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

February 2018

Abstract: Kremer and Snyder (2015) show that demand curves for a preventive and treatment, though they target the same disease, may have different shapes and show that this difference may bias the pharmaceutical manufacturer toward developing the lucrative rather than the socially desirable product. This paper tightens the theoretical bounds on the potential deadweight loss from such biases. Using a calibration of the global demand for HIV pharmaceuticals, we demonstrate the dramatically sharper analysis that is achievable with the new bounds, which allows us to pinpoint potential deadweight loss at 62% of the global gain from curing HIV. We use the calibration to assess the welfare effects of counterfactual policies such as: subsidies, reference pricing, and a price-discrimination ban. The fit of our calibration is good: We find that a hypothetical drug monopolist would price an HIV drug so high that only 4% of the infected population worldwide would purchase, matching actual drug prices and quantities in the early 2000s before subsidies in low-income countries ramped up.

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 general editor, Lawrence White, the special-issue editor, Victor Tremblay, and anonymous reviewers for suggestions that substantially improved the paper; to Margaret Kyle, Paul Novosad, Douglas Staiger, Robert Staiger, Heidi Williams, and seminar participants at Penn State 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

In previous work (Kremer and Snyder, 2015), we argued that a drug may be more lucrative than a vaccine, even when both products target the same disease and thus have the same health benefit—even in the absence of epidemiological externalities. The argument is that a vaccine is sold before consumers contract the disease, when consumers differ considerably in disease risk, which prevents the firm from easily extracting much of their surplus with a uniform price. On the other hand, a drug is sold after consumers have contracted the disease, when they no longer differ in disease risk, which allows the manufacturer to extract more surplus from these relatively homogeneous consumers. The drug may thus end up being more lucrative—and the manufacturer biased toward developing it over the vaccine—even if the vaccine is substantially more effective or lower cost.[1]

In that paper we calibrated the distribution of infection risk in the U.S. population for human immunodeficiency virus (HIV) and found it to be close to Zipf: a power-law distribution with exponent equal to 1. A Zipf distribution of disease risk intuitively means that each doubling of risk cuts the number of consumers with at least that risk in half, which leads to an iso-revenue property. In view of our theoretical result that (owing to this iso-revenue property) the Zipf risk distribution generates the worst bias against preventives, our previous work provided one explanation for why a variety of HIV drugs have been developed but as yet no vaccine has appeared.

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) and considering a range of values of the income elasticity of healthcare expenditures for robustness. For consistency, throughout most of the analysis we interpret calibrated world demand in a similar way as Kremer and Snyder (2015) did U.S. demand: We take demand to represent individual purchase decisions in a (perhaps counterfactual) private market in the absence of intervention by governments or insurance companies and in the absence of epidemiological externalities.[2] How-

[1] Kremer and Snyder (2015) show that this bias may be reversed when income (or more generally willingness to pay) covaries sufficiently negatively with disease risk (see their Proposition 18). The bias against vaccines described in the footnoted paragraph arises in a setting with little or no income variation (covered by their Proposition 3) or in a setting with independent income and disease-risk distributions (covered by their Proposition 16).

[2] For theoretical analyses of vaccine markets in the presence of epidemiological externalities, see Brito, Sheshinski, and Intrilligator (1991); Francis (1997); Geoffard and Philipson (1997); Gersovitz (2003); Gersovitz and Hammer

ever, the calibrations can be interpreted more broadly. We offer an alternative interpretation of world demand as reflecting purchases by national agencies on behalf of citizens in their health systems. This interpretation allows for consumer heterogeneity and epidemiological externalities within countries.

Using the calibrated global demand for an HIV drug, we show that—in the absence of international price discrimination, subsidies in low-income countries, or regulatory threats to intellectual property—a profit-maximizing monopolist would charge such a high drug price that only about 4% of infected individuals worldwide would buy the treatment. These estimates are remarkably close to the actual prices and quantities of antiretroviral therapies (ARTs) in 2003, the year we focus on in the subsequent analysis as arguably approximating the "state of nature" before subsidies or other policy interventions became widespread. We estimate that the deadweight loss from charging such a high price for an existing product (the area of the Harberger (1954) triangle) amounts to about half of the global benefit from curing HIV.

The resulting calibrated demands are employed to examine a range of policy issues: One exercise leverages Kremer and Snyder's (2015) formulas for worst-case bounds on deadweight loss from pharmaceutical sales in the private market. Since the greatest conceivable distortions occur at the extensive margin (whether a product is developed at all) rather than the intensive margin (how high the mark up is for an existing product, which generates a Harberger deadweight-loss triangle), 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, which we denote by 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 (v) or a drug (d). In our baseline calibration (indicated by superscript 0), we obtain a producer-surplus ratio of $\rho_v^0 = 44\%$ for an HIV vaccine. Considering the market for an HIV *vaccine* in isolation, we compute a worst-case bound on deadweight loss of $100-44 = 56\%$. In other words, the distorted incentives provided by the private market to the firm with regard to the vaccine's development and price could dissipate as much as 56% of the global benefit from curing HIV. Considering the market for an HIV *drug* in isolation, we compute a baseline estimate of the drug-producer-surplus

(2004, 2005); Chen and Toxvaerd (2014); as well as our own work (Kremer, Snyder, and Williams 2012).

ratio of $\rho_d^0 = 38\%$, which leads to a worst-case bound on deadweight loss 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 theoretical result relevant to this comprehensive case is Proposition 15 of Kremer and Snyder (2015). We have been able to tighten the bounds, as reported in Proposition 3 of this paper. 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; using the numbers from the calibration, $\max(56, 62) = 62\%$.

The previous result merely told us that worst-case deadweight loss in the comprehensive case is no less than the difference between the worst cases for the individual products: 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 pinpoint it at 62%.

The fit of our demand calibration appears to be good. Based on the calibrated global demand for an HIV drug, we calculate that—in the absence of international price discrimination, subsidies in low-income countries, or regulatory threats to intellectual property—a profit-maximizing monopolist would charge such a high drug price that only about 4% of infected individuals worldwide would buy the treatment. These price and quantity estimates are remarkably close to the actual prices and quantities of antiretroviral therapies (ARTs) in 2003—the baseline year for our analysis—which arguably approximates the "state of nature" before subsidies or other policy interventions became widespread.

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 key parameter is the income elasticity of healthcare expenditure: ϵ. We first consider the baseline case of unit income elasticity, which is close to leading estimates that use international data: e.g., Newhouse (1977) estimates $\epsilon = 1.3$. We provide calibrations for a wide range of ϵ that span 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, deadweight loss reaches 70% of the overall disease burden in the vaccine market.

We also use the calibration to assess the welfare effects of counterfactual policies such as: subsidies, reference pricing, and a price-discrimination ban. With regard to subsidies, 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.

With regard to reference pricing, if all other countries peg their prices to some proportion, u, of the U.S. price, a monopoly manufacturer of either a vaccine or a drug would prefer to serve the U.S. even if the ceiling were as low as $u = 0.2$. This result highlights the pivotal U.S. role in the pharmaceutical market and explains why other countries may be emboldened to free ride.

With regard to 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: it could lead to as much as a 56% decline in vaccine producer surplus and a 62% decline in drug producer surplus; these declines equal 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 generalizes the bounds on deadweight loss to product markets beyond pharmaceuticals.[4]

A detailed discussion of the connection between this line of our research and other authors' work 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

[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 that was 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 of the theoretical results are new developments.

[4]Kremer and Snyder (2016) also include calibrations of international demand. Since they analyze general product markets, their calibrations include only income—not disease risk. Unlike the present calibrations, that paper accounts for within-country heterogeneity by allowing each country to have a different lognormal distribution of income.

(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 the distortion in our model is due to under-consumption 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: The next section presents the model. Section 3 provides propositions that tighten the bound on deadweight loss from our previous work. The rest of the paper focuses on the calibrations. Section 4 describes the methods and Section (5) the data that are used in the calibrations. Section 6 presents the results from calibrations for the baseline scenario and Section 7 results for scenarios that involve price discrimination. Section 8 presents calibrations for alternative parameterizations to explore robustness. Section 9 concludes. Proofs of propositions are provided in the appendix.

