Preventives Versus Treatments Redux: Tighter Bounds on Distortions in Innovation Incentives with an Application to the Global Demand for HIV Pharmaceuticals

Michael Kremer

Harvard University
Center for Global Development
National Bureau of Economic Research

Christopher M. Snyder

Dartmouth College
National Bureau of Economic Research

September 2017

Abstract: Our previous work (Kremer and Snyder, 2015) argued that a pharmaceutical manufacturer, faced with the choice of developing a preventive or treatment, may be biased toward the treatment because the shape of its demand curve is more conducive to revenue extraction. The present paper provides new, tighter theoretical bounds on the potential deadweight loss from this bias. In an empirical application, we calibrate the global demand for HIV pharmaceuticals with country-level data on disease prevalence and income. Using the application, we demonstrate the dramatically sharper analysis achievable with the new bounds. We also perform policy counterfactuals, assessing welfare gains and losses from government interventions such as a subsidy, reference pricing, and price-discrimination ban.

JEL codes: O31 (Innovation and Invention: Processes and Incentives), L65 (Industry Studies: Manufacturing: Drugs), F23 (International Business), I14 (Health and Inequality)

Keywords: pharmaceuticals, deadweight loss, product development

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

Acknowledgments: The authors are grateful to the editor, Victor Trembley, and anonymous reviewers for suggestions that substantially improved the paper; to Eric Edmonds, Taryn Dinkelman, Margaret Kyle, Paul Novosad, Maxim Pinkovskiy, Doug Staiger, and Heidi Williams for insightful comments; and to Matthew Goodkin-Gold and Henry Senkfor for excellent research assistance. Henry Senkfor's assistantship was funded by Dartmouth College's Presidential Scholars program, which the authors gratefully acknowledge.

1. Introduction

Our previous work (Kremer and Snyder, 2015) argued that a drug can be much more lucrative than a vaccine, even when they target the same disease and thus have the same health benefit, even in the absence of epidemiological externalities, which could lead a vaccine to eradicate its own market. A preventive is sold before consumers contract the disease. At that point they may differ considerably in disease risk, preventing the firm from easily extracting much of their surplus with a uniform price. On the other hand, a treatment is sold to consumers after they have contracted the disease, at which point they no longer differ in disease risk, allowing the manufacturer to extract considerable surplus from these relatively homogeneous consumers. The treatment may thus end up being more lucrative than the preventive and the manufacturer biased toward developing it even if the preventive is substantially more effective or lower cost. We calibrated the risk distribution for human immunodeficiency virus (HIV) and found it to be close to Zipf. A Zipf distribution is the special case of a power-law distribution with an exponent equal to 1, which in the disease context intuitively means that each doubling of disease risk cuts the number of consumers with at least that risk in half, leading to an iso-revenue property.) In view of our theoretical result that the Zipf risk distribution generates the worst bias against preventives owing to its iso-revenue property, our previous work provided one explanation for why a variety of HIV drugs have been developed but as yet no vaccine.

The present paper contributes along both empirical and theoretical dimensions. The main contribution is empirical. We expand the calibrations of demand for HIV pharmaceuticals beyond the U.S. market, calibrating a global demand curve using country-level data on disease prevalence, factoring in the joint distribution of income using data on per capita gross domestic product (GDP), for robustness considering a range of values of the income elasticity of healthcare expenditures. To focus on the new surplus-extraction issues and away from epidemiological externalities well-studied in other papers,[1] the calibrations shut down any epidemiological externalities by considering pharmaceuticals that relieve symptoms without slowing person-to-person transmission.

The resulting calibrated demands are employed in a variety of exercises. One exercise leverages

[1] See Brito, Sheshinski, and Intrilligator (1991); Francis (1997); Geoffard and Philipson (1997); Gersovitz (2003); Gersovitz and Hammer (2004, 2005); Chen and Toxvaerd (2014); as well as our own work (Kremer, Snyder, and Williams 2012).

Kremer and Snyder's (2015) formulas for worst-case bounds on deadweight loss from pharmaceutical sales on the private market. Since the greatest conceivable distortions turn out to occur at the extensive margin (whether a product is developed at all) rather than the intensive margin (what price is set for an existing product), the formulas hinge on the percentage of the surplus that would be generated by completely relieving the disease burden that the producer can extract for itself (we refer to this as the producer-surplus ratio, denoted ρ). The producer-surplus ratio in turn depends on how the demand curve is shaped. Using the calibrated demands, we can compute ρ for a monopoly producer of either a vaccine or drug. In our baseline calibration, we obtain a producer-surplus ratio of $\rho_v^0 = 44\%$ for an HIV vaccine,[2] close to the estimate of $\rho_v^0 = 41\%$ from Kremer and Snyder (2015)—perhaps surprisingly close given the calibration is based on global rather than U.S. demand as in the previous paper. Considering the market for an HIV vaccine in isolation, this leads to a worst-case bound on deadweight loss of $100-44 = 56\%$. Our baseline estimate of the producer-surplus ratio for a drug is $\rho_d^0 = 38\%$, leading to a worst-case bound (considering the drug market in isolation) of $100-38 = 62\%$.

Rather than consider the vaccine and drug markets in isolation, it is natural to wonder what the worst-case deadweight loss is in the comprehensive case in which either or both products can be produced. The result relevant to this comprehensive case in Kremer and Snyder (2015) is Proposition 15. We have been able to tighten the bounds, reported in Proposition 1. For canonical values of model parameters, these tighter bounds lead to an exact expression for worst-case deadweight loss as reported in Proposition 2. The exact expression simply equals the larger of the two worst-case bounds for the isolated products; i.e., using the numbers from the calibration, $\max(56, 62) = 62\%$. The previous result merely told us that worst-case deadweight in the comprehensive case is no less than the difference between the worst cases for the individual products, i.e., no less than $62-56 = 6\%$. The new bound thus represents a dramatic sharpening of the analysis in this application: instead of saying that worst-case deadweight loss lies somewhere in the interval $[6\%, 100\%]$, we can now point-identify it at 62%.

Since the calibrations require assumptions about particular parameter values, we gauge the robustness of the results by providing calibrations for a range of values of these parameters. The

[2] The subscript on ρ_v^0 indicates the product, and the superscript indicates that it is an equilibrium value evaluated at benchmark parameter values, to be discussed.

key parameter turns out to be the income elasticity of healthcare expenditure, ϵ. We pick a round number, $\epsilon = 1.0$, for our baseline assumption, which happens to be close to leading estimates using international data (e.g., Newhouse (1977) estimates $\epsilon = 1.3$). We provide calibrations for a wide range of ϵ spanning these values as well as $\epsilon = 0$, which is equivalent to a model in which consumers vary only in disease risk. Because the disease-risk distribution is even more Zipf similar than the distribution of the product of risk and income, the potential for deadweight loss is enormous in the model with only disease risk, for example, reaching 70% of the overall disease burden in the vaccine market.

We also use the calibrations to perform a variety of counterfactual exercises, assessing the welfare gains and losses from government interventions such as a subsidy, reference pricing, and price-discrimination ban. Regarding a subsidy, owing to the Zipf-similar shape of demand in the baseline vaccine calibration, a small subsidy is enough to swing the equilibrium from a high price at which few consumers are served to universal vaccination. Thus universal vaccination may be a more robust public policy than previously thought, possible to rationalize even in the absence of epidemiological externalities, and possible to effectuate without a mandate. Regarding reference pricing, supposing all other countries peg their prices to some proportion, υ, of the United States', a monopoly manufacturer of either a vaccine or a drug would prefer to serve the United States even if the ceiling were as low as $\upsilon = 0.2$, i.e., 20% of the U.S. price. Regarding price discrimination, cross-country price discrimination can be quite lucrative in our model because each country's consumers are homogeneous, so one different price per country is all that is needed to accomplish perfect price discrimination and attain the social optimum. Hence, banning price discrimination can be quite distortionary, leading to a 56% decline in vaccine producer surplus and a 62% decline in drug producer surplus, declines equal to the potential increase in deadweight loss in these markets.

As discussed, the most closely related paper in the past literature is Kremer and Snyder (2015).[3] Other of our past related work includes Kremer, Snyder, and Williams (2012), which focuses on epidemiological externalities, and Kremer and Snyder (2016), which provides general results

[3] Kremer and Snyder (2015) was initially circulated as a series of National Bureau of Economic Research working papers (Kremer and Snyder 2003, 2013). The international calibration provided in Section 6 of the 2013 working paper but cut from the 2015 published version became the germ of the present paper. Besides Figure 4, the other calibrations as well as all the theoretical results are new developments.

bounding deadweight loss for product markets beyond pharmaceuticals and provides international calibrations allowing for a separate lognormal distribution of income within each country. A detailed discussion of the connection between this line of our research and other authors' work across a variety of literatures is provided in Kremer and Snyder (2015). Here we highlight just two key connections. The present 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 2015). 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, whereas distortion in our model is due to underconsumption and too little entry. The present paper also contributes to the literature on the revenue and welfare consequences of government interventions in the international pharmaceutical market. See, e.g., Kyle (2007), Sood *et al.* (2008), and a series of papers by Patricia Danzon and coauthors, including Danzon and Ketcham (2005) and Danzon, Wang, and Wang (2005).

The paper is organized as follows. Most of the theoretical analysis is presented in the next section. There, we set up the model providing the basis for the calibrations and derive a result (Proposition 1) tightening the bound on deadweight loss from our previous work. Section 5 presents the data, Section 6 presents calibrations for the baseline scenario, and Section 7 presents calibrations for a suite of alternative scenarios. Section 8 concludes. An appendix provides the proofs of all propositions.

2. Model

We base our analysis on the most general model in Kremer and Snyder (2015). This model, appearing in their Section V, is general in several respects. First, it allows for general values of the parameters c_j, s_j, and e_j introduced later. Second, it allows for multiple sources of consumer heterogeneity embodied in the random variables $X \in [0,1]$, representing disease risk, $Y \geq 0$, representing willingness to pay to avoid a unit of disease harm (for simplicity, this can be thought of as income or wealth), and $H \geq 0$, representing the consumer's draw of disease harm conditional on contracting it, where H has unit mean and be mean-independent of X and Y, i.e., $E(H) = E(H|Y = y) = E(H|X = x) = 1$. Ex ante, the consumer draws realizations x and y of random variables X and Y. Ex post, the consumer's disease status is a draw from a Bernoulli distribution with probability x of contracting it. Conditional on contracting the disease, the consumer draws a realization h of harm H. Letting B denote total disease burden from an ex ante perspective, normalizing the mass of consumers to 1, we have $B = E(XYH) = E(XY)$, where the last equality follows from the assumptions on H. Kremer and Snyder (2015) show in Proposition 14 that these three random variables are sufficient to capture quite general models of consumer heterogeneity.

A monopoly pharmaceutical manufacturer can develop two products indexed by j to sell on the private market: a vaccine ($j = v$) and/or a drug ($j = d$). If the firm chooses to develop product j, it must pay fixed cost $k_j \geq 0$, reflecting expenditures on research, capacity, etc. Let $c_j \geq 0$ denote product j's constant marginal production cost and $s_j \geq 0$ the harm from its side effects. Let $e_j \in [0,1]$ denote its efficacy or conversely $r_j = 1 - e_j$ denote its failure rate. To simplify the notation, we ignore discounting (or take all values to reflect ex ante present discounted values).

Taking a step back to assess what our model does and does not capture, the model abstracts from epidemiological externalities—which would obviously bias a firm against a vaccine reducing the spread of the disease (and consequently shrinking its own market)—by assuming that the products, while preventing individuals from experiencing disease symptoms, do not slow transmission.[4] The key difference between the products in the model is that a vaccine is purchased before the consumer has contracted the disease and so still faces an uncertain disease risk; a drug is purchased after disease risk has been resolved into binary disease status (infected or not). This difference can generate different shapes for the pharmaceuticals' demand curves, leading to differences between the producer's ability to appropriate surplus with the two products. It is well known since Arrow (1962) that the monopolist's inability to appropriate consumer surplus and Harberger deadweight loss results in suboptimal innovation incentives. An additional distortion here is created by the producer's bias toward the more lucrative product. Obviously, the producer will *prefer* the product

[4] We do not deny the importance of such externalities—indeed some of our other work (Kremer, Snyder, and Williams 2010) focuses exclusively on such externalities—but want to focus on other factors in this paper. An alternative way to shut down epidemiological externalities would be to focus on non-infectious conditions such as heart attacks.

with lower values of the R&D cost (k_j), marginal production cost (c_j), side effects (s_j), and ineffectiveness (r_j), but we do not call this preference a *bias* because it is shared by the social planner: both agree low values are better. Our analysis focuses on the wedge between private and social innovation incentives, arising here due to factors affecting surplus appropriability. Maximize the wedge between private and social innovation incentives will not be as simple as maximizing the difference between, say, k_v and k_d.

Letting σ denote the firm's product-development strategy, it has four possibilities: produce the vaccine alone ($\sigma = v$), the drug alone ($\sigma = d$), both products ($\sigma = b$), or nothing ($\sigma = n$). Consider the continuation game following its having developed exactly one of the products, i.e., $\sigma = j$ for $j \in \{v, d\}$. Let p_j denote price and $Q_j(p_j)$ the demand curve for product j. (Demand is also a function of s_j and r_j arguments besides p_j are suppressed for brevity in the notation.) Let $PS_j(p_j) = (p_j - c_j)Q_j(p_j)$ denote producer surplus, $CS_j(p_j) = \int_{p_j}^{\infty} Q_j(x)dx$ consumer surplus, and $TS_j(p_j) = PS_j(p_j) + CS_j(p_j)$ total surplus. Note PS_j and TS_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—i.e., 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 $W_j(p_j) = TS_j(p_j) - k_j$.

