tilde{c}_p$. Although $\tilde{c}_p$ is a function of $e_p$ in (B5), it can be varied independently by varying the parameters $s_p$ and $c_p$ for which $\tilde{c}_p$ has been substituted, so $e_p$ and $\tilde{c}_p$ should be viewed as independent parameters in (B7). Since $x_p^*$ is the maximizer of the second integral in (B7), clearly $x_p^* \geq \tilde{c}_p$. Thus the first integral in (B7) is weakly greater than the second, implying the factor in square brackets is non-negative, implying (B7) is non-decreasing in $e_p$, implying the supremum can be generated by setting $e_p = 1$. Differentiating (B7) with respect to $\tilde{c}_p$ yields

$$-e_p \left[ \int_{\tilde{c}_p}^{1} dF_X(x) + \int_{x_p^*}^{1} dF_X(x) \right] = e_p \left[ F_X(\tilde{c}_p) - F_X(x_p^*) \right], \tag{B8}$$

which is nonpositive because $x_p^* \geq \tilde{c}_p$. The derivative in (B8) uses the Envelope Theorem, allowing one to ignore the indirect effect of $\tilde{c}_p$ on (B7) through its effect on $x_p^*$: because $x_p^*$ is a maximizer of the second integral, changes in $x_p^*$ have a second-order effect on the integral. The supremum can be generated by the lowest feasible value of $\tilde{c}_p$, i.e., $\tilde{c}_p = 0$. Substituting $\tilde{c}_p = 0$ and $e_p = 1$ into (B7) yields

$$\begin{aligned} \sup_{\{c_p, s_p, e_p \geq 0 | e_p \geq 1\}} (TS_p^{**} - PS_p^*) &= \int_0^1 x \, dF_X(x) - \max_{x \in [0,1]} [x \Phi_X(x)] && \text{(B9)} \\ &= \mu_X - PS_p^o, && \text{(B10)} \end{aligned}$$

where the last line follows from Lemma 1.

Next, examine the case in which the firm sells both a preventive and a treatment ($\ell = b$) using backward induction. Let $p_{bj}$ be the price for product $j \in \{p, t\}$ given both are sold. A consumer who ends up contracting the disease (either because he did not be the preventive or it was unsuccessful) buys the treatment if $e_t - s_t \geq p_{bt}$. The optimal treatment price is $p_{bt}^* = e_t - s_t$. Note this price leaves the consumer with no surplus. Folding the game back, in deciding whether or not to buy the preventive, the presence of the treatment is thus irrelevant for the consumer. A consumer with disease risk $x$ buys the preventive if the net benefit $e_p x - s_p$ exceeds the price $p_{bp}$, or upon rearranging,

$$x \geq \frac{p_{bp} + s_p}{e_p} = x_{bp}. \tag{B11}$$

The producer surplus from selling both, $PS_b^*$, equals

$$\max_{p_{pb}\in[0,\infty)} \left\{ \int_{x_{bp}}^{1} \left[p_{pb} - c_p + (1-e_p)x(p_{bt}^* - c_t)\right] dF_X(x) + \int_0^{x_{bp}} x(p_{bt}^* - c_t) dF_X(x) \right\}, \tag{B12}$$

where $x_{bp}$ is regarded as the function of $p_{bp}$ given by (B11). The first integral incorporates the profit from consumer types who buy the preventive. Markup $p_{pb} - c_p$ is earned from each. With probability $(1-e_p)x$, the preventive does not work for him and he buys the treatment as well. The second integral incorporates the profit from consumer types who do not buy the preventive but end up contracting the disease and buy the treatment. Applying the change of variables in (B11) to (B12) and taking $x_{bp}^*$ to be the value of $x_{bp}$ maximizing the expression shows $PS_b^*$ equals

$$\begin{aligned} & \int_{x_{pb}^*}^{1} [e_p x_{bp}^* - s_p - c_p + (1-e_p)x(p_{bt}^* - c_t)] dF_X(x) + \int_0^{x_{bp}^*} x(p_{bt}^* - c_t) dF_X(x) && \text{(B13)} \\ = \;& e_t(1-\tilde{c}_t)\mu_X + e_p \int_{x_{bp}^*}^{1} [x_{bp}^* - \tilde{c}_p - x e_t(1-\tilde{c}_t)] dF_X(x). && \text{(B14)} \end{aligned}$$

The second line follows from substituting $\tilde{c}_t = (c_t + s_t)/e_t$, substituting $p_{bt}^* = e_t - s_t$, and rearranging.