2. Model

We base our analysis on the most general model in Kremer and Snyder (2015). This model, which appeared 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 defined below. Second, it allows for multiple sources of consumer heterogeneity that are embodied in the following random variables: $X \in [0,1]$, which represents disease risk; $Y \geq 0$, which represents willingness to pay to avoid a unit of disease harm (for simplicity, this can be thought of as income or wealth); and $H \geq 0$, which represents the consumer's draw of disease harm conditional on contracting it, where H has unit mean and is mean-independent of X and Y: $E(H) = E(H|Y = y) = E(H|X = x) = 1$.

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 and N be the mass of consumers, we have $B = NE(XYH) = NE(XY)$, where the last equality follows from the assumptions on H. Proposition 14 in Kremer and Snyder (2015) shows that the three random variables X, Y, and H are sufficient to capture quite general models of consumer heterogeneity.

A monopoly pharmaceutical manufacturer can develop two products, which are indexed by j, to sell in the private market to consumers: a vaccine ($j = v$) and/or a drug ($j = d$). If the firm chooses to develop product j, it must pay a fixed cost $k_j \geq 0$, which reflects 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).

For consistency with Kremer and Snyder (2015), throughout most of the discussion we maintain the interpretation of demand as reflecting the individual purchase decisions on a private market in the absence of government intervention, insurance, or epidemiological externalities.[5] There is an alternative interpretation of demand that allows these factors to be integrated into the analysis. Rather than assuming that consumers are homogeneous within each country, one can assume that national governments (or large national insurers) decide whether to order a pharmaceutical on behalf of the heterogeneous citizens who are enrolled their health systems. This interpretation handles epidemiological externalities under the assumption that they mostly flow within country borders and would be internalized by the national agent. We return to a discussion of this alternative interpretation in the Conclusion, with the caveats that are required for it to apply.

The key distortion in the model is not a government intervention, externality, or consumer behavioral bias but rather the problem of surplus appropriability. 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

[5] We do not deny the importance of epidemiological externalities—indeed some of our other work (Kremer, Snyder, and Williams 2012) focuses exclusively on such externalities—but want to focus on other distortions in this paper. Epidemiological externalities can be shut down as a source of distortion by assuming that the pharmaceuticals, while preventing individuals from experiencing disease symptoms, do not slow transmission of an infectious disease. An alternative way to shut down epidemiological externalities would be to consider non-infectious conditions such as heart attacks. The alternative interpretation of demand reflecting purchases by a national agent discussed next can incorporate some forms of epidemiological externality.

resolved into binary disease status (infected or not). This difference can generate different shapes for the pharmaceuticals' demand curves, which lead 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. Naturally, the producer will *prefer* the product 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 that lower values are better.

Our analysis focuses on the wedge between private and social innovation incentives. This wedge is driven not by simple differences between, say, the two products' fixed or production costs but by more subtle factors related to surplus appropriability.

We proceed with the specification of the model by introducing notation for the firm's decisions: Let σ denote the firm's product-development strategy. It has four possibilities: produce the vaccine alone ($\sigma = v$), produce the drug alone ($\sigma = d$), produce both products ($\sigma = b$), or produce neither product ($\sigma = n$).

Consider the continuation game that follows when the firm has developed exactly one of the products: $\sigma = j$ 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, but arguments besides p_j are suppressed in the notation for brevity.) 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 that 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 by $\Pi_j(p_j) = PS_j(p_j) - k_j$. Ex ante social welfare is denoted by $W_j(p_j) = TS_j(p_j) - k_j$.

Next, consider the continuation game that follows when the firm has developed both products: $\sigma = b$. Let p_{bv} denote the price that 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 that follows when the firm has developed no product: $\sigma = n$. Nothing can be sold if nothing has been developed. Thus, for all $p_n \geq 0$, $Q_n(p_n) = 0$, which implies

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

Equilibrium values are denoted with a single star. For example, $p_\sigma^* = \text{argmax}_p PS_\sigma(p)$ denotes the monopoly price; $q_\sigma^* = Q_\sigma(p_\sigma^*)$ denotes the monopoly quantity; $PS_\sigma^* = PS_\sigma(p_\sigma^*)$ denotes the monopoly producer surplus; W_σ^* denotes the 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 that are 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: $p^* = p_{\sigma^*}^*$; $q^* = q_{\sigma^*}^*$; $PS^* = PS_{\sigma^*}^*$; $W^* = W_{\sigma^*}^*$; 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: $p^{**} = p_{\sigma^{**}}^{**}$; $q^{**} = q_{\sigma^{**}}^{**}$; $PS^{**} = PS_{\sigma^{**}}^{**}$; $W^{**} = W_{\sigma^{**}}^{**}$; 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: They allow us to "level the playing field" for the two products in all respects save for the timing of when they are sold.[6] To denote equilibrium under benchmark parameters, we replace the star with a zero in the superscript.

[6] For example, if we allow 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).

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 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 central results of this section are a series of propositions that provide 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—which captures the deadweight loss that stems from prices above marginal cost—as Harberger deadweight loss. We denote it as $HDWL_\sigma^*$, 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 B 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 equilib-

rium product strategy σ^*.[7]

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

$$\frac{HDWL^*_\sigma}{B} = \frac{TS^{**}_\sigma}{B} - \frac{PS^*_\sigma}{B} - \frac{CS^*_\sigma}{B} = \frac{TS^{**}_\sigma}{B} - \rho^*_\sigma - \gamma^*_\sigma, \tag{3}$$

where $\rho^*_\sigma = PS^*_\sigma/B$ is relative producer surplus and $\gamma^*_\sigma = CS^*_\sigma/B$ is relative consumer surplus. When the firm pursues the equilibrium product strategy σ^*, we know 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 $\gamma^* = CS^*/B$. When we assume 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, which is reported in their Proposition 15.

In this subsection, we provide a series of propositions that tighten that bound. The series of propositions apply to increasingly rich environments. The first proposition assumes benchmark parameters and considers the market for one product j in isolation. The second proposition allows for multiple products to be developed. The third proposition generalizes the parameters beyond benchmark values.[8]

[7] 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.

[8] To trace out the precise connection between the series of propositions provided here and our past results, Propo-

Under benchmark parameters, we can derive exact expressions for the supremum on relative deadweight loss. The first proposition restricts attention to the market for a single product j in isolation, putting aside the possibility of developing the other product alone or both. The worst case for deadweight loss arises when k_j exceeds PS_j^0. The firm develops nothing in equilibrium, so the whole of first-best social welfare constitutes deadweight loss. In the limit as k_j approaches PS_j^0 from above, deadweight loss approaches $TS_j^{00} - PS_j^0 = B - PS_j^0$. Dividing by B to express in relative terms gives the expression for the supremum in the next proposition. A formal proof is provided in the appendix.

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

$$\sup_{\{k_j \geq 0\}} \left(\frac{DWL^0}{B} \right) = 1 - \rho_j^0. \tag{6}$$

The next proposition allows for the development of a second product, which requires some additional notation. 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 $PS^0_{\min}$, $PS^0_{\max}$, $\rho^0_{\min}$, and $\rho^0_{\max}$ be the corresponding variables in equilibrium under benchmark parameters.

The proof (which is provided in the appendix) establishes that the greatest distortion again arises when nothing is produced because development costs are too high. The whole of first-best social welfare again constitutes deadweight loss. The supremum on deadweight loss is approached in the limit as the development cost in the market for the less lucrative product approaches producer surplus in that market from above (keeping the development cost above producer surplus in the other market). Suppose for concreteness that the vaccine is the less lucrative of the two products. Then the supremum on deadweight loss is approached in the limit $k_v \downarrow PS^0_{\min}$ for k_d sufficiently high (taking the limit $k_d \uparrow \infty$ suffices). Expressed relative to B, this supremum equals $1 - \rho_v^0$. More

sition 1 here superficially resembles Proposition 2 of Kremer and Snyder (2015) but they are subtly different. The previous result applied to the case in which both products could be produced but there is heterogeneity in X alone. Proposition 1 here allows for heterogeneity in both X and Y but assumes only product j can be produced. In fact, Proposition 1 here is a corollary of Theorem 1 of Kremer and Snyder (2016) for the special case of benchmark parameters. The translation of that result into the present context is somewhat involved, so instead we provide a direct proof here. Propositions 2 and 3 are new results. The assumptions behind Proposition 3 are identical to those behind Proposition 15 of Kremer and Snyder (2015), so the results are directly comparable. Proposition 3 tightens the previous bound.

generally, the supremum is associated with whichever of the two isolated product markets is less lucrative. In such a market, a low development cost can generate a large equilibrium distortion without dissipating too much first-best social welfare, leaving the greatest potential for deadweight loss. The supremum on deadweight loss when there are two products equals the greater of the suprema in equation (6) for the two isolated markets: $\max(1-\rho_v^0, 1-\rho_d^0) = 1-\min(\rho_v^0, \rho_d^0) = 1-\rho_{\min}^0$.