Next consider the continuation game following the firm's having developed both products, i.e., $\sigma = b$. Let p_{bv} denote the price it sets for the vaccine and p_{bd} for the drug. With both products available, the demand for vaccines $Q_{bv}(p_b)$ and for drugs $Q_{bd}(p_b)$ in general are functions of the vector of both prices, $p_b = (p_{bv}, p_{bd})$. These demands can be tied back to the single-product demand curves: $Q_j(p_j) = Q_{bj}(p_j, \infty)$. Denote the vector-valued demand function when both products are produced as $Q_b(p_b) = (Q_{bv}(p_b), Q_{bd}(p_b))$. Letting $c_b = (c_v, c_d)$, we can write producer surplus from both products as the dot product $PS_b(p_b) = (p_b - c_b)' Q_b(p_b)$. Consumer surplus is

$$CS_b(p_b) = \int_{p_{bv}}^{\infty} Q_{bv}(p_v, p_{bd})\, dp_v + \int_{p_{bd}}^{\infty} Q_{bd}(p_d, p_{bv})\, dp_d,$$

and total surplus is $TS_b(p_b) = CS_b(p_b) + PS_b(p_b)$.

Next consider the continuation game following the firm's developing no product, i.e., $\sigma = n$.

Nothing can be sold if nothing has been developed. Thus, for all $p_n \geq 0$, $Q_n(p_n) = 0$, implying

$$PS_n(p_n) = CS_n(p_n) = TS_n(p_n) = 0. \tag{1}$$

Equilibrium values are denoted with a single star. For example, $p^*_\sigma = \text{argmax}_p PS_\sigma(p)$ denotes the monopoly price, $q^*_\sigma = Q_\sigma(p^*_\sigma)$ monopoly quantity, $PS^*_\sigma = PS_\sigma(p^*_\sigma)$ monopoly producer surplus, W^*_σ denote monopoly welfare, etc. (Note that p^*_σ and q^*_σ are vectors if $\sigma = b$.) Rather than an arbitrary product strategy σ, we will often be interested in equilibrium values associated with the equilibrium product strategy σ^*. To streamline notation, we denote these values by starring the relevant variable but dropping the product-strategy subscript: i.e., $p^* = p^*_{\sigma^*}$, $q^* = q^*_{\sigma^*}$, $PS^* = PS^*_{\sigma^*}$, $W^* = W^*_{\sigma^*}$, etc. Variables with two stars denote socially optimal values. For example, p^{**}_σ denotes the socially optimal price. We have $p^{**}_j = c_j$ if a single product j is produced and $p^{**}_b = (c_v, c_d)$, the vector of marginal costs, if both are produced. Furthermore, for any product strategy σ, we have $PS^{**}_\sigma = 0$ and $TS^{**}_\sigma = CS^{**}_\sigma$. Rather than an arbitrary product strategy, σ, we will often be interested in socially optimal values under the socially optimal product strategy, σ^{**}. To streamline notation, we denote these values by double starring the relevant variable but dropping the product-strategy subscript: i.e., $p^{**} = p^{**}_{\sigma^{**}}$, $q^{**} = q^{**}_{\sigma^{**}}$, $PS^{**} = PS^{**}_{\sigma^{**}}$, $W^{**} = W^{**}_{\sigma^{**}}$, etc.

While the analysis allows for general values of the parameters c_j, s_j, and r_j, the values $c_j = s_j = r_j = 0$, which we refer to as benchmark values, play a special role, allowing us to "level the playing field" for the two products in all respects save for the timing of when they are sold.[5] To denote equilibrium under benchmark parameters, we replace the star with a zero in the superscript. For example, p^0 denotes the monopoly price under the equilibrium product strategy for benchmark parameters, and PS^0 denotes the resulting equilibrium producer surplus. To denote the social optimum under benchmark parameters, we replace the two stars with two zeros in the superscript. For example, p^{00} denotes the socially optimal price under the socially optimal product strategy for benchmark parameters, and W^{00} denotes the resulting social welfare in this social optimum.

Suppose that the firm chooses a non-trivial product strategy $\sigma \in \{v, d, b\}$ under benchmark

[5] For example, allowing for positive values of c_v and c_d, the normalization $c_v = c_d$ equalizes the cost of producing a dose but introduces a bias in the aggregate cost of a universal pharmaceutical program. In particular, universal vaccination would be more costly than universal drug treatment by a factor equal to the reciprocal of the prevalence rate. In addition, the benchmark parameters are associated with the most extreme worst-case bounds under some conditions; see Proposition 12 from Kremer and Snyder (2015).

parameters. Since products are costless to manufacture, the socially optimal pricing policy is to give the products away: $p_\sigma^{00} = 0$. Since products have no side effects, all consumers purchase; and since products are perfectly effective, their consumption relieves the entire disease burden. Therefore,

$$TS_\sigma^{00} = B \quad \text{for } \sigma \in \{v, d, b\}. \tag{2}$$

3. Bounding Deadweight Loss

The highlight of this section are propositions providing new bounds on deadweight loss. Before presenting the propositions, we define several deadweight loss concepts that we work with.

3.1. Harberger Deadweight Loss

We refer to the area of Harberger's (1954) triangle, equal to the deadweight loss stemming from prices above marginal cost, as Harberger deadweight loss, denoted $HDWL_\sigma^*$, where σ indexes the product strategy under consideration. We have $HDWL_\sigma^* = TS_\sigma^{**} - TS_\sigma^* = TS_\sigma^{**} - PS_\sigma^* - CS_\sigma^*$. Deadweight loss from all sources including distortions due to a possible mismatch between the product strategy σ and the socially efficient one, σ^{**}, is denoted $DWL_\sigma^* = W^{**} - W_\sigma^*$. As before, we can condition on the firm's equilibrium product strategy σ^* rather than arbitrary product strategy σ, writing $HDWL^* = TS_{\sigma^*}^{**} - TS^*$ and $DWL^* = W^{**} - W^*$.

It will be useful to work with relative versions of these deadweight loss measures, normalizing by disease burden as a measure of the size of the market. Thus, relative Harberger and total deadweight losses are $HDWL_\sigma^*/B$ and DWL_σ^*/B, respectively, when conditioned on an arbitrary product strategy σ; and $HDWL^*/B$ and DWL^*/B, respectively, when conditioned on the equilibrium product strategy σ^*.[6]

The analysis proceeds by trying to find simple characterizations of these relative deadweight loss concepts. Substituting for $HDWL_\sigma^*$, we have

$$\frac{HDWL_\sigma^*}{B} = \frac{TS_\sigma^{**}}{B} - \frac{PS_\sigma^*}{B} - \frac{CS_\sigma^*}{B} = \frac{TS_\sigma^{**}}{B} - \rho_\sigma^* - \gamma_\sigma^*, \tag{3}$$

[6] Tirole (1988) proposes slightly different expressions for relative deadweight loss, dividing by first-best social surplus rather than disease burden. By equation (2), our relative concepts coincide with his for benchmark parameters.

where $\rho_\sigma^* = PS_\sigma^*/B$ is relative producer surplus and $\gamma_\sigma^* = CS_\sigma^*/B$ is relative consumer surplus. Allowing the firm to pursue the equilibrium product strategy σ^*, it follows from equation (3) that

$$\frac{HDWL^*}{B} = \frac{TS_{\sigma^*}^{**}}{B} - \rho^* - \gamma^*, \tag{4}$$

where relative surpluses are defined in the obvious way: $\rho^* = PS^*/B$ and $\gamma^* = CS^*/B$. Assuming benchmark parameter values and further that the equilibrium product strategy is non-trivial (i.e., $\sigma^0 \in \{v,d,b\}$), equation (2) implies

$$\frac{HDWL^0}{B} = 1 - \rho^0 - \gamma^0. \tag{5}$$

3.2. Comprehensive Deadweight Loss

No such simple formulas are available for the more comprehensive measure of relative deadweight loss, DWL^*/B, which involves both pricing and product-choice distortions. The approach of Kremer and Snyder (2015) was to focus instead on finding a simple formula for the worst case—the supremum—on this relative deadweight loss. Though they did not find an exact expression for the supremum, they found a lower bound on it, reported in their Proposition 15.

To understand the derivation of the bound, some notation is in order. Let $PS_{\min}^* = \min\{PS_v^*, PS_d^*\}$ and $PS_{\max}^* = \max\{PS_v^*, PS_d^*\}$, respectively, be the minimum and maximum producer surpluses from the two individual products; and let $\rho_{\min}^* = PS_{\min}^*/B$ and $\rho_{\max}^* = PS_{\max}^*/B$ be the corresponding relative measures. Let $PS_{\min}^0$, $PS_{\max}^0$, $\rho_{\min}^0$, and $\rho_{\max}^0$ be the corresponding variables in equilibrium under benchmark parameters. Proposition 15 stated that the supremum on deadweight loss is no less than $\rho_{\max}^0 - \rho_{\min}^0$. This bound was derived by restricting attention to the benchmark parameters and to just one possible distortion, arising when the firm develops the more lucrative product though its fixed cost, k_j, is higher than the other product's. The excess of one product's fixed cost over the other's constitutes deadweight loss because, at baseline parameter values, the social planner can generate the social optimum with either product by freely distributing it. In the limit, the highest the difference in fixed costs can be and still have the firm choose to develop the more lucrative product equals the difference between the two products' producer surpluses, or $\rho_{\max}^0 - \rho_{\min}^0$

expressed as a percentage of disease burden. This expression is a lower bound because considering more general parameters or additional sources of distortion would only serve to increase potential deadweight loss.

We have been able to tighten this bound in the next proposition. The proofs of this and subsequent propositions are provided in the appendix.

Proposition 1. *Consider a model of the pharmaceutical market with multiple sources of consumer heterogeneity and general values of the parameters* $c_j, s_j, r_j \geq 0$. *The supremum on relative deadweight loss is at least* $1 - \rho^0_{\min}$*:*

$$\sup_{\{k_j, c_j, s_j, r_j \geq 0 | j = v, d\}} \left(\frac{DWL^*}{B} \right) \geq 1 - \rho^0_{\min}. \tag{6}$$

The proof recognizes that there is a worse distortion than that due to the wrong product choice; an even greater distortion can be generated by adjusting k_v and k_d such that nothing is produced. The whole of first-best social surplus then constitutes deadweight loss. For benchmark parameters, we can compute the supremum on deadweight loss by solving the problem of choosing fixed costs to maximize first-best social welfare subject to the constraint that the firm develops no product in equilibrium. Suppose for concreteness that the vaccine is the less lucrative of the two products, implying $PS^0_v = PS^0_{\min}$. Then the constrained maximum on social welfare is approached in the limit as k_v approaches the margin below which the firm chooses to develops the vaccine ($k_v \downarrow PS^0_{\min}$) and k_d grows large ensuring the firm chooses not to develop the drug ($k_d \uparrow \infty$). The value from the constrained maximization problem thus equals $\lim_{k_v \downarrow PS^0_{\min}, k_d \uparrow \infty} (TS^{00} - k_v) = TS^{00} - PS^0_{\min}$, which constitutes deadweight loss since it is unrealized because no product is developed in equilibrium. Substituting from equation (2) and dividing by B gives the left-hand side of equation (6). It is a lower bound rather than an exact value because we have fixed the parameters at their benchmark values for the purposes of the discussion; the supremum over more general parameters may be higher yet.

It is obvious that the new bound is weakly tighter than the previous one since $\rho^0_{\max} \leq 1$ implies $1 - \rho^0_{\min} \geq \rho^0_{\max} - \rho^0_{\min}$. To understand how much tighter it can be in practice, consider a scenario in which a vaccine is about as lucrative as a drug (i.e., $\rho^0_v \approx \rho^0_d$), but neither is particularly lucrative (i.e., $\rho^0_v, \rho^0_d \ll 1$). Using the old bound, $\rho^0_{\max} - \rho^0_{\min} \approx 0$, we cannot come to any more than the sterile conclusion that the supremum on deadweight loss is positive. The new bound $1 - \rho^0_{\min} \gg 0$

allows us to conclude that deadweight loss can be quite bad in the extreme. Just such a scenario will play out in the calibrations.

The next proposition verifies that the expression $1-\rho^0_{\min}$, which merely bounds the supremum on deadweight loss in Proposition 1 for general parameters, is an exact expression for the supremum for benchmark parameters.

Proposition 2. *Consider a model of the pharmaceutical market with multiple sources of consumer heterogeneity, fixing the parameters at their benchmark values $c_j = s_j = r_j = 0$. The supremum on relative deadweight loss is exactly $1-\rho^0_{\min}$:*

$$\sup_{\{k_v,k_d \geq 0\}} \left(\frac{DWL^0}{B} \right) = 1-\rho^0_{\min}. \tag{7}$$

Some of the calibrations analyze the counterfactual case in which the firm is restricted to developing just one of the pharmaceutical products, $j = v$ or $j = d$. It is easy to see—and indeed is immediate from the proofs of the previous propositions—that the supremum on deadweight loss is exactly $1-\rho^0_j$ for benchmark parameters in this case.

Proposition 3. *Suppose that only product $j = v$ or $j = d$ can be developed, not both. Fix the parameters at their benchmark values $c_j = s_j = r_j = 0$. The supremum on relative deadweight loss is exactly $1-\rho^0_j$:*

$$\sup_{\{k_j \geq 0\}} \left(\frac{DWL^0_j}{B} \right) = 1-\rho^0_j. \tag{8}$$

The definition $\rho^0_{\min} = \min_{j\in\{v,d\}} \rho^0_j$ implies $1-\rho^0_{\min} = \max_{j\in\{v,d\}}(1-\rho^0_j)$. Combining this equation with the results from Propositions 2 and 3 it is immediate that we can compute the deadweight-loss supremum in the comprehensive case in which both products can be produced by computing the deadweight-loss suprema from each isolated product market and taking the larger of these suprema.