In the first-best policy involving both products, the preventive and treatment are sold at cost, $c_p$ and $c_t$, respectively. First-best surplus $TS_b^{**}$ thus equals

$$\begin{aligned} & \int_{\tilde{c}_p}^{1} [e_p x - s_p - c_p + (1-e_p)x(e_t - s_t - c_t)] dF_X(x) + \int_0^{\tilde{c}_p} x(e_t - s_t - c_t) dF_X(x) && \text{(B15)} \\ = \;& e_t(1-\tilde{c}_t)\mu_X + e_p \int_{\tilde{c}_p}^{1} [x - \tilde{c}_p - x e_t(1-\tilde{c}_t)] dF_X(x), && \text{(B16)} \end{aligned}$$

where (B16) follows from similar computations behind (B14). Subtracting (B14) from (B16),

$$\begin{aligned} & \sup_{\{c_j, s_j, e_j \geq 0 | j=p,t; e_j \leq 1\}} (TS_b^{**} - PS_b^*) \\ = \;& \sup_{\{\tilde{c}_j, e_j \geq 0 | j=p,t; e_j \leq 1\}} \left\{ e_p \left\{ \int_{\tilde{c}_p}^{1} [x - \tilde{c}_p - x e_t(1-\tilde{c}_t)] dF_X(x) - \int_{x_p^*}^{1} (x_p^* - \tilde{c}_p) dF_X(x) \right\} \right\}. && \text{(B17)} \end{aligned}$$

Since $x_{bp}^*$ is the maximizer of the integral in (B12), the firm would never chose $x_{bp}^*$ such that the integrand is negative for any $x$. Therefore the integrand is non-negative for $x = x_p^*$, implying $x_{bp}^* - \tilde{c}_p - x_{bp}^* e_t(1-\tilde{c}_t) \geq 0$, in turn implying $x_{bp} \geq \tilde{c}_p$. But $x_{bp} \geq \tilde{c}_p$ implies that the difference between the integrals in (B17) is non-negative, implying that the supremum in (B17) can be generated by setting $e_p = 1$. The derivative of (B17) with respect to $\tilde{c}_t$ is

$$\int_{\tilde{c}_p}^{1} x e_t dF_X(x) - \int_{x_p^*}^{1} x e_t dF_X(x), \tag{B18}$$

which is positive because $x_{bp} \geq \tilde{c}_p$. As in (B8), the derivative in (B18) uses the Envelope Theorem. We have shown that the supremum in (B17) can be generated by $e_p = 1$ and $c_t = 1$. Substituting these values into (B17) implies (B17) equals

$$\begin{aligned} \sup_{\{c_j, s_j, e_j \geq 0 | j=p,t; e_j \leq 1\}} (TS_b^{**} - PS_b^*) &= \sup_{\{\tilde{c}_p \geq 0\}} \left[ \int_{\tilde{c}_p}^{1} (x - \tilde{c}_p) dF_X(x) - \int_{x_p^*}^{1} (x_p^* - \tilde{c}_p) dF_X(x) \right] && \text{(B19)} \\ &= \mu_X - PS_p^o, && \text{(B20)} \end{aligned}$$

where (B20) follows from the same arguments used to derive (B10) from (B7). *Q.E.D.*

**Proof of Proposition 16:** Assume benchmark values of the parameters $c_j, s_j = 0$, $e_j = 1$, $j = p, t$. Suppose $Y$ is independent of $X$. Then