Proposition 2. *Consider a model of the pharmaceutical market with multiple sources of consumer heterogeneity, still fixing the parameters at their benchmark values $c_j = s_j = r_j = 0$ but allowing the firm to produce either or both of the products. The supremum on relative deadweight loss is exactly $1-\rho_{\min}^0$:*

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

The final proposition generalizes the parameters beyond benchmark values. The proof is immediate from the previous proposition. The supremum when parameters are restricted to benchmark values is weakly lower than the supremum when parameters are allowed to be free. Thus the supremum for benchmark parameters from the previous proposition is a lower bound on the supremum over general parameters in the next proposition.

Proposition 3. *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 $1-\rho_{\min}^0$:*

$$\sup_{\{k_j,c_j,s_j,r_j\geq 0 | j=v,d\}} \left(\frac{DWL^*}{B}\right) \geq 1-\rho_{\min}^0. \tag{8}$$

This bound is directly comparable to the bound that we derived for the case of multiple products and general parameters in our earlier work (Kremer and Snyder, 2015, Proposition 15). The earlier result stated that the supremum on relative deadweight loss is no less than $\rho_{\max}^0 - \rho_{\min}^0$. We derived that bound by focusing on just one possible distortion: arising from the firm's developing the "wrong" product, the less socially efficient but more lucrative one. The supremum on that distortion equals $\rho_{\max}^0 - \rho_{\min}^0$ in relative terms. The new bound focuses on a potentially larger source of deadweight loss. Society stands to lose more when the firm develops not just a different product but nothing in place of the socially efficient product. The new bound $1-\rho_{\min}^0$ reflects the larger gap between no product and the efficient one.

It is obvious that the new bound is weakly tighter 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 neither the vaccine nor the drug is particularly lucrative: $\rho^0_v \approx \rho^0_d \approx 0$. Using the old bound, $\rho^0_{\max} - \rho^0_{\min} \approx 0$, we arrive at the tautological conclusion that the supremum on deadweight must be positive. The new bound, $1-\rho^0_{\min} \approx 1$, raises the real prospect that nearly the whole disease burden can be dissipated in this scenario. Just such a scenario will arise in the calibrations.

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, which implies that all feasible prices generate the same revenue, and hence the same producer surplus under the 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 6; they can be visualized as rectangular hyperbolas truncated at the top and bottom. The truncations are technical features that keep 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: They provide the necessary adjustments to apply this formula to general settings with arbitrary scaling of the price and quantity axes. When we adapt the formula from equation (16) of Kremer and Snyder (2016) to reflect the present notation, we find the Zipf similarity of the demand

for pharmaceutical j under benchmark parameters, Z_j^0, equals

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

where $\underline{\rho}(\mu_j^0)$ is the producer-surplus ratio of the associated STRZ demand. By Proposition 1, 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 that the demand for pharmaceutical j is 50% Zipf similar. For the supremum on deadweight loss to be 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, which is equal to the ratio of the mean consumer value for product j in the social optimum to the highest value in the population.[9] 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.[10]

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

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

where $N = \sum_{i=1}^I N_i$ is total population size and $(XY)_{(I)}$ denotes the maximum order statistic for the product of X_i and Y_i: the maximum value for $X_i Y_i$ 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.

[10] 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

4. Calibration Methodology

This section describes the methods that we use to calibrate the new bounds on potential deadweight loss in the market for HIV pharmaceuticals.[11] 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 6, 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 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 that we 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 has the same disease risk $X_i \in (0, 1]$ and has the same income $Y_i > 0$. We have no reason to believe that the medical consequences of HIV infection differ across countries—indeed, untreated HIV is fatal everywhere. Thus, we abstract from cross-country variation in H_i, adopting the normalization $H_i = 1$. We defer to the Conclusion a discussion of a reinterpretation of the model that allows for heterogeneous consumers within a country, on the assumption that the national government makes purchases on their behalf.

As an intermediate step in deriving pharmaceutical demands, consider the counterfactual situation of a consumer who surely becomes infected. Assume that the elasticity of an individual's

and Snyder (2015, Figure 4) or taking the value from the tabulation in Kremer and Snyder (2016, Table 2).

[11] 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.

healthcare demand with respect to income is a constant ϵ across income and across individuals. For this to be the case, his willingness to pay to avoid harm from the disease as a function of his income Y_i must take the form Y_i^ϵ.[12]

Now move from the counterfactual case in which the consumer is surely infected 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: $X_i Y_i^\epsilon$. 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.

We next turn to the 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 that is sold at price p_d as long as $Y_i^\epsilon \geq p_d$. In expectation, $N_i X_i$ consumers in country i contract the disease. Thus global 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 products 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.[13]

[12]The most general form that preserves the property of constant income elasticity is $A_i Y_i^\epsilon$, which allows 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; this would contradict 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 $A Y_i^\epsilon$, with a leading coefficient A that is constant across consumers. Our specification of willingness to pay normalizes $A = 1$. For most of our analysis, this normalization is 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 divides out of the proportion. In our analysis of prices, which is done in levels, the combined normalizations $A = 1$ and $H_i = 1$ comport with World Health Organization procurement thresholds, as will be discussed below.

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

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

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 to calibrate vaccine demand. When we set ϵ to the value that we take for our baseline scenario, $\epsilon = 1$, and we maintain 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 $X_2 Y_2 = 450$. These willingness-to-pay numbers are the only relevant pricing points for a monopolist that wishes to extract as much surplus as possible. According to equation (11), vaccine demand equals the cumulative population of countries that have consumers with a willingness to pay that is at least as high as the given price: $Q_v(450) = N_2 = 1,500$ 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, $PS_v(450) = 450 \cdot 1,500 = 675,000$ and $PS_v(100) = 100 \cdot 2,500 = 250,000$, which implies $PS_v^0 = 675,000$, 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 these pricing points equals the cumulative infected population, $N_i X_i$, of countries with at least that willingness to pay: $Q_d(10,000) = N_1 X_1 = 10$ and $Q_d(9,000) = N_1 X_1 + N_2 X_2 = 85$. The resulting demand curve is shown in Panel B of Figure 1. Drug producer surplus is $PS_d(10,000) = 100,000$ and $PS_d(9,000) = 765,000$, which implies $PS_d^0 = 765,000$, 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 the time that the drug is sold because disease risk resolves into disease status, which allows the monopolist to extract almost all of the available

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 that turns out to be ineffective may later purchase the drug as well.

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 can be constructed in which the reverse inequality holds; the outcomes depend 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 (2015, Proposition 3) is that if income is homogeneous across countries—if $Y_i = \bar{Y}$—and at least two countries have different positive values of X_i, then $\rho_d^0 > \rho_v^0$. In essence, though incomes are equal in the two countries, the different disease risks generate different willingnesses to pay for a vaccine across the two countries, which (in the absence of price discrimination) limits the ability of the firm to capture all of the potential surplus; if a drug rather than a vaccine has been developed, after individuals have realized their probabilistic disease status, all of the infected individuals are willing (by assumption of equal incomes) to pay the same amount for the curative drug, which allows the firm to capture all of the potential surplus. By continuity, this inequality $\rho_d^0 > \rho_v^0$ continues to hold 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 that are 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 the demand for HIV pharmaceuticals, we would like some measure of the burden of the disease in the state of nature without any pharmaceuticals. As can be seen from 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). Unfortunately, the dip turned out to be short lived, rising again after 2010.

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).[14] Table 2 provides more details on the data sources and descriptive statistics for the main variables. Our dataset includes 158 countries.[15]

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 countries in the world, which reduces their willingness to

[14] The UNAIDS website that we 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 the country-level HIV data.

[15] Our data include 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.

pay in the calibration. Table 3 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 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 from Proposition 2. We conclude the section with an analysis of whether and how equilibrium in the baseline calibration can be improved with government subsidies.