3.3. Zipf Worst Case

The analysis so far asked what fixed costs and other parameters generate the worst deadweight loss for a given demand curve. Kremer and Snyder (2015) proceed to ask what demand curve shapes generate the worst deadweight loss among those with the same area underneath (the same B in our setting). The answer is what they term the symmetrically truncated Zipf (STRZ) demand.

A Zipf distribution is the special case of a power-law distribution with an exponent equal to 1, which in the disease context intuitively means that each doubling of disease risk cuts the number of consumers with at least that risk in half. The resulting demand curve is unit elastic, implying that all feasible prices generate the same revenue, and hence the same producer surplus under benchmark parameters. Put another way, no price is especially lucrative with this demand curve; all generate the same (low) producer surplus. Example STRZ demands are provided later in Figures 4 and 5; they can be visualized as rectangular hyperbolas truncated at the top and bottom. The truncations are technical features keeping both the highest conceivable consumer value and the area under the curve constant.

Kremer and Snyder (2015) proposed a measure of how close a demand curve comes to the STRZ worst case, called the Zipf similarity of demand. Kremer and Snyder (2016) extended this result, providing the necessary adjustments to apply this formula to general settings with arbitrary scaling of the price and quantity axes. Adapting the formula from equation (16) of Kremer and Snyder (2016) to reflect the present notation, the Zipf similarity of the demand for pharmaceutical j under benchmark parameters, Z_j^0, is

$$Z_j^0 = \frac{1-\rho_j^0}{1-\underline{\rho}(\mu_j^0)}, \tag{9}$$

where $\underline{\rho}(\mu_j^0)$ is the producer-surplus ratio of the associated STRZ demand. By Proposition 3, Zipf similarity is simply the supremum on deadweight loss for pharmaceutical j relative to the supremum on deadweight loss for the associated STRZ demand. If the supremum on deadweight loss for pharmaceutical j is 50% that of the STRZ worst case, then we say the demand for pharmaceutical j is 50% Zipf similar. To have a supremum on deadweight loss close to the STRZ worst case, all points on the demand curve for pharmaceutical j must lie in a neighborhood around the STRZ demand curve. We will use equation (9) to measure the Zipf similarity of calibrated demand curves in Section 6.

To fill in some technical details behind equation (9), by the "associated" STRZ demand curve, we mean the STRZ demand in the same class as the demand for pharmaceutical j, where the class is indexed by the so-called mean-to-peak ratio, μ_j^0, equal to the ratio of the mean consumer value for product j in the social optimum to the highest value in the population.[7] By Proposition 3 of Kremer and Snyder (2016), the producer-surplus ratio for this STRZ demand curve equals

$$\underline{\rho}(\mu_j^0) = \frac{-1}{LW_{-1}(-\mu_j^0/e)}, \tag{10}$$

where LW is the Lambert W function, the inverse relation $W(z)$ of the function $z = We^W$. The -1 subscript on LW denotes the lower branch of this relation.[8]

4. Calibration Methodology

This section describes the methods used to calibrate the new bounds on potential deadweight loss in the market for HIV pharmaceuticals.[9] There are a number of advantages to assuming benchmark values of the parameters: $c_j = s_j = r_j = 0$ for $j = v, d$. In addition to those discussed in footnote 5, another advantage is that for benchmark parameters Proposition 2 gives an exact expression for the supremum on deadweight loss rather than a bound. We thus maintain benchmark parameters throughout the remainder of the analysis. Following our earlier notation, we use a zero superscript (instead of a star) to indicate equilibrium values evaluated at the benchmark parameters and two zeros in the superscript (rather than two stars) to indicate socially optimal values under benchmark parameters.

By Proposition 2, calibration of the supremum on deadweight loss under benchmark parameters reduces to calibration of $\rho^0_{\min}$, which by the definition $\rho^0_{\min} = \min(\rho^0_v, \rho^0_d)$ in turn reduces to

[7] Adapting the formula for from Lemma 1 of Kremer and Snyder (2015) to the present context, for the vaccine market we have

$$\mu_v^0 = \frac{\sum_{i=1}^I N_i X_i Y_i}{(XY)_{(I)} \sum_{i=1}^I N_i},$$

where $(XY)_{(I)}$ denotes the maximum order statistic for the product of X_i and Y_i, i.e., the maximum value that X_iY_i takes on across countries. For the drug market,

$$\mu_d^0 = \frac{\sum_{i=1}^I N_i X_i Y_i}{Y_{(I)} \sum_{i=1}^I N_i X_i},$$

where $Y_{(I)}$ denotes the maximum value of Y_i across countries.

[8] The branches of LW are built-in functions in standard mathematical software packages including Mathematica, Matlab, and R. Other ways to compute $\underline{\rho}(\mu_j^0)$ besides equation (10) include reading the value from the graph in Kremer and Snyder's (2015) Figure 4 or taking the value from the tabulation in Kremer and Snyder's (2016) Table 2.

[9] We call this a "calibration" rather than an "estimation" exercise because we assume convenient forms for demand and cost and we fix certain important parameters (including c_j, s_j, r_j, and the income elasticity) rather than estimating them from price and quantity data.

calibration of demand for a vaccine alone and a drug alone. We are exempted from having to calibrate the more complicated demands that apply when both products are offered because they do not show up in the formula. To calibrate demand for a vaccine alone and drug alone, we need to return to the random variables X, Y, and H introduced at the outset of Section 2 and specify their distributions in the consumer market. We take the consumer market to consist of the entire world population. Country $i \in \{1, \ldots, I\}$ has a population of $N_i > 0$ risk-neutral consumers. Consumers within a country are homogeneous, each having the same disease risk $X_i \in (0, 1]$ and income $Y_i > 0$. Since we do not have data on harm, we abstract from cross-country variation in H_i, adopting the normalization $H_i = 1$.[10]

As an intermediate step in deriving pharmaceutical demands, consider the counterfactual situation of a consumer who surely contracts the disease. Assume that the elasticity of an individual's healthcare demand with respect to income is a constant ϵ across income and across individuals. For this to be the case, his benefit from avoiding harm from the disease as a function of his income Y_i must take the form Y_i^{ϵ}.[11] Now move from the counterfactual case in which the consumer surely contracts the disease to the factual case in which he has disease risk X_i. His expected benefit from a vaccine is the product of this risk and the benefit of avoiding sure harm, i.e., $X_i Y_i^{\epsilon}$. If the firm's product strategy is to produce the vaccine alone, global vaccine demand is

$$Q_v(p_v) = \sum_{i=1}^{I} \mathbf{1}(X_i Y_i^{\epsilon} \geq p_v) N_i, \tag{11}$$

where $\mathbf{1}(\cdot)$ is the indicator function, equal to 1 if the statement in parentheses is true and 0 otherwise. Turning to demand for a drug, because it is sold ex post, after disease status has been realized, any consumer who has contracted the disease buys a drug sold at price p_d as long as $Y_i^{\epsilon} \geq p_d$. In expectation, $N_i X_i$ consumers in country i end up contracting the disease. Thus global

[10] Country populations could be rescaled as $N_i / \sum_{k=1}^{I} N_k$ to arrive at a unit mass of consumers as in the model, but there is no need to do so because the calibration results will be expressed in percent terms, scaling them by first-best social surplus (equivalently, disease burden).

[11] The most general form preserving the property of constant income elasticity is $A_i Y_i^{\epsilon}$, allowing the leading coefficient to vary across consumers by taking it to be a random variable. We do not allow for that source of heterogeneity because doing so would introduce a third random variable characterizing consumers in a country, contradicting the maintained assumption that consumers are fully characterized by just X_i and Y_i. A form that is more general but does not introduce a third source of heterogeneity is AY_i^{ϵ}, with a leading coefficient A that is constant across consumers. We normalize $A = 1$ without loss of generality since all of our surplus calculations will be expressed as a proportion of disease burden; A is a scale factor which would divide out of the proportion.

drug demand is

$$Q_d(p_d) = \sum_{i=1}^{I} \mathbf{1}(Y_i^\epsilon \geq p_d) N_i X_i. \tag{12}$$

Demands when the firm produces both are more complicated. As mentioned above, the supremum on deadweight loss can be calibrated without having to solve for the $\sigma^0 = b$ continuation equilibrium. We report these demands for completeness but relegate them to a footnote.[12]

Equations (11) and (12) provide the foundation for our calibrations. To translate these formulas into demand units, the only additional elements required are country-level data on N_i, X_i, and Y_i (described in the next section) and an assumption about the income elasticity, ϵ. The practical implementation details are best discussed in the context of a simple example. Consider the toy example in Table 1 of a world with two countries. We first show how (11) can be used calibrate vaccine demand. Setting ϵ to the value we take for our baseline scenario, $\epsilon = 1$, and maintaining benchmark parameters (no side effects and perfect efficacy), country i's willingness to pay for a vaccine equals the product $X_i Y_i$. As shown in the last column, country 1 has lower willingness to pay, $X_1 Y_1 = 100$, than country 2, at $X_2 Y_2 = 400$. These willingness-to-pay numbers are the only relevant pricing points for a monopolist wishing to extract as much surplus as possible. According to equation (11), vaccine demand equals the cumulative population of countries having consumers with a willingness to pay at least as high as the given price: $Q_v(400) = N_2 = 1,500$ and $Q_v(100) = N_1 + N_2 = 2,500$. The resulting demand curve is the step function shown in Panel A of Figure 1. Under the benchmark assumption of costless production, vaccine producer surplus equals vaccine revenue, $PS_v(400) = 400 \cdot 1,500 = 675,000$ and $PS_v(100) = 100 \cdot 2,500 = 250,000$, implying $PS_v^0 = 675,000$, the area of the shaded rectangle in Panel A.

Next, we use (12) to calibrate drug demand. Willingness to pay for this product equals Y_i. The only two relevant pricing points for a monopolist are thus 10,000 and 9,000. Demand at

[12] With benchmark values of the parameters $s_j = r_j = 0$, one can show

$$\begin{aligned} \tilde{Q}_v(p_{bv}, p_{bd}) &= \sum_{i=1}^{I} \mathbf{1}(X_i \geq p_v/p_d) \mathbf{1}(X_i Y_i^\epsilon \geq p_v) N_i \\ \tilde{Q}_d(p_{bd}, p_{bv}) &= \sum_{i=1}^{I} \mathbf{1}(X_i \leq p_v/p_d) \mathbf{1}(Y_i^\epsilon \geq p_d) N_i X_i. \end{aligned}$$

With general values of s_j and r_j, the expressions become considerably more complicated. Among other things, with imperfect efficacy, a consumer who purchases a vaccine than turns out to be ineffective may later purchase the drug as well.

these pricing points equals the cumulative infected population, N_iX_i, of countries with at least that willingness to pay: $Q_d(10{,}000) = N_1X_1 = 10$ and $Q_d(9{,}000) = N_1X_1 + N_2X_2 = 85$. The resulting demand curve is shown in Panel B of Figure 1. Drug producer surplus is $PS_d(10{,}000) = 100{,}000$ and $PS_d(9{,}000) = 765{,}000$, implying $PS_d^0 = 765{,}000$, the area of the shaded rectangle in Panel B.

In this toy example, the drug is more lucrative than the vaccine. Intuitively, most of the heterogeneity across consumers in this toy example can be traced to X_i, which is five times higher in country 2 than 1. This heterogeneity disappears by time the drug is sold because disease risk resolves into disease status, allowing the monopolist to extract almost all the available surplus—$PS_d^0 = 765{,}000$ is 99% of the total disease burden, $B = 775{,}000$—from drug consumers, who as Panel B of Figure 1 shows are virtually homogeneous. More formally, a comparison of demand formulas (11) and (12) reveals that they are identical except for the shifting of X_i outside of the indicator function's argument in drug demand (12). This shift of course changes the shape of the demand curve as the move from Panel A to B in Figure 1 illustrates. Whether this shape change leads to a more or less lucrative demand curve in general is impossible to say—cases can be constructed in which $\rho_v^0 > \rho_d^0$ and and other cases in which the reverse inequality holds—depending on the joint distribution of N_i, X_i, and Y_i. More concrete conclusions are available when there is little heterogeneity in Y_i. An immediate corollary of Kremer and Snyder's (2015) Proposition 3 is that if income is homogeneous across countries, i.e., if $Y_i = \bar{Y}$, and at least two countries have different positive values of X_i, then $\rho_d^0 > \rho_v^0$. By continuity, this inequality still holds if Y_i varies across countries as long as the variance is sufficiently small. This is the case in the toy example, in which Y_1 is only 11% higher than Y_2.

We use the same methods to calibrate demand with actual cross-country data as used here in the toy example. The demand curve will still be a step function but with finer steps due to the large number of actual countries. As in the toy example, the number of relevant pricing points for the monopolist is no greater than the number of countries. Only this manageable number of prices needs to be checked to compute equilibrium (producer surplus maximizing) prices. Equilibrium prices and quantities can be combined with the calibrated demand curves to calculate surpluses, ratios of surpluses to disease burden, and the deadweight loss supremum.

5. Data

This section discusses the data sources for N_i, X_i, and Y_i used in the calibrations. A preliminary question is what year would provide the best snapshot of the market for HIV pharmaceuticals. While the best year is the current one for most applications, this is not necessarily the case here. To calibrate demand for HIV pharmaceuticals, we would like some measure of the burden of the disease in the state of nature without any pharmaceuticals. Referring to the lower dotted curve in Figure 2, a growing percentage of people living with HIV have been receiving antiretroviral treatment (ART), from the negligible percentage in 2000 to nearly half in the most recent data. This expansion of treatment may have reduced transmission, and coupled with other initiatives to curtail the spread of HIV and other epidemiological trends resulted in a decline in HIV prevalence after its peak in 2003 (see the upper solid curve in Figure 2). The dip turned out to be short lived, rising again after 2010. This more recent rise may belie good news, reflecting a decline in HIV mortality resulting from the expansion of ART treatment. We select 2003 as our target year for data collection, a time when HIV prevalence reached a local peak while ART coverage was still fairly negligible.