$$
\begin{aligned}
PS_p^o &= \max_{p_p \in [0,\infty)} \left\{ \int_{p_p/\bar{y}}^{1} \left[ \int_{p_p/x}^{\bar{y}} p_p \, dF_Y(y) \right] dF_X(x) \right\} && \text{(B21)} \\
&\leq \max_{p_p \in [0,\infty)} \left\{ \int_{p_p/\bar{y}}^{1} \max \left[ 0, \int_{p_p/x}^{\bar{y}} p_p \, dF_Y(y) \right] dF_X(x) \right\} && \text{(B22)} \\
&\leq \max_{p_p \in [0,\infty)} \left\{ \int_{0}^{1} \max \left[ 0, \int_{p_p/x}^{\bar{y}} p_p \, dF_Y(y) \right] dF_X(x) \right\} && \text{(B23)} \\
&\leq \int_{0}^{1} \left\{ \max_{p_p \in [0,\infty)} \left\{ \max \left[ 0, \int_{p_p/x}^{\bar{y}} p_p \, dF_Y(y) \right] \right\} \right\} dF_X(x) && \text{(B24)} \\
&= \int_{0}^{1} \left\{ \max_{p_p \in [0,\infty)} \left[ \int_{p_p/x}^{\bar{y}} p_p \, dF_Y(y) \right] \right\} dF_X(x) && \text{(B25)} \\
&= \int_{0}^{1} \left\{ \max_{p' \in [0,\infty)} \left[ \int_{p'}^{\bar{y}} p' x \, dF_Y(y) \right] \right\} dF_X(x) && \text{(B26)} \\
&= \mu_X \max_{p' \in [0,\infty)} \left[ p' \Phi_Y(p') \right] && \text{(B27)} \\
&= \mu_X REC_Y && \text{(B28)} \\
&= PS_t^o. && \text{(B29)}
\end{aligned}
$$

Equations (B21) and (B27) hold by applying the independence condition to the formulae (20). The rest of the steps up to the last are algebraic manipulations. The last step follows from (21). The inequality in (B24) is strict if there is nontrivial heterogeneity in the distribution of positive risks $X$. *Q.E.D.*

**Proof of Proposition 17:** Assume benchmark values of the parameters $c_j, s_j = 0$, $e_j = 1$, $j = p, t$. Let $PS_p^o$ and $PS_t^o$ be equilibrium producer surpluses in the model with no heterogeneity in $Y$ and $PSY_p^o$ and $PSY_t^o$ be equilibrium producer surpluses when $Y$ distributed independently from $X$ is added to the model. Given $h = 1$,

$$
\begin{aligned}
PSY_p^o &= p_U^o \Phi_U(p_U^o) && \text{(B30)} \\
&\geq p_X^o p_Y^o \Phi_U(p_X^o p_Y^o) && \text{(B31)} \\
&\geq p_X^o p_Y^o \Pr(x \geq p_X^o) \Pr(y \geq p_Y^o) && \text{(B32)} \\
&= PS_p^o PSY_t^o / PS_t^o, && \text{(B33)}
\end{aligned}
$$

where $p_\Theta^o = \text{argmax}_p [p \Phi_\Theta(p)]$ for $\Theta \in \{X, Y, U\}$. Equation (B30) follows from equation (20). Condition (B31) follows because $p_U^o$, as an argmax, produces a higher value for $p\Phi_U(p)$ than the product $p_X^o p_Y^o$. Condition (B32) follows from

$$
\begin{aligned}
\Phi_U(p_X^o p_Y^o) &= \Pr(u \geq p_X^o p_Y^o) \\
&= \Pr(xy \geq p_X^o p_Y^o) \\
&\geq \Pr(x \geq p_X^o) \Pr(y \geq p_Y^o),
\end{aligned}
$$

where the last step holds because $x \geq p_X^o$ and $y \geq p_Y^o$ implies $xy \geq p_X^o p_Y^o$. To see (B33), note first that the proof of Lemma 2 implies $PS_p^o = p_X^o \Pr(X \geq p_X^o)$. Note second $PS_t^o = \mu_X$ by equation (3). Note third $PSY_t^o = \mu_X p_Y^o \Pr(y \geq p_Y^o) = \mu_X p_Y^o \Pr(y \geq p_Y^o)$ applying the independence assumption to equation (21). Conditions (B30)–(B33) together imply $PS_p^o / PS_t^o \leq PSY_p^o / PSY_t^o$. If $X$ and $Y$ are continuous, then the inequality in (B32) is strict. *Q.E.D.*

**Proof of Proposition 18:** Suppose $Y = g(X)$, where $g$ is some increasing function. Let $p_p^*$ be the optimal preventive price. Preventive demand equals $\Phi_U(p_p^*) = \Phi_Y(\hat{y})$ for $\hat{y}$ given by the solution to $g^{-1}(\hat{y})\hat{y} = p_p^*$. Hence

$$PS_p^* = p_p^* \Phi_Y(\hat{y}) = g^{-1}(\hat{y})\hat{y}\Phi_Y(\hat{y}). \tag{B34}$$