This baseline specification has some practical appeal beyond its use of round numbers. Interpret the normalization of harm $H_i = 1$ as meaning that the pharmaceutical saves one Disability Adjusted Life Year (DALY). This is roughly the case for HIV drugs, as taking a year's course of ARTs roughly extends an infected person's life for that year. (Applying this interpretation to an HIV vaccine requires more care: We either have to assume that a yearly booster shot is needed or, if the vaccine provides more permanent protection, have to scale up the health benefit by the discounted flow of DALYs saved.) Under this interpretation, the baseline assumption that a consumer is willing to pay annual per-capita GDP Y_i to avoid H_i is the same as saying that he or she is willing to spend a year of income to save a year of life. The standard of the World Health Organization (WHO) is that a highly cost effective health intervention saves a disability adjusted life year (DALY) at a cost less than the country's GDP per capita (see Hutubessy, *et al.*, 2003). Thus our baseline assumption is simply that consumers purchase according to WHO standards.

As was discussed above (and will be discussed further in the Conclusion), the calibrations can be reinterpreted as applying to the purchase of pharmaceuticals by national governments on behalf of their citizens. This interpretation has the appeal of incorporating possible consumer

heterogeneity and epidemiological externalities within countries. According to this interpretation, the baseline assumption is simply that WHO standards guide national purchases.

6.1. Vaccine Market

The top panel of Figure 4 shows the calibrated (inverse) demand curve for a vaccine alone. From the methodology of 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 demarcating 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 was discussed in Section 3.3, the worst shape is the symmetrically truncated Zipf (STRZ) demand, which is drawn in the figure as the grey curve.

Notice that the calibrated demand curve largely overlaps the STRZ curve except where the United States appears, leading the demand curve to bulge there. The reason for the outsize influence of the U.S. on the market is that its consumers have very high incomes relative to most others, coupled with a not insubstantial HIV risk. These factors generate a large value of the product $X_i Y_i$ for the large population of U.S. consumers. U.S. consumers are 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 in the United States, their extremely high HIV prevalence rates result in their being higher demanders.

Further details of the calibration are provided in Table 4. The vaccine sells at a price of $130 to 344 million consumers. These numbers require some context—the price in particular—since, as was discussed in footnote 12, it depends on an arbitrary scaling of the willingness-to-pay function. The price—at almost eight times the mean vaccine-consumer value—is quite high, which leads to a quantity that is quite low: Less than 6% of the global population is served. All but 10 million of buyers live in South Africa and the United States.

Producer surplus is 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 the 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 1 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.

Using the formula for the Zipf similarity, Z_j^0, of the demand for product j provided by equation (9), we obtain $Z_v^0 = 0.66$: The calibrated vaccine demand curve is 66% similar to the STRZ worst case, in that it generates 66% of the deadweight loss bound that is 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.

From the methodology of Section 4, the drug demand curve is constructed by ordering countries in terms of Y_i (i.e., GDP per capita), which reflects 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 demarcating quantity $N_i X_i$, which is equal to the expected number of people in the country who contract the disease.

The producer-surplus-maximizing price is \$27,400. To provide some context for this price (which as explained in the previous subsection depends on an arbitrary scaling constant), it is extremely high: over ten times the mean of positive drug-consumer values. At this price, marginal consumers in Italy and inframarginal consumers in the 17 other higher-income countries (including the United States) purchase the drug. Although these countries have a combined population of 783

million, only 1.4 million units end up being sold. These rich countries have a relatively low HIV prevalence rate (an average of 0.12%), so relatively few people contract the disease and need an HIV drug. The 1.4 million consumers who are served represent a small fraction—less than 4%—of the 38 million infected individuals in the world market for the drug.

The difference in the shape of the distribution of consumer values for a drug versus a vaccine leads the 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 panel as compared to the upper panel.

Our finding that the drug is more lucrative than the vaccine in this calibration deserves further comment because it contrasts not only the "toy" example in Section 4 above but also the calibrations in Kremer and Snyder (2015), which were based on U.S. rather than global demand. The shape of the drug demand curve in the lower panel of Figure 4 is quite close to the STRZ worst case, which is drawn as a grey curve. Its shape is not conducive to inscribing a rectangle of substantial area underneath. The drug demand curve is more Zipf similar—$Z_d^0 = 0.74$—than the vaccine demand curve—$Z_v^0 = 0.66$.

The reason the drug demand curve Zipf similarity of the demand curves can be understood from their formulas. As shown in equation (11), a consumer's willingness to pay for a vaccine depends on the product of disease risk and income: $X_i Y_i$. The negative correlation between X_i and Y_i shown in Table reft:correlations concentrates consumer values from the extremes into the middle of the distribution, which leads to a lucrative bulge in the middle of the vaccine demand curve. From equation (12), a consumer's willingness to pay for a drug, by contrast, depends only on Y_i, so the negative correlation between X_i and Y_i does not induce a bulge in the middle of the drug demand curve.

Consumers obtain only 13% of B in the drug market. 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 that 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 1, this supremum equals $100-38=62\%$ of B.

6.3. Goodness of Fit

The predictions from our vaccine calibration cannot be compared to actual outcomes since an HIV vaccine has yet to be developed. We can perform the comparison for the ARTs that were developed. Our calibration, though simple, is able to match actual ART price and quantity quite closely. As discussed at the beginning of Section 6, we can interpret $27,400 as the model prediction of consumer expenditure per DALY for the HIV drug. Freedberg, *et al.* (2001, Table 2) estimate an actual expenditure per DALY in developed countries of $23,000. If the $23,000 number reflects price concessions in response to public pressure described by Reich and Bery (2005), our $27,400 estimate could be close to the counterfactual price without public pressure.

The calibrated quantity of 1.4 million can be compared to the actual quantity of 1.3 million, computed from Figure 2 as the reported 4% ART coverage in 2003 times the 0.5% HIV prevalence times 6.4 billion population. The predicted drug price and quantity from our calibrations both match the corresponding actual variables remarkably closely.

6.4. 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 that allows 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 was argued in Section 3, it simply equals the greater of the deadweight-loss suprema from the isolated product markets: the greater of 56% and 62%, and thus equal to 62%, which is reported in the last row of Table 4. This supremum can be approached in the limit as k_v approaches infinity—so that the vaccine market is certainly non-viable—and as k_d approaches 38% of B—so that the drug market falls just short of the margin of viability.[16]

[16] 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 supremum on deadweight loss lies somewhere in the interval between 6% and 100% of B, a fairly uninformative statement in this calibration.

6.5. 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 the former to the latter, which would eliminate 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, which leads to a large Harberger triangle: 41% of B. Introducing a government subsidy—a per-unit subsidy, which for accounting purposes let us assume is paid directly to the firm—and gradually raising it in penny increments has no effect on the 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 that is 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 \$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 the subsidy 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 the first-best social surplus since the 11% of consumers who remain unserved have very low incomes and thus very low drug demand.[17] 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

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

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 improved entry incentives—thus substantially reducing deadweight loss at the extensive margin—would call for much larger subsidies.

7. International Price Discrimination

Pharmaceutical manufacturers currently have considerable ability to price discriminate across countries, but there is an active policy debate on whether this ability should be curtailed—for example, in the contexts of parallel trade for pharmaceuticals within the European Union (Danzon 1998) or re-importation of Canadian pharmaceuticals in the United States (Pecorino 2002). In contrast to the baseline calibration, which assumed that the monopolist charges a uniform price across countries, the calibrations in this section allow for some form of international price discrimination, whether perfect or within limits set by reference pricing. Comparing the results across the two sections will allow us to assess the welfare effects of policies that impede price discrimination.

The results for various scenarios that involve international price discrimination are presented in Table 5. For brevity, the table just presents relative producer-surplus ratios ρ_v^0 and ρ_d^0 for each scenario. The producer-surplus ratios 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 1. 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 baseline calibration, which assumes uniform pricing.

The second row of the table presents results for a calibration that allows the firm 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: It allows 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 $p_{vi}^0 = X_i Y_i^\epsilon$ and 100% from the drug market by selling to all infected

individuals in country i at a price of $p_{di}^0 = Y_i^\epsilon$. Thus, as reported in Table 5, $\rho_v^0 = \rho_d^0 = 100\%$. This would eliminate any possibility of deadweight loss, whether due to inefficient pricing or entry decisions.