The key inputs into our policy simulations are calibrations of the demand for a vaccine in (11) and a drug in (12). These equations contain three variables for which we need data: population N_i, disease risk X_i, and income Y_i for countries $i = 1, \ldots, I$. We obtained N_i and Y_i from the World Bank Open Data website, using gross domestic product (GDP) per capita for Y_i. We obtained X_i from a UNAIDS publication (UNAIDS 2004a).[13] Table 2 provides more details on the data sources and descriptive statistics for the main variables. Our dataset includes 158 countries.[14]

Figure 3 maps the geographical distribution of X_i and Y_i. The top panel shows the well-known concentration of HIV risk in sub-Saharan Africa. This would lead these countries to be the highest demanders of HIV pharmaceuticals in the calibration but for the fact, shown in the lower panel, that many are among the lowest income in the world, reducing their willingness to pay in the

[13] The UNAIDS website used as the source for the aggregate trends displayed in Figure 2 would be a natural source for country-level HIV data. However, the website seems to have expunged current and historical data for a substantial number of countries including the United States. We thus relied on a historical publication (UNAIDS 2004a) to recover country-level HIV data.

[14] Our data includes all countries with substantial populations except for Iraq, North Korea, Saudi Arabia, and Turkey, which are excluded because of missing HIV data. Other sources (UNAIDS 2004b) report very low prevalence rates for these countries, so their omission likely has little effect on our results.

calibration. Table 2 reports a negative correlation between X_i and Y_i of –0.19. Some countries such as the United States and South Africa are above the median in both X_i and Y_i and presumably will be centers of high demand for HIV pharmaceuticals in the calibration.

6. Calibrations for Baseline Scenario

This section presents calibrations for the baseline scenario in which countries are heterogeneous in both disease risk X_i and income Y_i, the income elasticity is taken to be $\epsilon = 1$, and the parameters are set at benchmark values $c_j = s_j = r_j = 0$. Under these assumptions, a consumer in country i has ex ante willingness to pay $X_i Y_i$ for a vaccine and ex post willingness to pay Y_i for a drug conditional on being infected. Our analysis of the baseline calibration divided proceeds by first considering the vaccine market in isolation and then the drug market in isolation. We then combine these separate results to compute a comprehensive bound on deadweight loss using Proposition 2. We conclude the section with an analysis of whether and how equilibrium in the baseline calibration can be improved with government subsidies.

6.1. Vaccine Market

The top panel of Figure 4 shows the calibrated (inverse) demand curve for a vaccine alone. Following the methodology laid out in Section 4, it is constructed by ordering countries by the product $X_i Y_i$ and then sequentially plotting that value on the vertical axis after stepping off a quantity given the country's population N_i on the horizontal axis.

Since we have assumed production is costless ($c_v = 0$), the price that maximizes producer surplus for the vaccine monopolist equivalently maximizes its revenue. Finding the price that maximizes producer surplus boils down to the geometric problem of finding the rectangle of largest area that can be inscribed underneath the demand curve. Holding the total area underneath the demand curve constant, some shapes are of course more conducive to inscribing large rectangles underneath than others. As discussed in Section 3.3, the worst case for inscKremer and Snyder (2015) solve for the worst case for the firm, which they term symmetrically truncated Zipf (STRZ) demand. The STRZ demand corresponding to the vaccine market is drawn as the grey curve in the figure. Notice that the calibrated demand curve largely overlaps the STRZ curve except where the United States appears, leading the demand curve to “belly out” there. The reason for the outsize influence of the United States on the market is that its consumers have very high incomes relative to most others, coupled with a moderate HIV risk. These factors generate a large value of the product X_iY_i for the large population of U.S. consumers. U.S. consumers turn out to be the marginal ones in the calibration: the producer-surplus-maximizing price just induces them to buy and strictly induces purchases by consumers in Botswana, South Africa, Swaziland, Bahamas, Namibia, Trinidad and Tobago, and Gabon. While consumers in these other countries are poorer than the United States, their extremely high HIV prevalence rates lead their demands to be higher demanders than in the United States.

Further details of the calibration are provided in Table 4. The vaccine sells at a price (in 2003 U.S. dollars) at \$130 to 344 million consumers, all but 10 million of these consumers from South Africa and the United States. Producer surplusis 44% of the social surplus B from completely relieving the disease burden. This 44% is the ratio of the shaded rectangle to the area under the inverse demand curve in the top panel of Figure 4. Consumers obtain 16% of B. The residual, 41%, is Harberger deadweight loss $HDWL_v^0$. The supremum on DWL_v^0 is achieved in the limit as k_v approaches producer surplus—44% of B—from above. The formula in Proposition 3 yields that this supremum equals $100-44 = 56\%$ of B. A lack of adequate incentives to enter the market could dissipate more than half of the social surplus from completely relieving the disease burden.

Equation (9) provided a formula for the Zipf similarity, Z_j^0, of the demand for product j, measuring how close the demand’s shape comes to the STRZ worst case. We $Z_v^0 = 0.66$, i.e., the calibrated vaccine demand curve is 66% similar to the STRZ worst case in that it generates 66% of the deadweight loss bound generated by the STRZ curve. This measure quantifies the moderately Zipf similar shape of the calibrated vaccine demand curve.

6.2. Drug Market

We next turn to the market for the drug alone. The calibrated (inverse) demand curve for a drug is shown in the lower panel of Figure 4. Given that a drug is only sold to consumers who contract the disease, it would not be surprising to see it sell at a much higher price to a much smaller group of consumers than a vaccine. The scale for drug price on the vertical axis is 100 times that for vaccine price, but the scale for drug quantity on the horizontal axis is only $1/100$ that for vaccine quantity. The combined scaling of the axes maintains the property that a unit of area reflects the same revenue and same surplus in both panels. Following the methodology laid out in Section 4, the drug demand curve is constructed by ordering countries in terms of Y_i (i.e., GDP per capita), reflecting consumers' ex post willingness to pay for a drug conditional on contracting the disease. We then sequentially plot that value on the vertical axis after stepping off quantity N_iX_i, equal to the expected number of people in the country who contract the disease.

The producer-surplus-maximizing price is \$27,400 (in 2003 U.S. dollars), at which marginal consumers in Italy and 17 other countries with higher incomes (including the United States) purchase. Although these countries have a combined population of 783 million, only 1.4 million doses end up being sold. These rich countries have a relatively low HIV prevalence rate (an average of 0.12%), so relatively few people end up contracting the disease and needing an HIV drug. The difference in the shape of the distribution of consumer values for a drug versus a vaccine leads drug to be less lucrative than the vaccine. The drug producer obtains only 38% of the social surplus B from completely relieving the disease burden, six percentage points less than with a vaccine, corresponding to the smaller area of the shaded producer-surplus rectangle in the lower compared to the upper panel. The shape of the drug demand curve is quite close to the STRZ worst case, drawn as a grey curve. Formally, the index of Zipf similarity is $Z_d^0 = 0.74$. Its shape is not conducive to inscribing a rectangle of substantial area underneath. Consumers obtain only 13% of B. The residual 49% is Harberger deadweight loss $HDWL_d^0$. Nearly half of B is lost because the firm's most lucrative strategy is to target the high-income market, which results in the exclusion of countries like South Africa with a slightly lower income but a tremendous disease burden. But $HDWL_d^0$ understates the potential for deadweight loss from all sources, DWL_d^0. Given the market is not very lucrative, the manufacturer may not have an incentive to enter at all. The supremum on DWL_d^0 is achieved in the limit as k_d approaches producer surplus—38% of B—from above. Using the formula in Proposition 3, this supremum equals $100-38 = 62\%$ of B.

6.3. Comprehensive Bound

To obtain a comprehensive bound on deadweight loss, we move from calibrations of the vaccine and drug market in isolation to a calibration allowing for the possibility that either or both could be produced. Proposition 2 provides an exact expression for the supremum on deadweight loss in this comprehensive situation. As argued in Section 3, it simply equals the greater of the deadweight-loss suprema from the isolated product markets, i.e., the greater of 56% and 62%, thus equal to 62%, reported in the last row of Table 4. This supremum can be approached in the limit as k_v approaches infinity, so the vaccine market is certainly non-viable, and k_d approaches 38% of B, so the drug market falls just short of the margin of viability.[15]

6.4. Government Subsidies

Kremer and Snyder (2016) note in Section 6 that Harberger deadweight loss can be highly unstable with a Zipf-similar demand curve. The monopolist may be almost indifferent between the equilibrium strategy that targets a small segment of high demand consumers and one that targets the bulk of consumers with a low price. A small subsidy may be enough to flip the equilibrium from one to the other, eliminating most of the Harberger deadweight-loss triangle.

Our baseline calibrations display exactly this property. Consider the calibrated vaccine market. Less than 6% of potential consumers are served at the high equilibrium price, leading to a large Harberger triangle, amounting to 41% of B. Introducing a government subsidy—a per-unit subsidy, which for accounting purposes assume is paid directly to the firm—and gradually raising it in penny increments has no effect on equilibrium price or quantity until it reaches \$7.69. The next penny increment to \$7.70 causes the firm to cut the price from \$130 to zero. This relatively modest subsidy, just 6% of the pre-subsidy equilibrium price, is enough to eliminate the Harberger triangle entirely. The subsidy is socially efficient for a wide range of parameters. Even with a social cost of public funds as high as 1.8 (so raising \$1 of taxes costs society \$1.80), social welfare would be higher under the \$7.70 subsidy than in the equilibrium without it. Interestingly, what was intended to be a subsidy policy with the target of mitigating Harberger deadweight loss ends up being equivalent to a universal vaccination program.

The effect of a subsidy on the drug market is similar. Introducing a government subsidy (again, a per-unit subsidy paid directly to the firm) and gradually raising it in penny increments has no effect on equilibrium price or quantity until it reaches \$941.34. The next penny increment to

[15] To see the improvement that Proposition 2 entails over previous results, compare the tight bound of 62% reported here to the bound from Proposition 15 of Kremer and Snyder (2015), equal to $\rho^0_{\max} - \rho^0_{\min} = 44 - 38 = 6\%$. The new bound point-identifies worst-case deadweight loss at 62% of B. The old bound tells us that the deadweight-loss supremum lies somewhere in the interval between 6% and 100% of B, a fairly uninformative statement in this calibration.

$941.35 causes the firm to cut the price from $27,387 to $265. While a subsidy of over $900 might seem high, since drug prices are much higher than vaccine prices, in fact it is only 3% of the pre-subsidy equilibrium drug price. Though the subsidy does not fully eliminate the Harberger triangle, it reduces it to less than 1% of first-best social surplus as the 11% of consumers who remain unserved have very low incomes and thus very low drug demand.[16] The subsidy would be socially efficient even if the social cost of public funds were as high as 2.5.

While the subsidy policy nearly eliminates Harberger deadweight loss conditional on the development of the product, it does little to eliminate the potential for deadweight loss at the extensive margin regarding whether the product is developed at all. Intuitively, the subsidies are too small to have much effect on incentives to enter the market. In the vaccine market, the $7.70 subsidy reduces the supremum on deadweight loss by just two percentage points, from 56% to 54% of B. In the drug market, the $941.35 subsidy also reduces the worst-case bound on deadweight loss by just two percentage points, from 62% to 60%. Substantially improve entry incentives—thus substantially reducing deadweight loss at the extensive margin—would call for much larger subsidies.

7. Results from Alternative Scenarios

Table 5 presents a variety of alternatives to our main calibration. For brevity, the table just presents relative producer-surplus ratios ρ_v^0 and ρ_d^0. These are summary indicators of how lucrative the respective markets for a vaccine and drug are, and can easily be converted into bounds on deadweight loss using the formula in Proposition 3. The larger of the two of these gives a comprehensive bound on deadweight loss by Proposition 2. For reference, the first row of the table repeats the relative producer-surplus ratio from the main calibration.

7.1. Scenarios Maintaining Uniform Pricing

We start by analyzing alternative scenarios continuing with the maintained assumption that the monopolist engages in uniform pricing. The first alternative calibration examines what happens if countries have the same incomes, differing only in disease risk. This change makes the vaccine market much less lucrative. The top panel of Figure 5, which plots the inverse demand for this

[16] Eliminating the Harberger triangle entirely would require almost double the subsidy, $1,867.

calibration, shows why. Whereas the high income and moderate HIV risk compounded to make the United States a substantial source of demand in the main calibration, the United States has now been pushed down the demand curve, removing the curve's "belly" that helped generate revenue before. Global disease risk X_i closely follows a power law with exponent 1, the so-called Zipf distribution. Vaccine demand inherits this property, leading an almost perfect symmetrically truncated Zipf (STRZ) shape in Kremer and Snyder's (2015) terms. This is the worst possible shape for trying to inscribe a rectangle underneath capturing surplus for the producer. As reported in the second row of Table 5, the vaccine producer only obtains 30% of the social surplus B from completely relieving the disease burden. The firm's strategy is to serve Uganda and higher-risk countries.

The opposite picture emerges with a drug. As disease risk is resolved into disease status ex post, consumers with the same incomes have no heterogeneity. The resulting inverse demand curve for the drug, shown in the lower panel of Figure 5, is a rectangle. The drug monopolist is able to extract 100% of B with a price set at consumers' homogeneous willingness to pay. There is no deadweight loss from any source, pricing or product strategy, on this market. Unfortunately, Propositions 2 and 3 tell us that the potential for deadweight loss in the comprehensive setting in which either or both products can be developed depends not on the best but on the worst outcome in the two isolated markets. Here, the comprehensive bound on relative deadweight loss equals $100-30=70\%$. The bound can be approached in the limit as k_v approaches 30% of B and k_d approaches infinity. More than two thirds of total surplus could be dissipated because of inadequate incentives to develop the less lucrative product, the vaccine, in this calibration.