Turning to producer surplus from a treatment,

$$\begin{aligned} PS_t^* &\geq \hat{y} \int_{\hat{y}}^{\bar{y}} g^{-1}(y_i)\, dF_Y(y_i) && \text{(B35)} \\ &\geq \hat{y} \int_{\hat{y}}^{\bar{y}} g^{-1}(\hat{y})\, dF_Y(y_i) && \text{(B36)} \\ &= g^{-1}(\hat{y})\hat{y}\Phi_Y(\hat{y}) && \text{(B37)} \\ &= PS_p^*. && \text{(B38)} \end{aligned}$$

Equation (B35) holds because producer surplus from a treatment $PS_t^*$ at the maximizing price at least weakly exceeds producer surplus on the right-hand side from selling a treatment at price $\hat{y}_i$. To see that the right-hand side of (B35) is the correct expression for this producer surplus, note that all types $y_i > \hat{y}$ buy the drug if they contract the disease. Each contracts the disease with probability $x_i = g^{-1}(y_i)$. Integrating over types gives the expression for producer surplus. Equation (B36) 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 (B37) is a straightforward calculation. Equation (B38) follows from (B34). The inequality in (B36) is strict if there is nontrivial heterogeneity in $X$ for preventive consumers. *Q.E.D.*

**Proof of Proposition 19:** Assume benchmark values of the parameters $c_j, s_j = 0$, $e_j = 1$, $j = p, t$. Suppose that $Y = \mu_Y$ and that $X$ and $H$ are independent. First compute the producer surplus from a preventive. The consumer buys a preventive if $p_p$ is less than the expected benefit from his or her ex ante perspective, $E(XYH|X = x, Y = \mu_Y) = x\mu_Y E(H|X = x) = x\mu_Y$, where the second equality holds by Proposition 14. Thus, preventive demand is $\Phi_X(p_p/\mu_Y)$. Producer surplus from a preventive is $p_p\Phi_X(p_p/\mu_Y) = \mu_Y x\Phi_X(x)$, making the change of variables $x = p_p/\mu_Y$. In equilibrium,

$$PS_p^o = \mu_Y REC_X. \tag{B39}$$

Next compute the producer surplus from a treatment. The consumer buys a treatment conditional on contracting the disease if $p_t$ is less than his or her ex post benefit $yh = \mu_Y h$. Because $X$ and $H$ are independent, we can multiply probabilities to obtain the individual consumer's treatment demand: $x\Phi_H(p_t/\mu_Y)$. Market demand for a treatment is thus $\int_0^1 x\Phi_H(p_t/\mu_Y)dx = \mu_X\Phi_H(p_t/\mu_Y)$, implying that producer surplus is $\mu_X p_t \Phi_H(p_t/\mu_Y) = \mu_X\mu_Y h\Phi_H(h)$, making the change of variables $h = p_t/\mu_Y$. In equilibrium,

$$PS_t^o = \mu_X\mu_Y REC_H. \tag{B40}$$

Thus $PS_t^o > PS_p^o$ if and only if (B40) strictly exceeds (B39), which in turn holds if and only if $REC_H/\mu_H > REC_X/\mu_X$.

The last step is to use decomposition results from Section 3.2 to find an equivalent expression for the ratios $REC_H/\mu_H$ and $REC_X/\mu_X$. The decomposition results apply to random variables with support on the unit interval. This is the reason for introducing the rescaled variable $\tilde{H} = H/h^{\max}$. We need to verify

$REC_H/\mu_H = REC_{\tilde{H}}/\mu_{\tilde{H}}$. Obviously $\mu_H = h^{\max}\mu_{\tilde{H}}$. A series of steps shows $REC_H = h^{\max} REC_{\tilde{H}}$:

$$\begin{aligned}
REC_H &= \max_{h\geq 0}[h\Phi_H(h)] \\
&= h^{\max}\max_{h\geq 0}[(h/h^{\max})\Phi_H(h)] \\
&= h^{\max}\max_{h\geq 0}[(h/h^{\max})\Phi_{\tilde{H}}(h/h^{\max})] \\
&= h^{\max}\max_{\tilde{h}\in[0,1]}[\tilde{h}\Phi_{\tilde{H}}(\tilde{h})] \\
&= h^{\max} REC_{\tilde{H}},
\end{aligned}$$

where the fourth step holds by the change of variables $\tilde{h} = h/h^{\max}$. Hence

$$\frac{REC_{\tilde{H}}}{\mu_{\tilde{H}}} = \frac{REC_H/h^{\max}}{\mu_H/h^{\max}} = \frac{REC_H}{\mu_H}.$$