Governments have a variety of policies that can interfere with firms' ability to price discriminate perfectly. An international ban could shut down price discrimination entirely. A policy allowing pharmaceutical imports with no trade frictions could lead to the same result. Whether directly or indirectly implemented, a ban on price discrimination re-introduces the potential for substantial deadweight loss that we found in the calibrations for uniform pricing. Comparing the first two rows of Table 5. 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 range of other policies that have been adopted by countries that impede perfect price discrimination in pharmaceutical markets. Some countries control prices directly. Others use international reference pricing: They tie 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—that has been pursued to a greater or lesser extent by nearly 80% of the countries for which Kyle (2007) could obtain dispositive evidence. We model this by assuming that all of the countries in the world use the U.S. as the reference country: They cap prices in their countries to be no greater than some proportion $u \in (0,1)$ of the U.S. price.

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 U.S. 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—which would allow the firm to price discriminate perfectly across all remaining countries.

We proceed by calibrating the outcome from these two strategies: one that serves the U.S. and the other that excludes it. We compare the producer surplus from each strategy and select 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 u increasing in 0.5 increments from 0 to 2. The $u = 0$ 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. Yet it is not much less than the producer surplus in the uniform-pricing baseline. In other words, the firm does just about as well if it has to rely on the U.S. as its sole revenue source as it would if it served the whole world at a uniform price. This conclusion is particularly true in the drug market, where calibrated producer surplus ratios ρ_d^0 differ only by one percentage point between the uniform pricing scenario and scenario with $u = 0$. One can see the basis for this result in Figure 4. In either panel, the shaded rectangle—which indicates the 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 u causes producer surplus to asymptote to the 100% maximum for perfect price discrimination; this pattern can be seen graphically in Figure 5. Because the United States is such a high-demand country, constraining price to be half that in the U.S. ($u = 0.5$) 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 that results from 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 that are 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 $u \geq 0.2$, the producer earns more from serving than excluding the United States whether it produces a vaccine or a drug.

8. Alternative Parameterizations

This section conducts a series of comparative-statics exercises. We see how the calibrations change when we allow for heterogeneity in just disease risk or just income, when we incorporate an alternative data series for disease risk, and when we vary the income elasticity.

The results are presented in Table 6. For brevity, again, the table just presents relative producer-surplus ratios ρ_v^0 and ρ_d^0. For reference, the first row of the table repeats the relative producer-surplus ratios from the baseline calibration.

The first comparative-statics exercise examines what happens if countries have the same incomes and differ only in disease risk. This change makes the vaccine market much less lucrative. The top panel of Figure 6, which plots the inverse demand for this 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 bulge 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, which leads to an almost perfect STRZ shape. This is the worst possible shape for trying to inscribe a rectangle underneath the demand curve, so as to capture surplus for the producer. As reported in the second row of Table 6, the vaccine producer obtains only 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—which is shown in the lower panel of Figure 6—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—in this market. Unfortunately, Propositions 2 and 1 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 comparative-statics exercise gives all countries the same HIV risk and has them just vary by income, Y_i. We omit the graph of the demand curves 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 comparative-statics exercise returns to the baseline with heterogeneity in both X_i and Y_i but uses the revised data for 2003 HIV prevalence 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 that were published in UNAIDS (2006) rather than the initial data that were published in UNAIDS (2004a). The correlation between the initial and revised HIV prevalence series is quite high: 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 comparative-statics exercise varies the income elasticity that appears 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 exogenous 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 6 reports calibrations for 0.5 increments in ϵ from $\epsilon = 0$ to $\epsilon = 2$. Figure 7 shows how the pattern appears 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.

The overall pattern in Figure 7 is that 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, 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 in the private market: in the state of nature absent government procurement or insurance coverage. Such a state

is far from modern conditions, at least in the U.S. Getzen (2000) provides a survey of empirical studies of the income elasticity of health expenditures, locating a handful of studies that use 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, which correspond to the calibration with only disease-risk heterogeneity. Cross-country studies typically produce higher estimates of ϵ. The handful of cross-country studies that are surveyed by Getzen (2000) provide estimates of ϵ in the $[1.2, 1.4]$ range. The 1.3 midpoint was estimated by Newhouse (1977). Figure 7 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.

9. 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 pharmaceuticals 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: We combine 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.

Overall, the analysis reveals 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: this is the shape that was 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 that 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 the 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 that we assume, and compute the deadweight loss.

As a reality check, we compared the predictions from the calibrations to actual experiences in the case of HIV pharmaceuticals—ARTs—that were available circa 2003. Predictions were close to actual for price ($27,400 versus $23,000) as well as quantity (1.4 million versus 1.3 million).

In both the calibrated vaccine and drug markets, 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 the deadweight loss 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 also ran a suite of calibrations that allow for price discrimination. These results highlighted the pivotal role that the U.S. plays in global pharmaceutical demand. In these calibrations the firm finds that targeting just the U.S. is almost as lucrative as selling to the world at a uniform price. If other countries employ reference pricing that is pegged to the U.S., for most reasonable scenarios it is more lucrative for the firm to serve the U.S. and accept the ceiling that this imposes on other prices than to exclude the U.S. and price discriminate perfectly 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 U.S. would have a pivotal role whether or not other countries regulated prices, which explains in part why other countries may be emboldened to free ride.

Our baseline analysis abstracted from the involvement of governments and insurance companies as well as from epidemiological externalities. These factors can be integrated back into the

analysis under a suitable reinterpretation of the model: Rather than assuming that consumers are homogeneous within each country, assume instead that national governments (or large national insurers) decide whether to order a pharmaceutical on behalf of the heterogeneous citizens who are enrolled their health systems. Assume further that the governments purchase only pharmaceuticals that satisfy the WHO's standard of high cost effectiveness: saving a DALY at a cost that is less than the country's GDP per capita. Assume finally that taking either pharmaceutical saves one DALY (so, for example, taking an HIV drug for a year extends the infected person's life for a year).

Country i's pharmaceutical demands are the same under these assumptions as in our baseline model with direct sales to homogeneous consumers in the country with disease risk X_i, willingness to pay Y_i (given by GDP per capita), harm H_i normalized to 1, and income elasticity set to $\epsilon = 1$. The model of national-government purchases can also handle epidemiological externalities, since externalities mostly flow within a country and would be internalized by the government.[18]

In future work, we plan to derive a quantitative measure of how pivotal a country is 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 the global demand for an arbitrary product by assuming a lognormal distribution of income within each country and 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 surveys and a variety of other data sources to estimate the five requisite parameters for each country.

[18]Several caveats apply to the model of national-government purchases. First, governments must be assumed to purchase at a uniform posted international price. If the firm were instead allowed to post country-specific prices, the outcome would be equivalent to the scenario with perfect price discrimination. If governments were instead allowed to bargain with the firm, this raises a new scenario not yet analyzed. It is easy to see that if parties engage in Nash bargaining, they will arrive at the social optimum. Let α be the firm's bargaining share. Then the producer-surplus ratios in the bargaining model would be $\rho_v^0 = \rho_d^0 = \alpha$. Whether the deadweight-loss supremum is higher or lower in this bargaining scenario than in the uniform-posted-price baseline depends on α: if $\alpha = 0.5$, then the deadweight-loss supremum is lower in the bargaining scenario, but the reverse is true for sufficiently low α. A second caveat is that the national government's purchase must be tied to a commitment to universal access for all citizens—as, for example, Brazil committed to for ARTs in 1996 (Reich and Bery 2005)—rather than targeting the rich or otherwise higher demand consumers. A third caveat regards the interpretation of harm relieved by the pharmaceuticals. As noted at the beginning of Section 6, if we interpret the $H_i = 1$ normalization as a year's course of the drug, which extends life by a year, the parallel interpretation for the vaccine would involve a booster each year to maintain protection. If the vaccine is assumed to provide permanent protection, relieved harm would have to be scaled up by the discounted stream of expected DALYs saved.

Appendix: Proofs

Proof of Proposition 1: Assume benchmark values of the parameters. Suppose that the firm's only choice is to develop product j or nothing. We will show that the supremum on relative deadweight loss is bounded by $1-\rho_j^0$ above and below, proving the two are equal and thus (6) holds.