The next alternative calibration gives all countries the same HIV risk and has them just vary by income, Y_i. We omit the graph of the demand curves from now on for brevity and just look at the relative producer-surplus ratio. In this calibration, both products generate the same producer surplus. There is no change in the nature of consumer heterogeneity from the ex ante to the ex post period. The market for both products is fairly lucrative, with the producer able to capture 57% of B.

The next calibration returns to the baseline with heterogeneity in both X_i and Y_i but uses the revised data for 2003 HIV prevalence that we use for X_i. Our goal is to see how robust our calibrations are to variation in HIV risk, which is measured with considerable error. In this alternative calibration, we use the revised data published in UNAIDS (2006) rather than the initial data published in UNAIDS (2004a). The correlation between the initial and revised HIV prevalence series is quite high, at 0.985. There is a noticeable change in the relative producer-surplus ratios, but they remain within four percentage points of the baseline.

The next set of alternative calibrations vary the income elasticity appearing in the formulas for the willingness to pay for a vaccine ($X_i Y_i^{\epsilon}$) and a drug (Y_i^{ϵ}). Given that ϵ is an exogeneous rather than estimated parameter in our calibrations, and it may have important effects on demand, it is important to examine the robustness of the results to variation in ϵ. Table 5 reports calibrations for 0.5 increments in ϵ from $\epsilon = 0$ to $\epsilon = 2$. Figure 6 shows what the pattern looks like graphically (as well as showing finer increments). The calibration for $\epsilon = 0$ is a repetition of the earlier one with heterogeneity only in disease risk. The calibration for $\epsilon = 1$ is a repetition of the baseline.

Looking at the overall pattern in Figure 6, ρ_v^0 is fairly flat in ϵ over the range $\epsilon \in [0.0, 0.7]$, hovering around $\rho_v^0 = 0.30$, falling to $\rho_v^0 = 0.28$ for $\epsilon = 0.5$. On the other hand, ρ_d^0 falls precipitously over the interval $\epsilon \in [0.0, 0.7]$, from $\rho_d^0 = 1.00$ to $\rho_d^0 = 0.35$. Both ρ_v^0 and ρ_d^0 turn upward for $\epsilon > 0.7$ and eventually overlap each other. Intuitively, for large ϵ, heterogeneity in income starts to matter more than heterogeneity in disease risk. But we know from Kremer and Snyder (2015) (see the first paragraph of Section V.B) that the two products are similarly good at generating producer surplus when there is heterogeneity in income alone. Indeed, for very large values of ϵ, the very highest income country starts to dominate demand as its income is raised to an increasingly large power.

Which ϵ is empirically most plausible? We are interested in pharmaceutical sales on the private market, so in the state of nature absent government procurement or insurance coverage. Such a state is far from modern conditions at least in the United States. Getzen (2000) provides a survey of empirical studies of the income elasticity of health expenditures, locating a handful of studies using U.S. micro data from an historical period when most of the population was uninsured. Getzen finds estimates from these studies in the $[0.2, 0.7]$ range. The midpoint of this interval, 0.4, was estimated by Anderson, Collette, and Feldman (1960) using 1953 data. Micro studies using U.S. data from the modern era with more insured consumers surveyed by Getzen (2000) find income elasticities near zero, corresponding to the calibration with only disease-risk heterogeneity. Cross-country studies typically produce higher estimates of ϵ. The handful of cross-country studies surveyed by Getzen (2000) estimated values of ϵ in the $[1.2, 1.4]$ range. The 1.3 midpoint was estimated by

Newhouse (1977). Figure 6 shows that relative producer surplus is higher (and the potential for deadweight loss consequently lower) at $\epsilon = 1.3$ than $\epsilon = 1.0$, though the estimates are fairly similar.

7.2. Scenarios Introducing Price Discrimination

We next turn to calibrations of alternative scenarios allowing the producer to engage in price discrimination. The first of these calibrations allows the producer full freedom to charge different prices across countries. Since our model takes consumers to be homogeneous within a country, this pricing strategy is equivalent to perfect price discrimination, allowing the producer to extract 100% of first-best social surplus regardless of the pharmaceutical it produces. It can extract 100% of first-best social surplus from the vaccine market by selling to all individuals in country i at a price of $p^0_{vi} = X_i Y_i^\epsilon$ and 100% from the drug market by selling to all infected individuals in country i at a price of $p^0_{di} = Y_i^\epsilon$. Thus, as reported in Table 5, $\rho^0_v = \rho^0_d = 100\%$. This would eliminate any possibility of deadweight loss, whether due to inefficient pricing or entry decisions.

Government policies sometimes interfere with firms' ability to perfectly price discriminate. A policy that allows pharmaceutical imports with no trade frictions would return the equilibrium to the baseline scenario with uniform pricing. An international ban on price discrimination would have the same effect. Either policy would re-introduce the potential for substantial deadweight loss in the market due to pricing and entry distortions. In the vaccine market, the deadweight-loss supremum rises from 0% to 56% when price discrimination is banned; and in the drug market, the supremum rises from 0% to 62%.

Kyle (2007) documents the panoply of other policies adopted by countries that preclude perfect price discrimination in pharmaceutical markets. Some countries control prices directly. Others use international reference pricing, tying domestic prices to prices in peer countries. Other policies include volume rebates, profit controls, reimbursement rate adjustments, etc. Space considerations prevent us from analyzing all of these policies. We focus on just one, international reference pricing, pursued to a greater or lesser extent by nearly 80% of the countries for which Kyle (2007) could obtain dispositive evidence. We model this simply, assuming that all the countries in the world use the United States as the reference country, capping prices in their countries to be no greater than some proportion of the U.S. price, denoted $\upsilon \geq 0$ (the Greek letter upsilon).

Calibrating this scenario is complicated by the endogeneity of the reference price. The firm may increase the reference price to relax the ceiling in other countries even at the sacrifice of some reference-country profits. This endogeneity is easily addressed in our simple model. There are only two relevant prices to consider charging in the United States. The firm can either decide to serve some U.S. consumers, in which case it should charge their maximum willingness to pay for product j and serve all of them. If it charges a penny more, no U.S. consumers would buy. In that case, it may as well exclude the United States entirely, or equivalently set the U.S. price to infinity, allowing the firm to perfectly price discriminate across all remaining countries. We proceed by calibrating the outcome from these two strategies, one serving the United States and the other excluding it, comparing the producer surplus from each, and selecting the more lucrative one as the equilibrium.

The resulting producer-surplus ratios ρ_v^0 and ρ_d^0 are shown in Table 5 for υ increasing in 0.5 increments from 0 to 2. The $\upsilon = 0$ case corresponds to the case in which the United States is the sole commercial market for the good; the producer either ignores all other countries or gives the product away there. Given costless production, $c_j = 0$, the producer is indifferent between ignoring these other countries and freely supplying them (of course consumer surplus is much higher with the latter option). Producer surplus is 37% of B in this scenario whether a vaccine or drug is produced. The firm can perfectly extract the whole surplus from the U.S. market with either product because consumers in the country are homogeneous. This is a huge reduction in producer surplus compared to the 100% from perfect price discrimination. In other words, the firm does just about as well if it has to rely on the United States as its sole revenue source as it would if it served the whole world at a uniform price. But it is not too far from the producer surplus in the uniform-pricing baseline. This conclusion is particularly true in the drug market, where calibrated producer surplus ratios ρ_d^0 only differ by one percentage point between the uniform-pricing scenario and scenario with $\upsilon = 0$. One can see the basis for this result in Figure 4. In either panel, the shaded rectangle indicating equilibrium producer surplus under uniform pricing, is not much bigger than the rectangle that could be drawn under just the slice of U.S. consumers.

Relaxing the relative-pricing constraint by increasing υ causes producer surplus to asymptote to the 100% maximum for perfect price discrimination, a pattern that can be seen graphically in Figure 7. Because the United States is such a high-demand country, constraining price to be half that in the U.S. ($\upsilon = 0.5$) still allows the firm to extract 80% of B with a vaccine and 95% with a drug.

The last row of the table reports the producer-surplus ratio excluding the United States and perfectly price discriminating among the remaining countries. This strategy allows the firm to extract 63% of B. That this falls well short of 100% indicates the relative importance of the U.S. market. Comparing the strategic options available to the firm, we see that the producer would prefer excluding the United States to serving only the United States. However, as long as the reference price is $\upsilon \geq 0.2$, the producer earns more from serving than excluding the United States whether it produces a vaccine or a drug.

8. Conclusion

In most countries' healthcare markets, government programs overlay a complicated public or private insurance system. In this setting, it is difficult to assess how well an important market such as that for pharamceuticals would perform in a simpler "state of nature" in which firms sold directly to consumers on a private market. This paper attempts such an assessment combining cross-country data on disease risk and income with some simple modeling assumptions to calibrate global pharmaceutical demand. As a proof of concept, we calibrate demand for an HIV vaccine and drug. Using these calibrated demands, we can compare how lucrative the products are, bound deadweight loss from pricing and entry distortions, and simulate the welfare effects of government policies.

Our analysis abstracted not only away from involvement of governments and insurance companies but also away from epidemiological externalities and other distortions in the market. The analysis still revealed a worrisome potential for distortion in global pharmaceutical markets owing simply to the shape of pharmaceutical demand. The global demand curves for both a vaccine and a drug are Zipf similar, a shape shown in our earlier work (Kremer and Snyder 2015) to have the greatest potential for deadweight loss at both the intensive/pricing margin (i.e., the largest Harberger triangle) and the extensive/entry margin (i.e., the largest gap between private and social incentives to develop the pharmaceutical).

Our baseline calibration assumed consumers in a country share the same disease risk (equal to the prevalence of HIV in 2003, when the HIV epidemic was still growing but before antiretroviral treatments became widespread) and same income (equal to per-capital GDP) and have unit income elasticity. These simple modeling assumptions allowed us to graph the demand curves for a vaccine and a drug, inscribe the producer-surplus-maximizing rectangle underneath, derive price and quantity for the simple monopoly equilibrium we assume, and compute deadweight loss. In both the vaccine and the drug market, the monopolist ends up selling to a small fraction of high-demand consumers. Deadweight loss from above-cost pricing in the vaccine market is 41% of the social surplus B from completely relieving the disease burden and in the drug market is 49%. Potential deadweight loss from inadequate entry incentives is 56% in the vaccine market and 62% in the drug market. We proved a new proposition showing that this 62% is an exact expression for the deadweight-loss supremum in the comprehensive case in which either or both products can be developed.

We provided a series of alternative calibrations to gauge the robustness of the baseline results. The income elasticity ϵ turned out to be perhaps the key determinant of relative producer surplus. Relative producer surplus bottoms out at $\epsilon = 0.7$ for both a vaccine and a drug. At this value of ϵ, the bound on potential deadweight loss is 82% of B in the vaccine market and 65% in the drug market, indicating tremendous potential for deadweight loss. The picture is rosier using Newhouse's (1977) estimate of $\epsilon = 1.3$ based on cross-country data, in which case the bound of potential deadweight loss of 41% in the vaccine market and 48% in the drug market. Our results suggest the value of credible estimates of the income elasticity of healthcare expenditures. In future work, we plan to perform more precise calibrations accounting for heterogeneity within as well as between countries. These more precise results could be less sensitive to variations in the income elasticity.

We also ran some calibrations allowing for price discrimination. These results highlighted the pivotal role played by the United States in global pharmaceutical demand. The firm finds targeting just the United States almost as lucrative as selling to the world at a uniform price. If other countries employ reference pricing pegged to the U.S., for most reasonable scenarios it is more lucrative for the firm to serve the United States and accept the ceiling this imposes on other prices than to exclude the United States and perfectly price discriminate across the remaining countries. The U.S. Department of Commerce (2004) has complained that OECD countries behave as free riders, relying on the high prices in relatively open U.S. markets to fund innovation while enjoying low regulatory prices themselves. Our calibrations suggest that the United States would have a pivotal role whether or not other countries regulated prices, explaining in part why other countries may be emboldened to free ride.

In future work, we plan to derive a quantitative measure of a country's pivotalness for innovation incentives. We also plan to enrich the international calibration by accounting for the heterogeneity of consumers within each country. Kremer and Snyder (2016) calibrate global demand for an arbitrary product assuming a lognormal distribution of income within each country, using Pinkovskiy and Sala-i-Martin's (2009) estimates of the two lognormal parameters for each country. We are exploring the approach of calibrating global demand for a pharmaceutical by modeling disease risk and income as bivariate lognormal random variables and combining a surveys and a variety of other data sources to estimate the five requisite parameters for each country.

Appendix: Proofs

Proof of Proposition 1: For concreteness, suppose throughout the proof that

$$PS_v^0 \leq PS_d^0. \tag{13}$$

Arguments establishing the bound for the reverse inequality are similar and omitted for brevity.

A series of steps can be used to bound the supremum:

$$\begin{aligned}
\sup_{\{k_j,c_j,s_j,r_j \geq 0 | j=v,d\}} \left(\frac{DWL^*}{B}\right) &\geq \sup_{\{k_v,k_d \geq 0\}} \left(\frac{DWL^0}{B}\right) & (14)\\
&\geq \lim_{k_v \downarrow PS_v^0, k_d \uparrow \infty} \left(\frac{DWL^0}{B}\right) & (15)\\
&= \frac{B-PS_v^0}{B} & (16)\\
&= 1-\rho_v^0 & (17)\\
&= 1-\rho_{\min}^0. & (18)
\end{aligned}$$

Equation (14) follows since the right-hand side is the supremum over a smaller parameter set. Equation (15) follows since the limit point on the right-hand side is just one element of the closure of the larger set over which the supremum on the left-hand side is being taken. We defer the more detailed argument behind (16) for the moment. Equation (17) holds by definition and (18) by (13).