We are set to apply the decomposition results from Section 3.2. Substituting the definition $\rho^*_X = REC_X/\mu_X$ into the decomposition formula in (10) gives $REC_X/\mu_X = 1 - Z_X[1 - \underline{\rho}(\mu_X)]$. Similarly, $REC_{\tilde{H}}/\mu_{\tilde{H}} = 1 - Z_{\tilde{H}}[1 - \underline{\rho}(\mu_{\tilde{H}})]$. Substituting, we have that $\underline{\rho}(\mu_{\tilde{H}})/Z_{\tilde{H}} > \underline{REC}(\mu_X)/Z_X > REC_X/\mu_X$ if and only if $Z_X[1 - \underline{\rho}(\mu_X)] > Z_{\tilde{H}}[1 - \underline{\rho}(\mu_{\tilde{H}})]$, which is thus a necessary and sufficient condition for $PS^o_t > PS^o_p$. The reverse inequalities are proved similarly. *Q.E.D.*

## Online Appendix C: Computing $\beta$ in Calibrations of the Kaplan Model

This online appendix provides details on the derivation of $\beta$ used in the calibrations based on the Kaplan (1990) model.

**HIV2 Calibration:** Calibration HIV2 assumes a homogeneous $\beta$ parameter across the U.S. population, computed by multiplying an estimate of the per-partner transmission rate by an estimate of the HIV prevalence rate. We take the per-partner transmission rate to be 10% from Rockstroh *et al.* (1995). The overall prevalence rate is computed from subgroup estimates in Purcell *et al.* (2012), a meta-analysis of seven surveys covering the 37 states with confidential, name-based HIV infection reporting. The following table reproduces estimates from their Table 5 of the number of HIV cases [in column (1)] and rates [in column (2)] by year end 2007 for the three mutually exclusive subgroups in the row headings.

**Table C1:** Computation of overall U.S HIV prevalence rate

| | Cases | HIV Rates<br>(per 100,000) | Population<br>(100,000) |
|---|---|---|---|
| | (1) | (2) | $(3) = (1) \div (2)$ |
| Women | 153,814 | 173 | 889.1 |
| MSM | 265,330 | 7,929 | 33.5 |
| Other men | 152,468 | 187 | 815.3 |
| Combined | 571,612 | 329 | 1,737.9 |

Population in column (3) is derived from columns (1) and (2). Summing down columns (1) and (3) given total number of cases and population, which can be divided to give the overall prevalence rate, 329 per 100,000, rounding to 0.033% per individual. Multiplying by the 10% per-partner transmission rate gives the 0.0033% value of $\beta$ used in the calibration. □

**HIV3 Calibration:** Calibration HIV3 allows $\beta$ to vary by sexual orientation, race,and gender. The homogeneous population parameter $\beta = 0.033\%$ is transformed into ones varying by subgroup in two stages. In the first stage we scale $\beta$ by the following subgroup factors, reflecting the HIV prevalence for each subgroup relative to that in the overall population.

**Table C2:** HIV prevalence relative to overall U.S. rates

| | MSM | Women and other men |
|---|---|---|
| Race | (1) | (2) |
| White | 16.792 | 0.141 |
| Hispanic | 32.231 | 0.814 |
| Black | 63.902 | 2.466 |

Column (1) can be computed fairly directly from Purcell *et al.* (2012), specifically, dividing the HIV rates per 100,000 for MSM by race reported in their Table 5 by the population rate of 329 from Table C1.

Column (2) requires more calculation. We obtain 2007 populations by race from the U.S. Census Bureau (2013) and 2007 HIV cases from Centers for Disease Control (2012), Table 5b. The number of HIV cases among women in Purcell *et al.* is half that in Centers for Disease Control (2012) and is also approximately half for other subgroups. Assume, therefore, that the 37 states covered in Purcell *et al.* account for half the U.S. population and, furthermore, that they are representative of this larger population. Then the statistics for MSM males by race from Purcell *et al.* can be combined with the overall statistics by race from Centers for Disease Control (2012) to back out statistics for non-MSM individuals by race, reported in column (2).