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

$$\begin{aligned}
\sup_{\{k_j \geq 0\}} \left(\frac{DWL^0}{B} \right) &\geq \lim_{k_j \downarrow PS_j^0} \left(\frac{DWL^0}{B} \right) && \text{(A1)} \\
&= \lim_{k_j \downarrow PS_j^0} \left(\frac{W^{00}}{B} \right) && \text{(A2)} \\
&= \lim_{k_j \downarrow PS_j^0} \left[\frac{\max(TS_j^{00} - k_j, 0)}{B} \right] && \text{(A3)} \\
&= \frac{TS_j^{00} - PS_j^0}{B} && \text{(A4)} \\
&= \frac{B - PS_j^0}{B} && \text{(A5)} \\
&= 1 - \rho_j^0. && \text{(A6)}
\end{aligned}$$

Equation (A1) 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. Equation (A2) holds because nothing is developed in equilibrium in the limit, which implies $W^0 = 0$ and thus $DWL^0 = W^{00} - W^0 = W^{00}$ in the limit. To see (A3), note that either product j is developed, which yields the social first-best surplus $TS_j^{00} - k_j$, or there is no product, which yields 0. The social optimum generates the maximum of these two social surpluses. Equation (A4) follows from evaluating the limit, (A5) follows from (2), and (A6) follows from the definition of ρ_j^0.

We next show that the supremum is bounded from above by $1-\rho_j^0$. If $W_j^{00} = 0$, then $DWL_j^0 = 0 \leq 1-\rho_j^0$, and we are done. So suppose instead that $W_j^{00} > 0$. We have the following series of steps:

$$\begin{aligned}
DWL^0 &= W^{00} - W^0 && \text{(A7)} \\
&= W_j^{00} - W^0 && \text{(A8)} \\
&\leq W_j^{00} - \Pi_j^0 && \text{(A9)} \\
&= TS_j^{00} - k_j - (PS_j^0 - k_j) && \text{(A10)} \\
&= B - PS_j^0. && \text{(A11)}
\end{aligned}$$

Equation (A7) holds by definition. Equation (A8) holds since $W^{00} = \max(0, W_j^{00}) = W_j^{00}$ by the maintained assumption that $W_j^{00} > 0$. Equation (A9) holds since $W_j^0 \geq \Pi_j^0$. Equation (A10) follows from substituting relevant definitions and (A11) substituting from (2) and canceling terms. Dividing (A7)–(A11) by B yields

$$\frac{DWL^0}{B} \leq 1 - \rho_j^0. \tag{A12}$$

Conditions (A6) and (A12) sandwich the supremum between $1-\rho_j^0$ above and below, which yields (6) as an exact equality. *Q.E.D.*

Proof of Proposition 2: Similar to the previous proof, we will show that the supremum on relative deadweight loss is bounded by $1-\rho_{\min}^0$ from above and below, which proves that the two are equal and thus (7) holds. For concreteness, suppose throughout the proof that

$$PS_v^0 \leq PS_d^0. \tag{A13}$$

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 from below:

$$\begin{aligned}
\sup_{\{k_v,k_d\geq 0\}}\left(\frac{DWL^0}{B}\right) &\geq \lim_{k_v\downarrow PS_v^0,k_d\uparrow\infty}\left(\frac{DWL^0}{B}\right) && \text{(A14)}\\
&= \lim_{k_v\downarrow PS_v^0}\left(\frac{W^{00}}{B}\right) && \text{(A15)}\\
&= \lim_{k_v\downarrow PS_v^0}\left[\frac{\max(W_v^{00},0)}{B}\right] && \text{(A16)}\\
&= \lim_{k_v\downarrow PS_v^0}\left[\frac{\max(TS_v^{00}-k_v,0)}{B}\right] && \text{(A17)}\\
&= 1-\rho_v^0 && \text{(A18)}\\
&= 1-\rho_{\min}^0. && \text{(A19)}
\end{aligned}$$

The arguments for (A14) and (A15) are similar to those for (A1) and (A2), respectively. Equation (A16) holds because the socially efficient product strategy cannot involve development of a drug, alone or together with a vaccine, for sufficiently large k_d. Hence $\sigma^{00}\in\{v,n\}$, which implies that $W^{00}=\max(W_v^{00},0)$. Equation (A17) holds by definition. Equation (A18) follows from steps that are similar to (A3)–(A6), and (A19) holds by assumption (A13).

We next bound the supremum from above. We start by establishing that the following inequality,

$$DWL^0 \leq B-PS_{\min}^0, \tag{A20}$$

holds regardless of which value—$\sigma^{00}\in\{v,d,b,n\}$—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 (A20) 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:

$$\begin{aligned}
DWL^0 &= W^{00}-W^0 && \text{(A21)}\\
&= TS^{00}-k_{\sigma^{00}}-(\Pi^0+CS^0) && \text{(A22)}\\
&\leq TS^{00}-k_{\sigma^{00}}-\Pi^0 && \text{(A23)}\\
&\leq TS^{00}-k_{\sigma^{00}}-\Pi_{\sigma^{00}}^0 && \text{(A24)}\\
&= TS^{00}-PS_{\sigma^{00}}^0. && \text{(A25)}\\
&= B-PS_{\sigma^{00}}^0. && \text{(A26)}
\end{aligned}$$

Equations (A21) and (A22) follow from the substitution of the definitions of the relevant variables, and (A23) follows from $CS^0 \geq 0$. Equation (A24) holds because the equilibrium product strategy is the most profitable, implying $\Pi^0 \geq \Pi^0_{\sigma^{00}}$. Equation (A25) follows from $\Pi^0_{\sigma^{00}} = PS^0_{\sigma^{00}} - k_{\sigma^{00}}$ and (A26) from (2).

We next show that

$$PS^0_{\sigma^{00}} \geq PS^0_{\min} \tag{A27}$$

for all $\sigma^{00} \in \{v,d,b\}$. If $\sigma^{00} = v$, then $PS^0_{\sigma^{00}} = PS^0_v \geq PS^0_{\min}$, which implies that (A27) holds. Similar arguments show that (A27) 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 $PS^0_b \geq PS^0_d \geq PS^0_{\min}$, which implies that (A27) holds, which completes the proof that (A27) holds for all $\sigma^{00} \in \{v,d,b\}$.

We can now complete the proof: Combining (A26) and (A27), we have $DWL^0 \leq B - PS^0_{\min}$ for all $\sigma^{00} \in \{v,d,b\}$. Combining this fact with the argument in the text following (A20) implies that (A20) holds for all $\sigma^{00} \in \{v,d,b,n\}$. Dividing (A20) by B,

$$\frac{DWL^0}{B} \leq 1 - \rho^0_{\min} \tag{A28}$$

for all $k_v, k_d \geq 0$, which implies that

$$\sup_{\{k_v,k_d \geq 0\}} \left(\frac{DWL^0}{B} \right) \leq 1 - \rho^0_{\min}. \tag{A29}$$

Conditions (A19) and (A29) sandwich the supremum from above and from below at $1 - \rho^0_{\min}$, which yields (7) as an exact equality. *Q.E.D.*

References

Acemoglu, D., & Linn, J. (2004). Market size in Innovation: Theory and evidence from the pharmaceutical industry. *Quarterly Journal of Economics*, 119, 1049–1090.

Anderson, O. W., Collette, P., & Feldman, J. J. (1960). *Expenditure patterns for personal health services, 1953 and 1958: Nationwide survey*. New York: Health Information Foundation.

Arrow, K. (1962). Economic welfare and the allocation of resources for inventions. In R. Nelson (Ed.), *The rate and direction of inventive activity: Economic and social factors* (pp. 609–626). Princeton, New Jersey: Princeton University Press.

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

Budish, E., Roin, B. N., & Williams, H. L. (2015). Do firms underinvest in long-term research? Evidence from cancer clinical trials. *American Economic Review*, 105, 2044–2085.

Chen, F., & Toxvaerd, F. (2014). The economics of vaccination. *Journal of Theoretical Biology*, 363, 105–117.

Danzon, P. M. (1998). The economics of parallel trade. *PharmacoEconomics*, 13, 293–304.

Danzon, P. M., & and Ketcham, J. D. (2004). Reference pricing of pharmaceuticals for medicare: Evidence from Germany, the Netherlands, and New Zealand. *Frontiers in Health Policy Research*, 7, 1–54.

Danzon, P. M., Wang, Y. R., & Wang, L. (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.

Finkelstein, A. (2004). Static and dynamic effect of health policy: Evidence from the vaccine industry. *Quarterly Journal of Economics*, 119, 527–564.

Francis, P. J. (1997). Dynamic epidemiology and the market for vaccinations. *Journal of Public Economics*, 63, 383–406.

Freedberg, K. A., *et al.* (2001). The cost effectiveness of combination antiretroviral Therapy for HIV disease. *New England Journal of Medicine*, 344, 824–831.