To verify equation (16), note that $DWL^0 = W^{00} - W^0$ by definition. But $\lim_{k_v \downarrow PS_v^0, k_d \uparrow \infty} W^0 = 0$ since nothing is produced in equilibrium in this limit. Hence

$$\begin{aligned}
\lim_{k_v \downarrow PS_v^0, k_d \uparrow \infty} DWL^0 &= \lim_{k_v \downarrow PS_v^0, k_d \uparrow \infty} W^{00} & (19)\\
&= \lim_{k_v \downarrow PS_v^0, k_d \uparrow \infty} (TS^{00} - k_v) & (20)\\
&= TS^{00} - PS_v^0 & (21)\\
&= B - PS_v^0, & (22)
\end{aligned}$$

where equation (20) holds by definition of W^{00}, (21) by evaluating the limit, and (22) by substituting from (2). *Q.E.D.*

Proof of Proposition 2: Assume benchmark parameter values. We will establish the following inequality,

$$DWL^0 \leq B - PS_{\min}^0, \tag{23}$$

holds regardless of which of the four values, $\sigma^{00} \in \{v, d, b, n\}$, the socially optimal product strategy takes on. First consider the trivial case in which $\sigma^{00} = n$. Then $DWL^0 = W^{00} - W^0 = 0$ since $W^{00} = W^0 = 0$. But then (23) trivially holds because $B - PS_{\min}^0 \geq 0 = DWL^0$.

Next, consider the non-trivial case in which $\sigma^{00} \in \{v,d,b\}$. We can establish the following series of steps:

$$\begin{aligned} DWL^0 &= W^{00}-W^0 && (24)\\ &= TS^{00}-k_{\sigma^{00}}-(\Pi^0+CS^0) && (25)\\ &\leq TS^{00}-k_{\sigma^{00}}-\Pi^0 && (26)\\ &\leq TS^{00}-k_{\sigma^{00}}-\Pi^0_{\sigma^{00}} && (27)\\ &= TS^{00}-PS^0_{\sigma^{00}}. && (28)\\ &= B-PS^0_{\sigma^{00}}. && (29)\end{aligned}$$

Equations (24) and (15) follow from substituting the definitions of the relevant variables and (26) from $CS^0 \geq 0$. Equation (27) holds because the equilibrium product strategy is the most profitable, implying $\Pi^0 \geq \Pi^0_{\sigma 00}$. Equation (28) follows from $\Pi^0_{\sigma^{00}} = PS^0_{\sigma^{00}} - k_{\sigma^{00}}$ and (29) from (2).

We will show

$$PS^0_{\sigma^{00}} \geq PS^0_{\min} \qquad (30)$$

for all $\sigma^{00} \in \{v,d,b\}$. If $\sigma^{00} = v$, then $PS^0_{\sigma^{00}} = PS^0_v \geq PS^0_{\min}$, implying (30) holds. Similar arguments show (30) holds if $\sigma^{00} = d$. Suppose $\sigma^{00} = b$. The firm can replicate producer surplus from a drug if both products have been developed by setting $p_{bv} = \infty$ and $p_{bd} = p^0_d$. Hence $PS^0_b \geq PS^0_d \geq PS^0_{\min}$, implying (30) holds, completing the proof that (30) holds for all $\sigma^{00} \in \{v,d,b\}$.

We can now complete the proof. Combining (29) and (30), we have $DWL^0 \leq B - PS^0_{\min}$ for all $\sigma^{00} \in \{v,d,b\}$ and thus for all $\sigma^{00} \in \{v,d,b,n\}$ by the first paragraph. Dividing by B,

$$\frac{DWL^0}{B} \leq 1-\rho^0_{\min} \qquad (31)$$

for all $k_v, k_d \geq 0$, implying

$$\sup_{\{k_v,k_d \geq 0\}} \left(\frac{DWL^0}{B}\right) \leq 1-\rho^0_{\min}. \qquad (32)$$

Conditions (6) and (32) sandwich the supremum between $1-\rho^0_{\min}$ above and below, yielding (7) as an exact equality. *Q.E.D.*

Proof of Proposition 3: Arguments paralleling (15)–(17) can be used to show

$$\sup_{\{k_j \geq 0\}} \left(\frac{DWL^0_j}{B}\right) \geq \lim_{k_j \downarrow PS^0_j} \left(\frac{DWL^0_j}{B}\right) = \frac{B-PS^0_j}{B} = 1-\rho^0_j. \qquad (33)$$

Arguments paralleling (24)–(29) can be used to show $DWL^0_j \leq B - PS^0_j$, or after dividing by B,

$$\frac{DWL^0_j}{B} \leq 1-\rho^0_j. \qquad (34)$$

Conditions (33) and (34) sandwich the supremum between $1-\rho^0_j$ above and below, yielding (8) as an exact equality. *Q.E.D.*

References

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

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

Arrow, Kenneth. (1962) “Economic Welfare and the Allocation of Resources for Inventions,” in Richard Nelson, ed., *The Rate and Direction of Inventive Activity*. Princeton, N.J.: Princeton University Press.

Brito, Dagobert. L., Eytan Sheshinski, and Michael D. Intrilligator. (1991) “Externalities and Compulsory Vaccination,” *Journal of Public Economics* 45: 69–90.

Budish, E., B. N. Roin, and H. Williams. (2015) “Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials,” *American Economic Review* 105: 2044–2085. no. 19430.

Chen, Frederick and Flavio Toxvaerd. (2014) “The Economics of Vaccination,” *Journal of Theoretical Biology* 363: 105–117.

Danzon, Patricia M. and Jonathan. D. Ketcham. (2004) “Reference Pricing of Pharmaceuticals for Medicare: Evidence from Germany, the Netherlands, and New Zealand,” *Frontiers in Health Policy Research* 7: 1–54.

Danzon, Patricia M., Y. Richard Wang, and Liang Wang. (2005) “The Impact of Price Regulation on the Launch Delay of New Drugs—Evidence from Twenty-Five Major Markets in the 1990s,” *Health Economics* 14: 269–292.

European Commission (2004). *The 2004 EU Industrial R&D Investment Scoreboard (Volume II: Company Data)*. Technical Report EUR 21399 EN.

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

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

Garber, Alan M., Charles I. Jones, and Paul M. Romer. (2006) “Insurance and Incentives for Medical Innovation,” *Forum for Health Economics & Policy* 9: 1–27.

Geoffard, Pierre-Yves and Tomas Philipson. (1997) “Disease Eradication: Public vs. Private Vaccination,” *American Economic Review* 87: 222–230.

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

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

Getzen, Thomas 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.

Harberger, Arnold C. (1954) "Monopoly and Resource Allocation," *American Economic Review* 44: 77–92.

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

Kremer, Michael and Christopher M. Snyder. (2013) "When Is Prevention More Profitable than Cure? The Impact of Time-Varying Consumer Heterogeneity," National Bureau of Economic Research working paper no. 18861.

Kremer, Michael and Christopher M. Snyder. (2015) "Preventives Versus Treatments," *Quarterly Journal of Economics* 130: 1167–1239.

Kremer, Michael and Christopher M. Snyder. (2016) "Worst-Case Bounds on R&D and Pricing Distortions: Theory and Disturbing Conclusions if Consumer Values Follow the World Income Distribution," Harvard University working paper.

Kremer, Michael, Christopher M. Snyder, and Heidi L. Williams. (2012) "Optimal Subsidies for Prevention of Infectious Disease," Harvard University working paper.

Kyle, Margaret K. (2007) "Pharmaceutical Price Controls and Entry Strategies," *Review of Economics and Statistics* 89: 88–99.

Lakdawalla, Darius and Neeraj Sood. (2013) "Health Insurance as a Two-Part Pricing Contract," *Journal of Public Economics* 102: 1–12.

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

Newhouse, Joseph. P. (1977) "Medical Care Expenditure: A Cross-National Survey," *Journal of Human Resources* 12: 115–125.

Pinkovskiy, Maxim and Xavier Sala-i-Martin. (2009) "Parametric Estimations of the World Distribution of Income," National Bureau of Economic Research working paper no. 15433.

Sood, Neeraj, Han de Vries, Italo Gutierrez, Darius N. Lakdawalla, and Dana P. Goldman. (2008) "The Effect of Regulation on Pharmaceutical Revenues: Experience in Nineteen Countries," *Health Affairs* 28: w125–w137.

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

UNAIDS. (2004a) *2004 Report on the Global AIDS Epidemic*. Geneva: Joint United Nations Programme on HIV/AIDS.

UNAIDS. (2004b) *Epidemiological Fact Sheets on HIV/AIDS and Sexually Transmitted Infections—2004 Update*. Geneva: Joint United Nations Programme on HIV/AIDS.

UNAIDS. (2006) *2006 Report on the Global AIDS Epidemic: A UNAIDS 10th Anniversary Special Edition*. Geneva: Joint United Nations Programme on HIV/AIDS.

U.S. Department of Commerce. (2004) *Pharmaceutical Price Controls in OECD Countries: Implications for U.S. Consumers, Pricing, Research and Development, and Innovation*. Washington, D.C.

Tables and Figures

Table 1: Toy Example Illustrating Calibration Method

i	N_i	X_i	Y_i	N_iX_i	X_iY_i	$N_iX_iY_i$
1	1,000	0.01	10,000	10	100	100,000
2	1,500	0.05	9,000	75	400	675,000

Note: Fictitious data for a toy example of the calibration method.

Table 2: Descriptive Statistics for Cross-Country Data

Variable	Notation	Mean	Std. dev.	Min.	Max.
Population (million)	N_i	38.9	138.5	0.3	1,288.4
HIV prevalence	X_i	0.014	0.032	0.000	0.202
GDP per capita	Y_i	7,653	12,375	106	64,670

Notes: Entries are descriptive statistics for 2003 data for sample of 158 countries. We compute HIV prevalence by dividing "Estimated number of people living with HIV, adults and children, end 2003" from UNAIDS (2004), by population. GDP per capita from "GDP per capita (current US$)" entry of World Bank Open Data, downloaded May 10, 2017 from http://data.worldbank.org/indicator/NY.GDP.PCAP.CD. Population from "Population, total" entry of World Bank Open Data, downloaded May 10, 2017 from http://data.worldbank.org/indicator/SP.POP.TOTL.

Table 3: Correlations in Cross-Country Data

	N_i	X_i	Y_i
N_i	1.00		
X_i	–0.07	1.00	
Y_i	–0.04	–0.19	1.00

Table 4: Baseline Calibration Results

Description	Formula	Result
Vaccine alone		
Price	p_v^0	\$130
Quantity	q_v^0	344 million
Consumer surplus	$\frac{CS_v^0}{B} = \gamma_v^0$	0.16
Producer surplus	$\frac{PS_v^0}{B} = \rho_v^0$	0.44
Harberger DWL	$\frac{HDWL_v^0}{B} = 1 - \gamma_v^0 - \rho_v^0$	0.41
DWL supremum	$\sup\left(\frac{DWL_v^0}{B}\right) = 1 - \rho_v^0$	0.56
Zipf similarity	$Z_v^0 = \frac{1-\rho_v^0}{1-\underline{\rho}(B)}$	0.66
Drug alone		
Price	p_d^0	\$27,387
Quantity	q_d^0	1.4 million
Consumer surplus	$\frac{CS_d^0}{B} = \gamma_d^0$	0.13
Producer surplus	$\frac{PS_d^0}{B} = \rho_d^0$	0.38
Harberger DWL	$\frac{HDWL_d^0}{B} = 1 - \gamma_d^0 - \rho_d^*$	0.49
DWL supremum	$\sup\left(\frac{DWL_d^0}{B}\right) = 1 - \rho_v^0$	0.62
Zipf similarity	$Z_d^0 = \frac{1-\rho_d^0}{1-\underline{\rho}(B)}$	0.74
Both products possible		
DWL supremum	$\sup\left(\frac{DWL^0}{B}\right) = 1 - \rho_{\text{min}}^0$	0.62

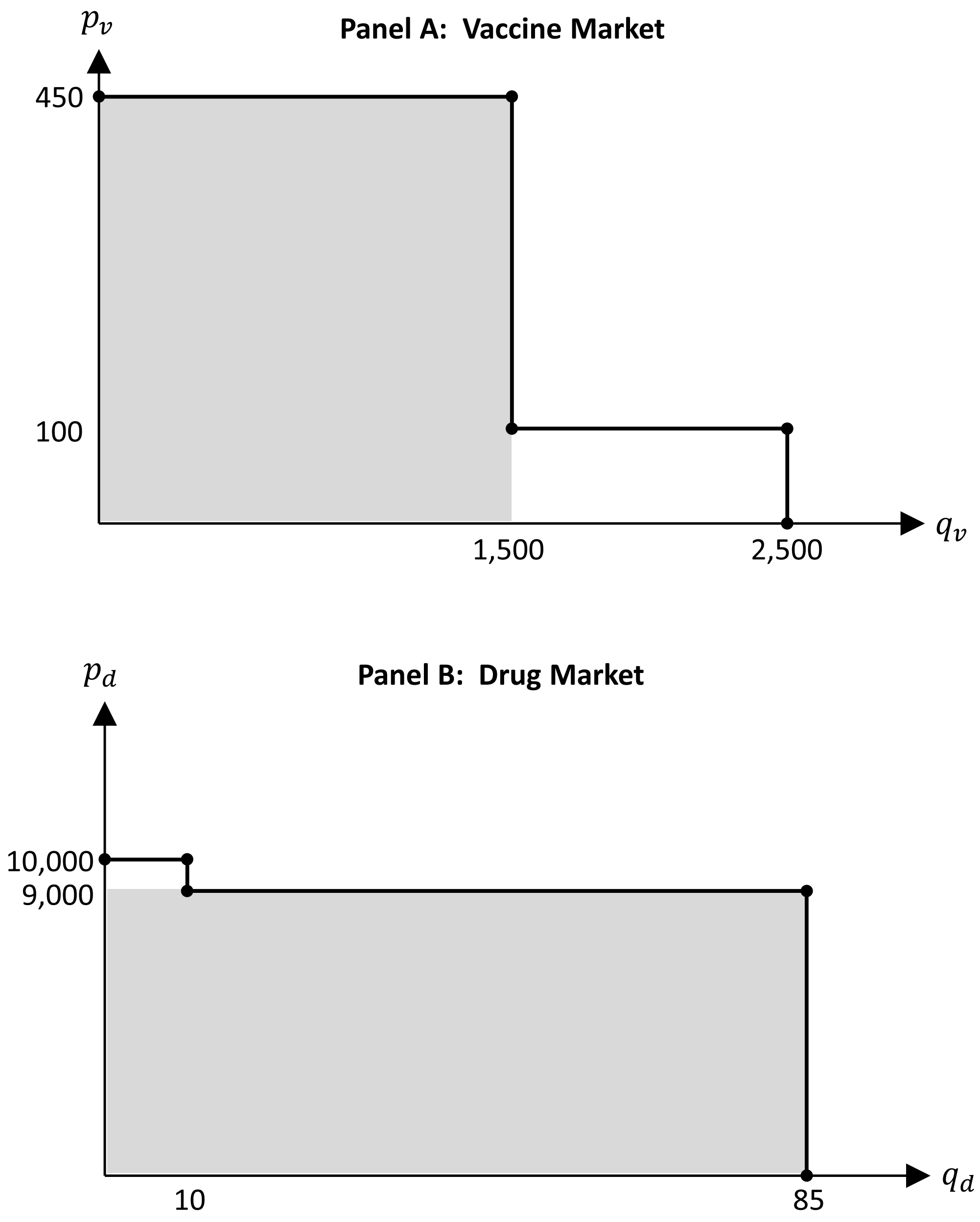
Notes: Baseline calibration in which consumers are heterogeneous in both disease risk (X_i) and income (Y_i), have unit income elasticity, and are willing to pay up to one year's income to avoid disease harm. All surpluses expressed as proportion of first-best social surplus (equivalently, as proportion of disease burden). The suprema are taken over $k_v, k_d \geq 0$.