In the second stage, for all MSM categories, after scaling by the factor in column (1), we further scale by a factor of three to compute the associated $\beta$ 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.

Some examples will help clarify the risk calculations. Consider a white woman with $n = 15$ lifetime sexual partners. Her estimated HIV risk is

$$1-(1-0.000047)^{15} = 0.07\%,$$

where the value 0.000047 for her $\beta$ comes from multiplying the population $\beta$ by 0.141 from column (2) of Table C2. Consider an Hispanic man reporting two male and three female partners. His estimated HIV risk is

$$1-(1-0.031909)^2(1-0.000269)^3 = 6.36\%,$$

where 0.031909 is the $\beta$ associated with his male partners, computed by multiplying the population $\beta$ by 32.231 from Table C2 and again by the male-male transmission factor of 3, and 0.000269 is the $\beta$ associated with his female partners, computed by multiplying the population $\beta$ by 0.814 from Table C2. □

**HPV1 Calibration:** The $\beta$ for the HPV1 calibration is computed by multiplying the transmission rate times the prevalence rate. Data from Hernandez *et al.* (2008) 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. 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%. We take this as the HPV prevalence rate. 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). Our estimate of $\beta$ is $88.8\% \times 15.2\% = 13.5\%$. □

**Additional References:** References to all cited work are provided in the published paper except the following.

Ault, Kevin A. (2007) "Human Papillomavirus Vaccines and the Potential for Cross-protection Between Related HPV Types," *Gynecologic Oncology* 107: S31–S33.

## Online Appendix D: Empirical Entry Model

To connect the results on producer surplus from Section 3 to the probability of product development studied in our regressions, in this subsection we introduce a simple entry game played by a preventive and treatment producer. The game nests the case in which only one product is viable, in which case the relevant decision would be that producer's monopoly entry decision. To keep the game, which involves incomplete information, as simple as possible otherwise, we will consider a static version.

**Setup:** Each producer receives a draw of a development cost $k_{jm} \geq 0$, where $j = p, t$ indexes products and $m$ indexes the particular disease market under consideration. Assume $k_{jm}$ is private information for the producer of $j$, a continuous random variable with cumulative distribution function $F_K(k_{jm}, \omega_{jm})$, where $\omega_{jm}$ are factors that shift the development technology for product $j$ in market $m$. If producer $j$ enters, its expected profit net of the development cost is

$$E(\Pi^*_{jm}) = PS^*_{jm}(1 - \eta_{-j,m}\gamma_j) - k_{jm}, \tag{D1}$$

where the expectation operator is taken over realizations of the development cost of $j$'s rival. Producer $j$'s surplus $PS^*_{jm}$ is scaled down by $\gamma_j$, measuring losses due to rival's business stealing, if $j$'s rival enters, which happens with probability $\eta_{-j,m}$. (If producer $j$ is a monopolist with no rival entry threat, then we can set either $\eta_{-j,m}$ or $\gamma_j$ to 0.)

**Equilibrium:** Producer $j$ enters if (D1) is non-negative, or rearranging, $k_{jm} \leq PS^*_{jm}(1 - \eta_{-j,m}\gamma_j)$. The probability of this event is

$$\eta_{jm} = F_K(r^*_{jm}(1 - \eta_{-j,m}\gamma_j), \omega_{jm}). \tag{D2}$$

Simultaneous solution of the set of equations (D2) for $j = p, t$ gives the equilibrium entry probabilities $\eta^*_{jm}$. Because these are the variables of interest, for our purposes the game of entry under incomplete information can be reduced to a static game of complete information in which firms choose entry probabilities $\eta_{jm}$ ex ante.