Garber, A. M., Jones, C. I., & Romer, P. M. (2006). Insurance and incentives for medical innovation. *Forum for Health Economics & Policy*, 9, 1–27.

Geoffard, P.-Y., & Philipson, T. (1997). Disease eradication: Public vs. private vaccination. *American Economic Review*, 87, 222–230.

Gersovitz, M. (2003). Births, recoveries, vaccinations, and externalities. In R. Arnott, B. Greenwald, R. Kanbur, & B. Nalebuff (Eds.), *Economics for an imperfect world: Essays in honor of Joseph E. Stiglitz* (pp. 469–483). Cambridge, Massachusetts: MIT Press.

Gersovitz, M., & Hammer, J. S. (2004). The economical control of infectious diseases. *Economic Journal*, 114, 1–27.

Gersovitz, M., & Hammer, J. S. (2005). Tax/subsidy policy toward vector-borne infectious diseases. *Journal of Public Economics*, 89, 647–674.

Getzen, T. E. (2000). Health care is an individual necessity and a national luxury: Applying multilevel decision models to the analysis of health care expenditures. *Journal of Health Economics*, 19, 259–270.

Harberger, A. C. (1954). Monopoly and resource allocation. *American Economic Review*, 44, 77–92.

Hutubessy, R., Chisholm, D., Tan-Torres Edejer, T., & WHO-CHOICE. (2003). Generalized cost-effectiveness analysis for national-level priority-setting in the health sector. *Cost effectiveness and resource allocation*, 1, 8.

Kremer, M., & Snyder, C. M. (2003). Why are drugs more profitable than vaccines? National Bureau of Economic Research working paper no. 9833.

Kremer, M., & Snyder, C. M. (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, M., & Snyder, C. M. (2015). Preventives versus treatments. *Quarterly Journal of Economics*, 130, 1167–1239.

Kremer, M., & Snyder, C. M. (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, M., Snyder, C. M., & Williams, H. L. (2012). Optimal subsidies for prevention of infectious disease. Harvard University working paper.

Kyle, M. K. (2007). Pharmaceutical price controls and entry strategies. *Review of Economics and Statistics*, 89, 88–99.

Lakdawalla, D. N., & Sood, N. (2013). Health insurance as a two-part pricing contract. *Journal of Public Economics*, 102, 1–12.

Newell, R. G., Jaffe, A. B., & and Stavins, R. N. (1999). The induced innovation hypothesis and energy-saving technological change. *Quarterly Journal of Economics*, 114, 907–940.

Newhouse, J. P. (1977). Medical care expenditure: A cross-national survey. *Journal of Human Resources*, 12, 115–125.

Pecorino, P. (2002). Should the US allow prescription drug reimports from Canada? *Journal of Health Economics*, 21, 699–708.

Pinkovskiy, M., & Sala-i-Martin, X. (2009). Parametric estimations of the world distribution of income. National Bureau of Economic Research working paper no. 15433.

Reich, M. R., & Bery, P. (2005). Expanding global access to ARVs: The challenges of prices and patents. In K. H. Mayer and H. F. Pizer (Eds.), *The AIDS pandemic: Impact on science and society* (pp. 324–350). New York: Academic Press.

Sood, N., de Vries, H., Gutierrez, I., Lakdawalla, D. N., & and Goldman, D. P. (2008). The effect of regulation on pharmaceutical revenues: Experience in nineteen countries. *Health Affairs*, 28, w125–w137.

Tirole, J. (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	450	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 the sample of 158 countries. We compute HIV prevalence by dividing "Estimated number of people living with HIV, adults and children, end 2003" from UNAIDS (2004a), 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	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	$\sup\left(\frac{DWL_d^0}{B}\right) = 1-\rho_d^0$	0.62
Zipf similarity	Z_d^0	0.74
Both products possible		
DWL supremum	$\sup\left(\frac{DWL^0}{B}\right) = 1-\rho_{\min}^0$	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 are expressed as a proportion of first-best social surplus (equivalently, as a proportion of disease burden). The suprema are taken over $k_v, k_d \geq 0$.

Table 5: Calibrations of Producer-Surplus Ratios for Various Price-Discrimination Scenarios

Scenario	Vaccine market ρ_v^0	Drug market ρ_d^0
Uniform-pricing baseline	0.44	0.38
Perfect price discrimination	1.00	1.00
Price ceiling tied to varying proportions of U.S. price		
$u = 0.0$	0.37	0.37
$u = 0.5$	0.80	0.95
$u = 1.0$	0.84	1.00
$u = 1.5$	0.88	1.00
$u = 2.0$	0.91	1.00
Excluding U.S. to evade the price ceiling	0.63	0.63

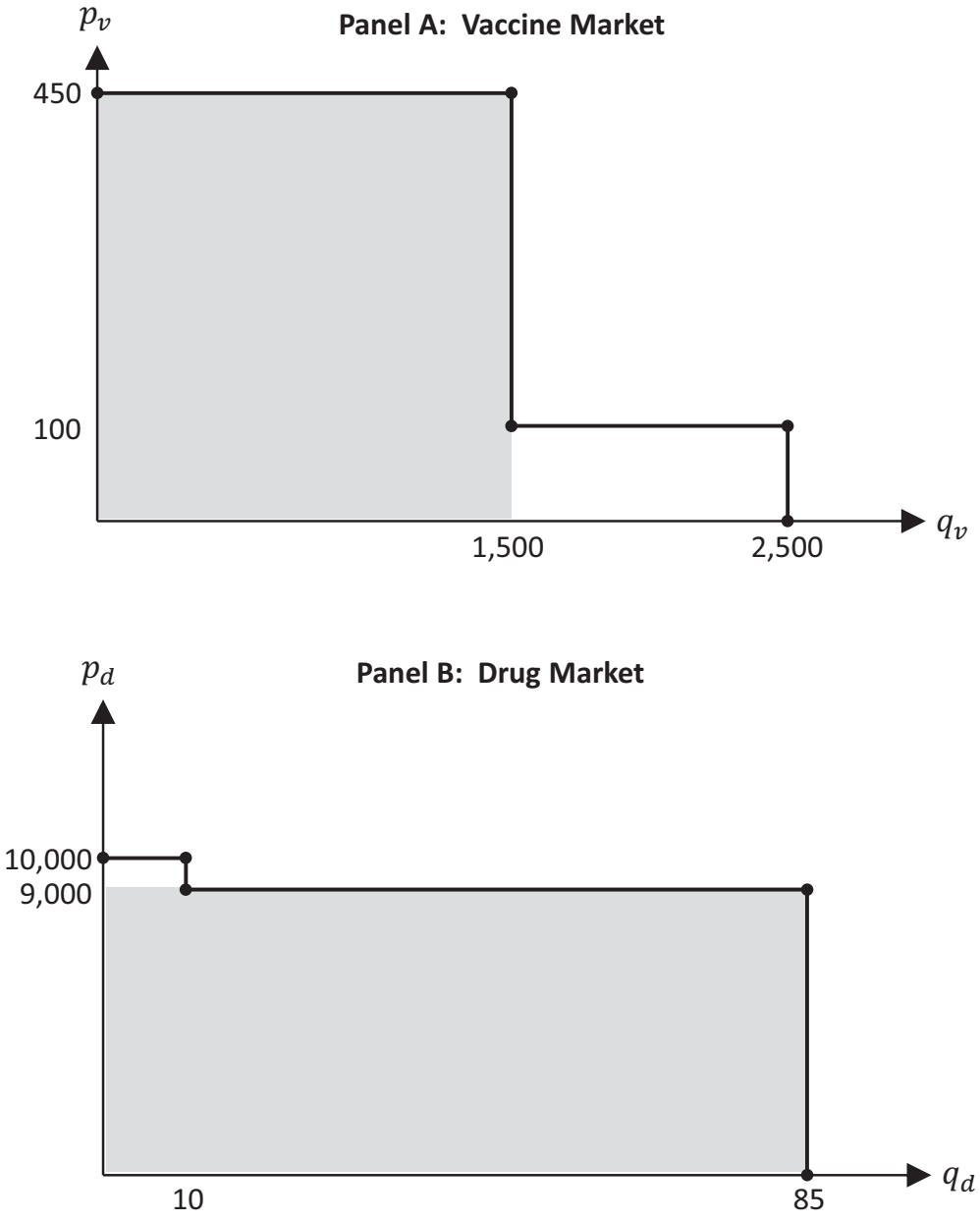
Notes: All entries are producer surplus as a proportion of first-best social surplus (equivalently, as a proportion of disease burden). Scenarios have the same parameters and functional form as the baseline, just changing the pricing strategy from uniform to one that allows prices to differ across countries.