Table 5: Calibrating Producer-Surplus Ratio in Alternative Scenarios

Scenario	Vaccine market ρ_v^0	Drug market ρ_d^0
Uniform pricing scenarios		
• Baseline	0.44	0.38
• Just disease-risk heterogeneity	0.30	1.00
• Just income heterogeneity	0.57	0.57
• Revised disease-risk data	0.48	0.42
• Varying income elasticity		
$\epsilon = 0.0$	0.30	1.00
$\epsilon = 0.5$	0.28	0.40
$\epsilon = 1.0$	0.44	0.38
$\epsilon = 1.5$	0.67	0.59
$\epsilon = 2.0$	0.69	0.70
Price-discrimination scenarios		
• Perfect price discrimination	1.00	1.00
• Price ceiling tied to varying proportion of U.S. price		
$\upsilon = 0.0$	0.37	0.37
$\upsilon = 0.5$	0.80	0.95
$\upsilon = 1.0$	0.84	1.00
$\upsilon = 1.5$	0.88	1.00
$\upsilon = 2.0$	0.91	1.00
• Excluding U.S. to evade price ceiling	0.63	0.63

Notes: All entries are producer surplus as proportion of first-best social surplus (equivalently, as proportion of disease burden). Scenarios return to baseline in all dimensions except for stated alternative. In particular, the price-discrimination scenarios have the same parameters and functional form as the baseline, just changing the pricing strategy from uniform to one allowing prices to differ across countries.

Figure 1: Demand Curves in Toy Example

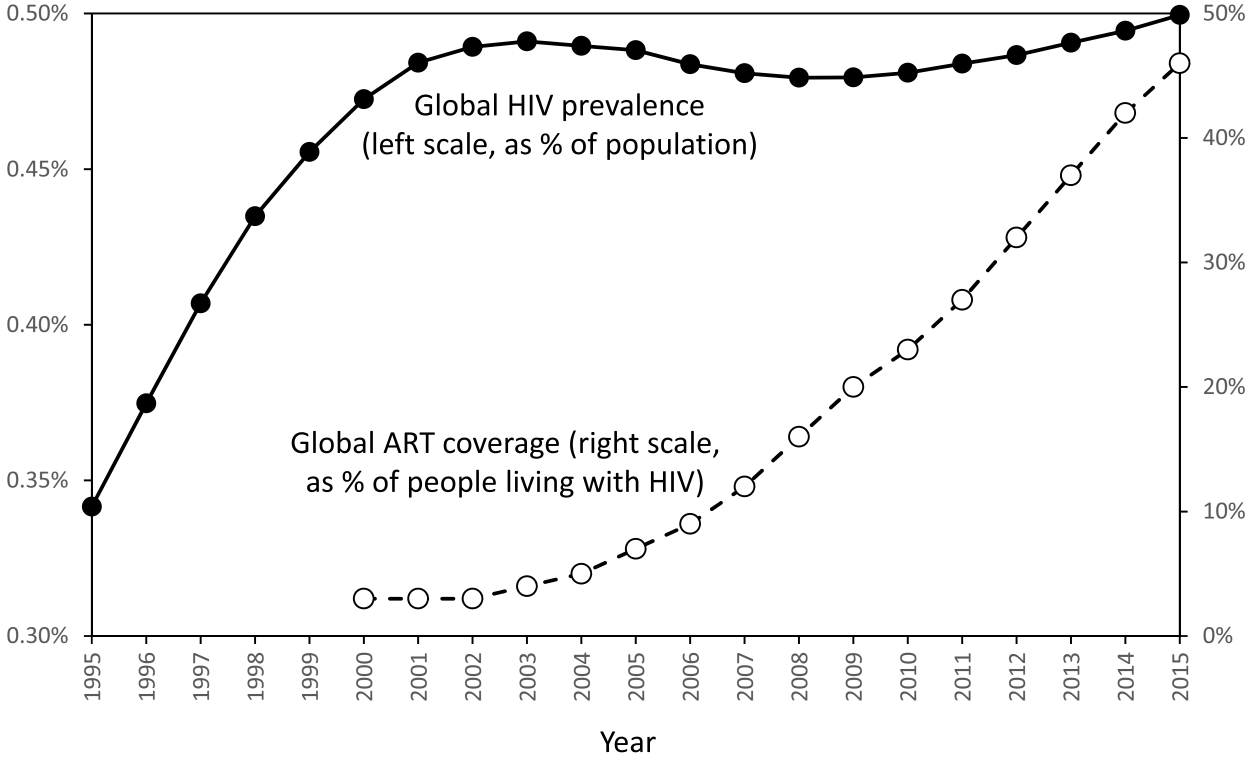


Panel A: Vaccine Market

Panel B: Drug Market

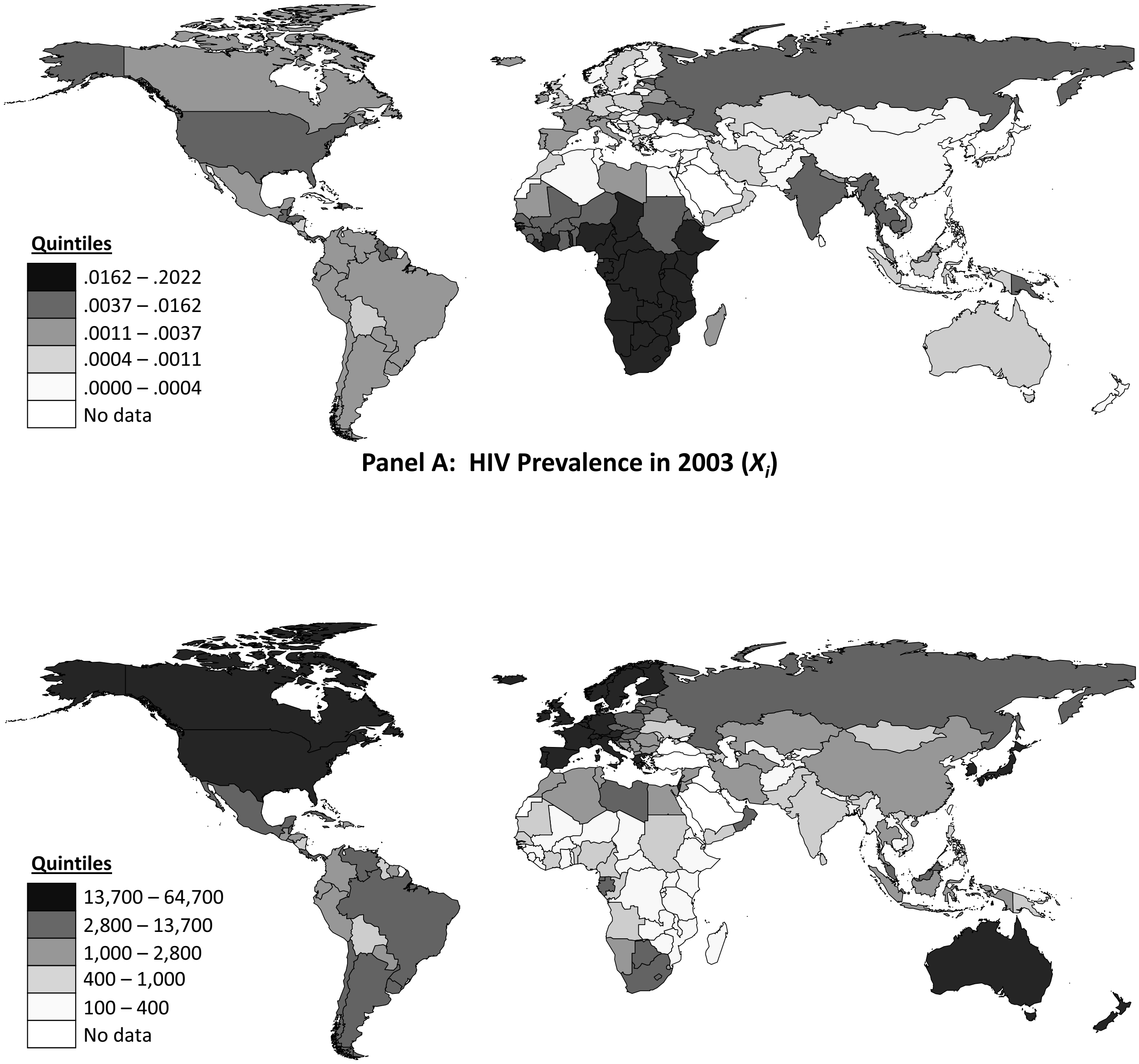
Notes: Pharmaceutical demands derived from fictitious data in Table 1. Axes scaled so that a unit of area represents the same producer surplus in both panels. Producer surplus equal to area of shaded region.

Figure 2: Trends in Global HIV Prevalence and Treatment



Sources: Left-axis variable equals HIV cases (from UNAIDS data on "Number of people living with HIV," downloaded from http://aidsinfo.unaids.org) as a percent of population (from World Bank data on "Population, total," downloaded from http://data.worldbank.org/indicator/SP.POP.TOTL). Right-axis variable from UNAIDS data reported by the World Bank in "Antiretroviral therapy coverage (% of people living with HIV)," downloaded from http://data.worldbank.org/indicator/SH.HIV.ARTC.ZS. All downloads on May 8, 2017.

Figure 3: Geographical Distribution of Main Variables

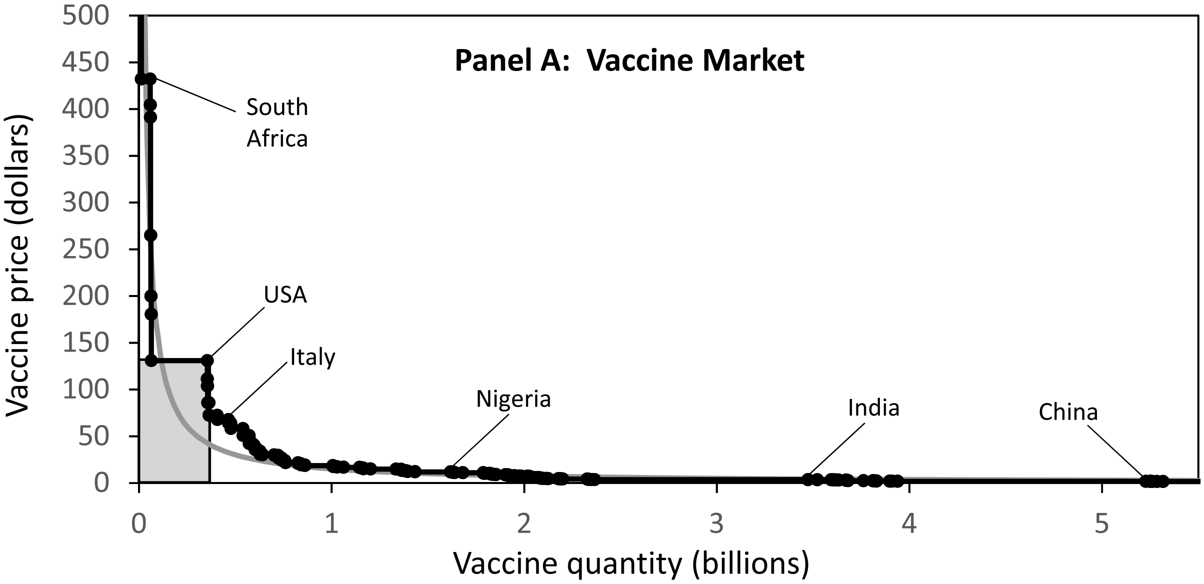


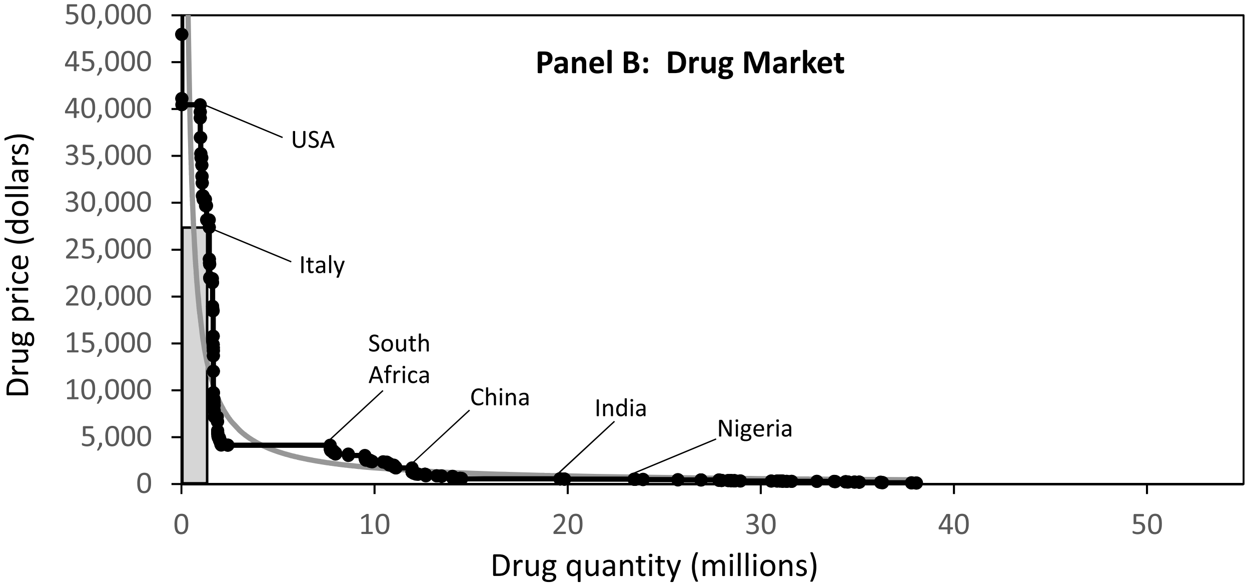
Panel A: HIV Prevalence in 2003 (X_i)

Panel B: GDP Per Capita in 2003 (Y_i)

Sources: See Table 2.