**Comparative Statics:** Comparative statics are straightforward to derive viewing the right-hand side of (D2) as a standard best-response function: i.e., $\eta_{jm} = BR_{jm}(\eta_{-j,m})$, where

$$BR_{jm}(\eta_{-j,m}) = F_K(r^*_{jm}(1 - \eta_{-j,m}\gamma_j), \omega_{jm}).$$

Figure D1 provides a schematic diagram of the best responses. In the absence of heterogeneity, best responses are given by the solid curves. Equilibrium entry probabilities are given by the intersection at point $A$. To verify that the best-response functions are downward sloping as drawn, note

$$BR'_{jm}(\eta_{-j,m}) = -\gamma_j f_K(r^*_{jm}(1 - \eta_{-j,m}\gamma_j), \omega_{jm}) \leq 0,$$

where $f_K(k_{jm}, \omega_{jm})$ is the density function associated with $k_{jm}$. Hence entry probabilities are strategic substitutes in the sense of Bulow, Geanakoplos, and Klemperer (1985).

The introduction of heterogeneity corresponding to an increase in $IZ_m$ from 0 to 1 results in a shift in the best responses. The shift is mediated through the term $PS^*_{jm}$ in (D2). By definition,

$$PS^*_{jm} = \begin{cases} PS^*_{tm} & j = t \\ \rho^*_m PS^*_{tm} & j = p, \end{cases} \tag{D3}$$

where $\rho^*_m$ is the producer-surplus ratio in the market under consideration, $m$. Proposition 3 implies $\rho^*_m = 1$ if $IZ_m = 0$ and $\rho^*_m < 1$ if $IZ_m = 1$. Hence an increase in $IZ_m$ from 0 to 1 reduces $\rho^*_m$, in turn reducing $PS^*_{pm}$,

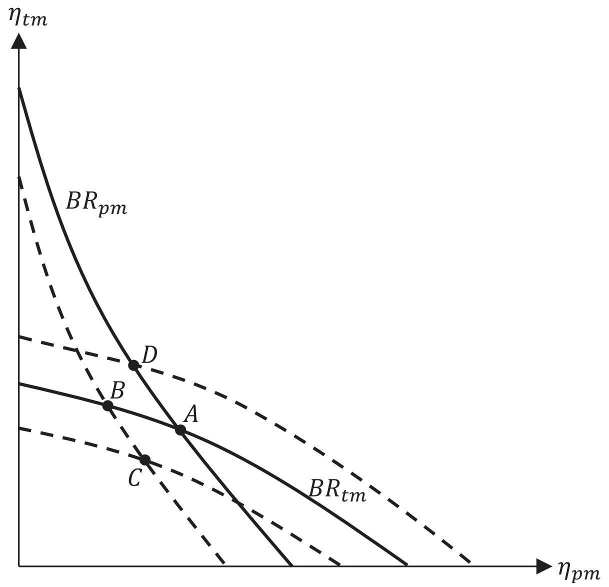

**Figure D1:** Best-response functions yielding equilibrium entry probabilities. (Market subscript *m* suppressed for brevity. Best responses after $IZ_m$ increased from 0 to 1, reflecting the introduction of disease-risk heterogeneity, given by dashed curves.)

in turn shifting the best response for the preventive $BR_{pm}$ down from the solid to the corresponding dashed curve. The best response for the treatment $BR_{tm}$ does not shift because $PS^*_{tm}$ is independent of $IZ_m$. Thus an increase in $IZ_m$ moves the equilibrium from point *A* to *B*, resulting in a decline in the probability that the preventive producer enters and, if anything, an increase in the probability that the treatment producer enters. Producer surplus from a monopoly treatment does not shift, but the decrease in the probability of entry by the preventive, which may steal some business from the treatment, makes entry indirectly more attractive for the treatment producer.

**Threats to Identification:** One threat to identification of the bias against preventives arises if $IZ_m$ is correlated with unobservable factors related to demand for all products in market *m*. For a stark example, suppose that a higher value of $IZ_m$ is associated with lower demand in market *m*, reducing $PS^*_{tm}$, but has no effect on the relative bias against preventives, so $\rho^*_m = 1$ independent of $IZ_m$. Increasing $IZ_m$ from 0 to 1 would shift both best responses, moving the equilibrium from *A* to *C* in Figure D1. One would not want to attribute the associated reduction in the probability of preventive development to a bias against preventives in this case. While we will be interested how an increase in $IZ_m$ affects the probability of preventive development $\eta^*_{pm}$, these considerations will lead us to focus more on difference in probabilities $\eta^*_{pm} - \eta^*_{tm}$. The difference-in-differences result will help identify the relative bias against preventives by effectively purging market fixed effects. The identification strategy is not iron-clad because the distribution functions in (D2) are not necessarily the same across products nor linear, so a proportional change in their arguments may not translate into an equal effect on resulting probabilities. One can simply assume that the density functions do not differ much in the relevant ranges for the two products. We will go further in the empirical analysis by controlling for technological factors $\omega_{jt}$ that might affect the distribution of development costs separately for each product.