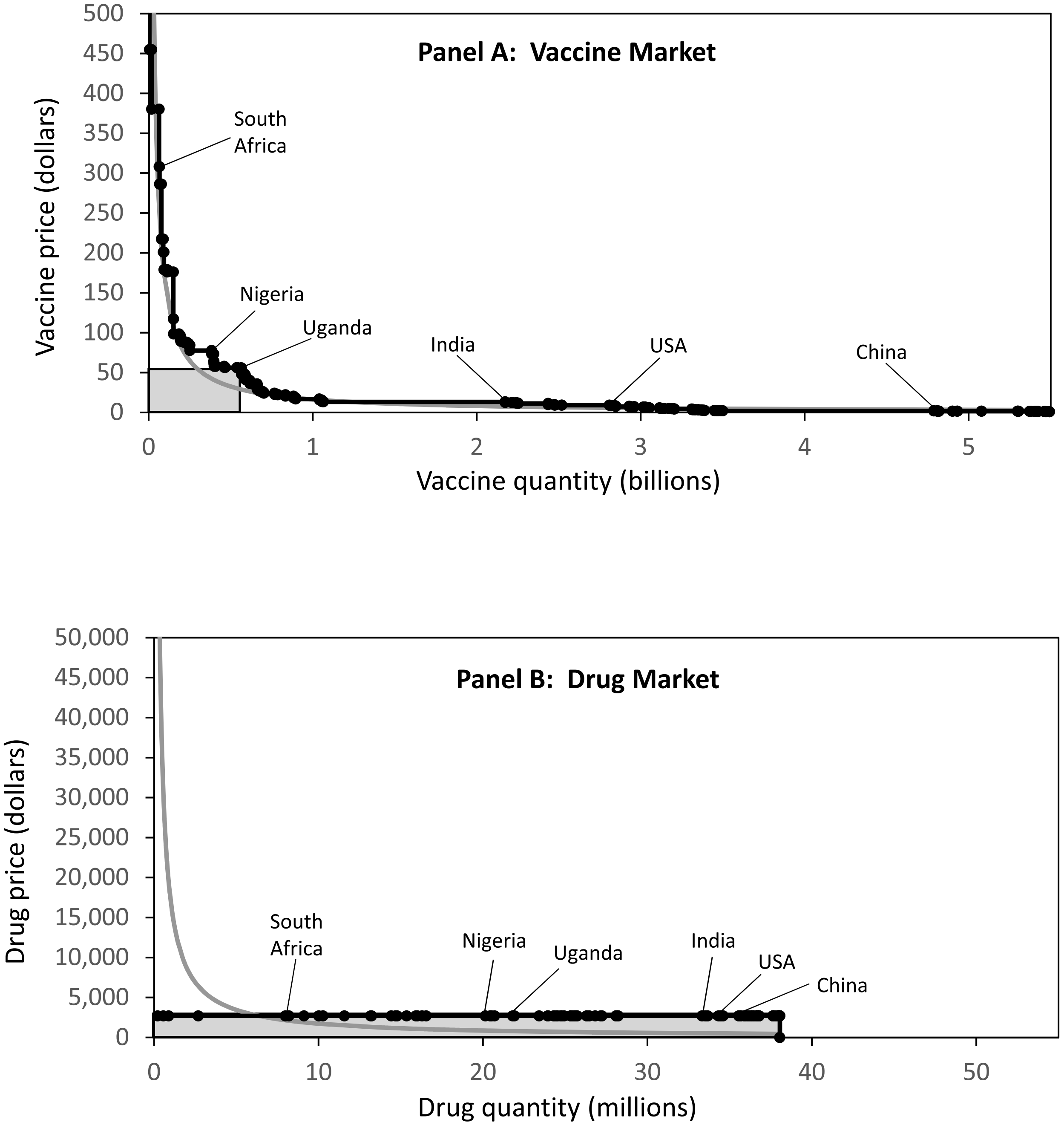
Figure 4: Demand Curves for HIV Pharmaceuticals in Baseline Calibration





Notes: Baseline calibration, assuming heterogeneity in both disease risk (X_i) and income (Y_i) and income elasticity $\epsilon = 1$. Axes scaled so that a unit of area represents the same producer surplus in both panels. Both axes truncated to aid visualization. Shaded rectangle represents equilibrium pricing decision; its area equals producer surplus. Grey line is STRZ demand curve with same area underneath (and thus same total surplus) as calibrated demand.

Figure 5: Demand Curves for HIV Pharmaceuticals in Calibration with Heterogeneity in Disease Risk Alone



Notes: Calibration with heterogeneity in disease risk (X_i) alone. As in previous figure, income elasticity is $\epsilon = 1$. Income set to the constant $\bar{Y} = \sum_{i=1}^{I} N_i X_i Y_i / \sum_{i=1}^{I} N_i X_i$ such that the social surplus B from completely relieving the disease burden is same as in the previous figure. One can show $\bar{Y}$ is the weighted harmonic mean of income, where country i's weight is its share of the world disease burden, $B_i / \sum_{k=1}^{I} B_k$. Axes have same scales as previous figure; as there, a unit of area represents the same producer surplus in both panels. Both axes truncated to aid visualization. Shaded rectangle represents equilibrium pricing decision; its area equals producer surplus. Grey line is STRZ demand curve with same area underneath (and thus same total surplus) as calibrated demand.

Figure 6: Effect of Income Elasticity on Producer-Surplus Ratio

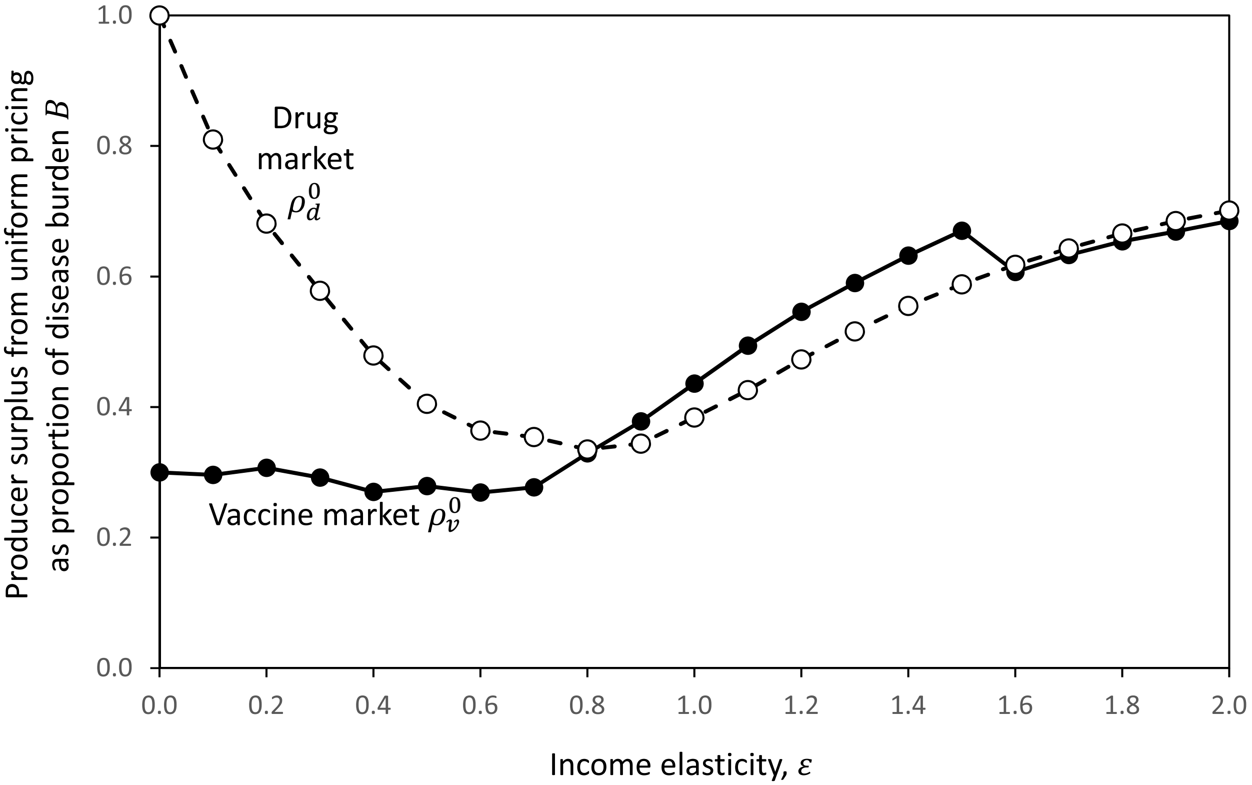
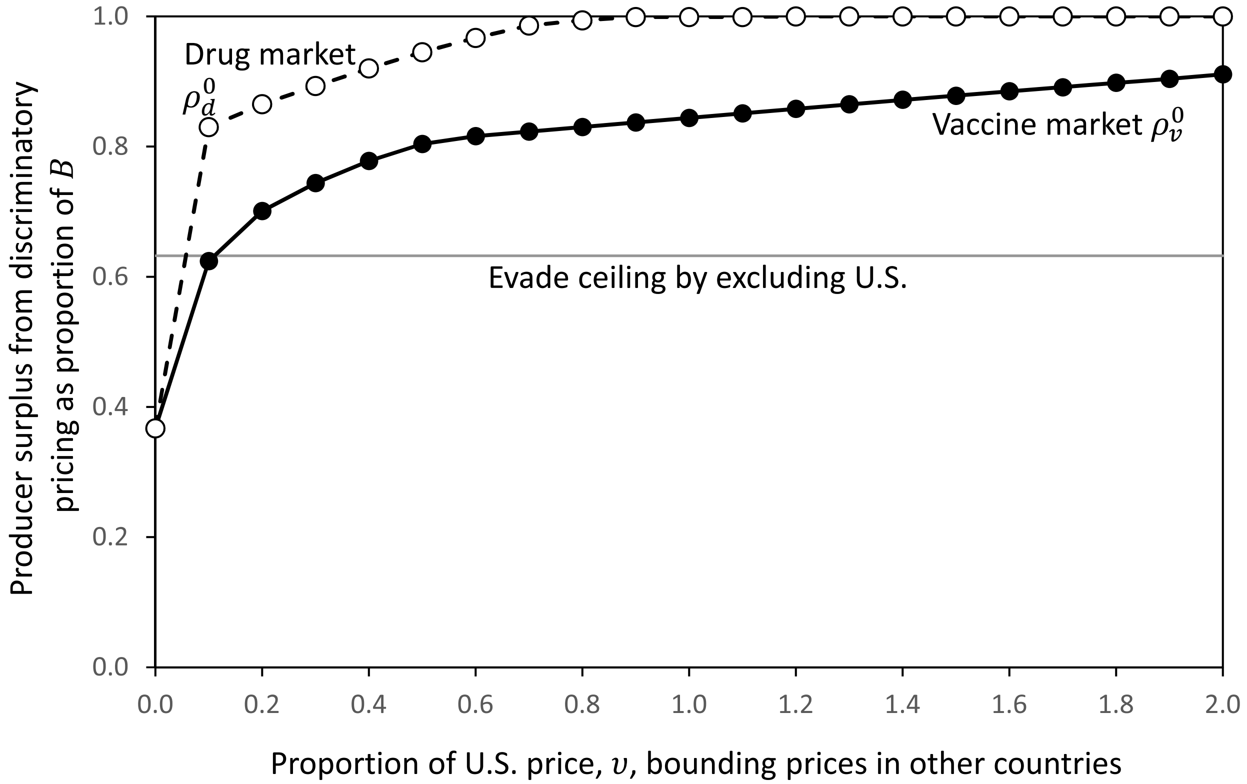


Figure 7: Effect of Reference Pricing Pegged to the United States on Producer-Surplus Ratio


Drug market
ρ_d^0
Vaccine market ρ_v^0
Evade ceiling by excluding U.S.
Producer surplus from discriminatory pricing as proportion of B
Proportion of U.S. price, υ, bounding prices in other countries