**Endogenizing Business Stealing:** For simplicity, we have assumed $\gamma_j$ is exogenous, independent of $IZ_m$. More generally, the proportion of business one product steals from another could depend endogenously on heterogeneity in disease risk. Although we will not develop the full-blown model of competition necessary to endogenize $\gamma$ in the general case here, we will argue that endogenizing $\gamma_j$ in the benchmark model (perfectly effective products, possible heterogeneity only in disease risk) reinforces the comparative-static effects derived from Figure D1. In the benchmark model, entry by a preventive would steal all of the market from a treatment with no heterogeneity but would leave some of the market for the treatment with some heterogeneity in disease risk. Thus changing $IZ_m$ from 0 to 1 would reduce $\gamma_t$. On the other hand, entry by a treatment would have no effect on producer surplus from a preventive because the treatment's price extracts consumers' entire ex post surplus, so treatment entry would not reduce their ex ante demand for the preventive. Hence changing $IZ_m$ from 0 to 1 would not affect $\gamma_p$. Reflecting these considerations in Figure D1, the shift in $BR_{pm}$ would be the same whether or not $\gamma$ is endogenized, in either event shifting in from the solid to the dotted curve. $BR_{tm}$, which did not shift when $\gamma$ was taken to be exogenous would shift out from the solid to the dotted curve further from the origin. The equilibrium shifts from $A$ to $D$, yielding qualitatively the same comparative-statics conclusions as when the equilibrium shifted from $A$ to $B$ when $\gamma$ was taken to be exogenous.

## Online Appendix E: Additional Robustness Checks for Linear Probability Model

The following table reports alternative specifications for the spare specification reported in column (1) of Table 5 of the published paper. The first row repeats the baseline specification from the published table for comparison. The next two rows run a probit or logit instead of linear probability models. The remaining rows return to linear probability models but are run on restricted samples.

**Table E1:** Alternatives to spare specification in Table 5

| | $IZ_m$ coefficient from spare specification (1) | | | | |
|---|---|---|---|---|---|
| Specification | Vaccine developed (1a) | Drug developed (1b) | Difference (1c) = (1a) − (1b) | Observations in (1a), (1b) | $R^2$ or pseudo $R^2$ in (1c) |
| Baseline linear probability model | $-0.409^{***}$ | $-0.050$ | $-0.358^{*}$ | 58 | 0.232 |
| | (0.115) | (0.116) | (0.184) | | |
| Probit (marginal effect) | $-0.381^{***}$ | $-0.050$ | $-0.322^{*}$ | 58 | 0.180 |
| | (0.087) | (0.113) | (0.178) | | |
| Logit (marginal effect) | $-0.380^{***}$ | $-0.050$ | $-0.317^{*}$ | 58 | 0.180 |
| | (0.086) | (0.113) | (0.180) | | |
| Excluding STIs | $-0.396^{***}$ | $-0.115$ | $-0.281$ | 50 | 0.185 |
| | (0.127) | (0.136) | (0.213) | | |
| Excluding diseases spread by animal contact | $-0.491^{***}$ | 0.004 | $-0.496^{***}$ | 46 | 0.248 |
| | (0.114) | (0.135) | (0.183) | | |
| Excluding diseases affecting concentrated populations | $-0.396^{***}$ | $-0.073$ | $-0.323^{***}$ | 56 | 0.211 |
| | (0.118) | (0.120) | (0.190) | | |
| Excluding diseases with restricted ecological habitats | $-0.381^{***}$ | $-0.009$ | $-0.372^{*}$ | 54 | 0.215 |
| | (0.123) | (0.118) | (0.195) | | |
| Excluding parasitic or fungal diseases | $-0.500^{***}$ | $-0.032$ | $-0.468^{**}$ | 51 | 0.217 |
| | (0.121) | (0.128) | (0.195) | | |
| Excluding bacterial diseases | $-0.321$ | $-0.083$ | $-0.238$ | 26 | 0.057 |
| | (0.023) | (0.203) | (0.324) | | |

Notes: See notes to Table 5.