Table 6: Calibrations of Producer-Surplus Ratios for Alternative Parameterizations

Scenario	Vaccine market ρ_v^0	Drug market ρ_d^0
Baseline parameterization	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

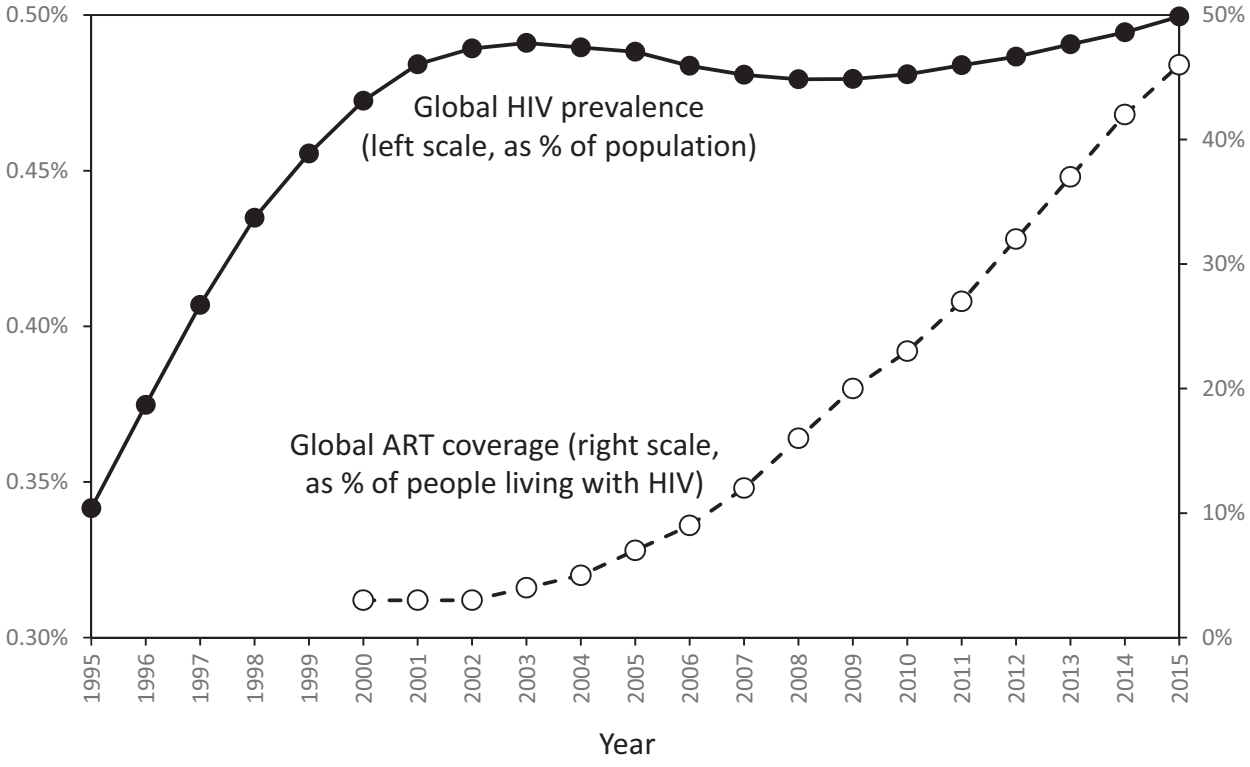
Notes: All entries are producer surplus as proportion of first-best social surplus (equivalently, as proportion of disease burden). Scenarios return to baseline assumption of uniform pricing. Scenarios return to baseline in all other dimensions as well except for the stated alternative.

Figure 1: Demand Curves in "Toy" Example



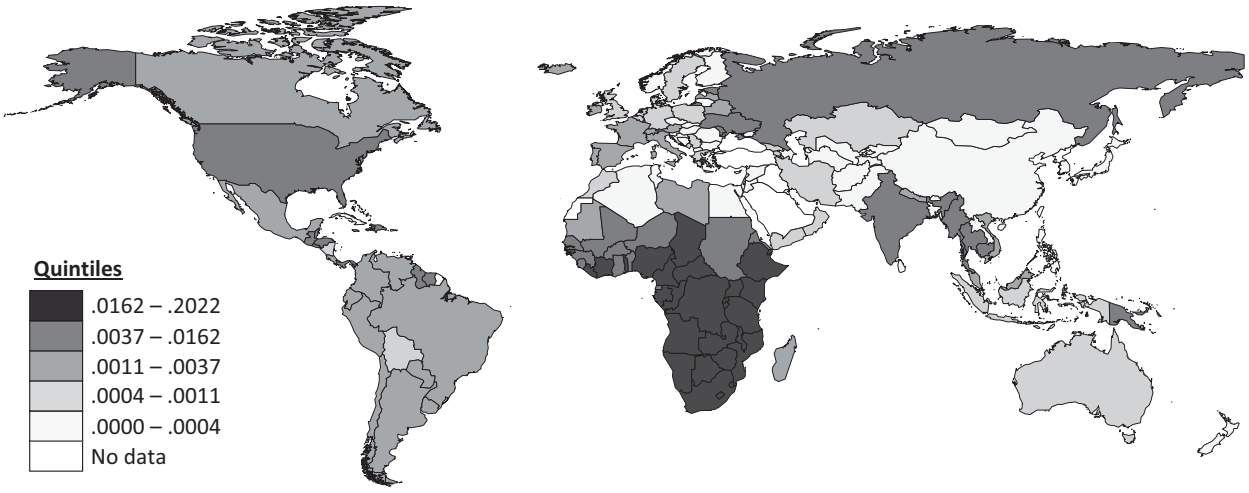
Notes: Pharmaceutical demands derived from fictitious data in Table 1. Axes scaled so that a unit of area represents the same producer surplus in both panels. Producer surplus equal to area of shaded region.

Figure 2: Trends in Global HIV Prevalence and Treatment

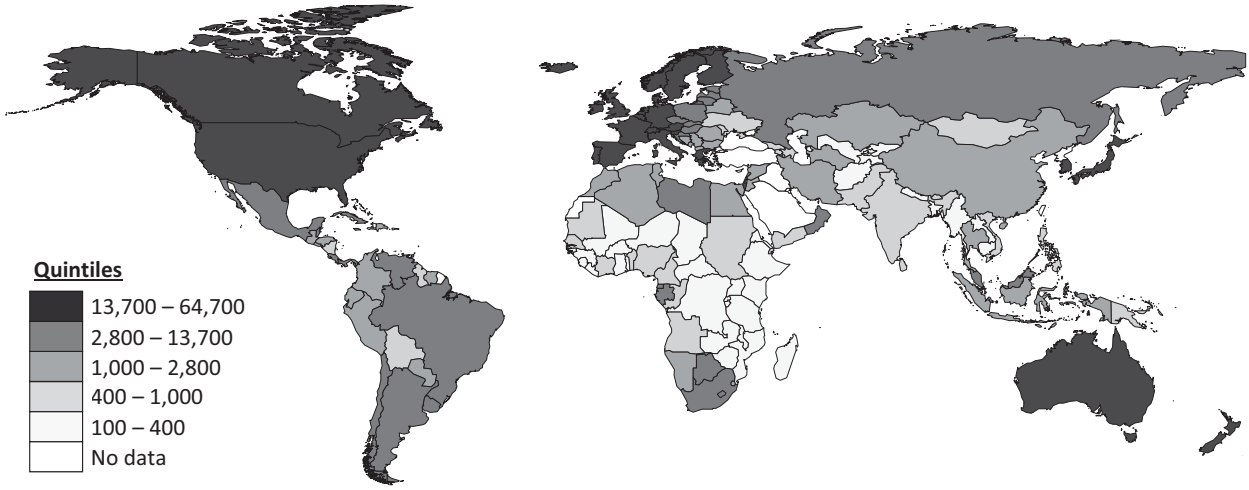


Sources: Left-axis variable equals HIV cases (from UNAIDS data on "Number of people living with HIV," downloaded from http://aidsinfo.unaids.org) as a percent of population (from World Bank data on "Population, total," downloaded from http://data.worldbank.org/indicator/SP.POP.TOTL). Right-axis variable 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



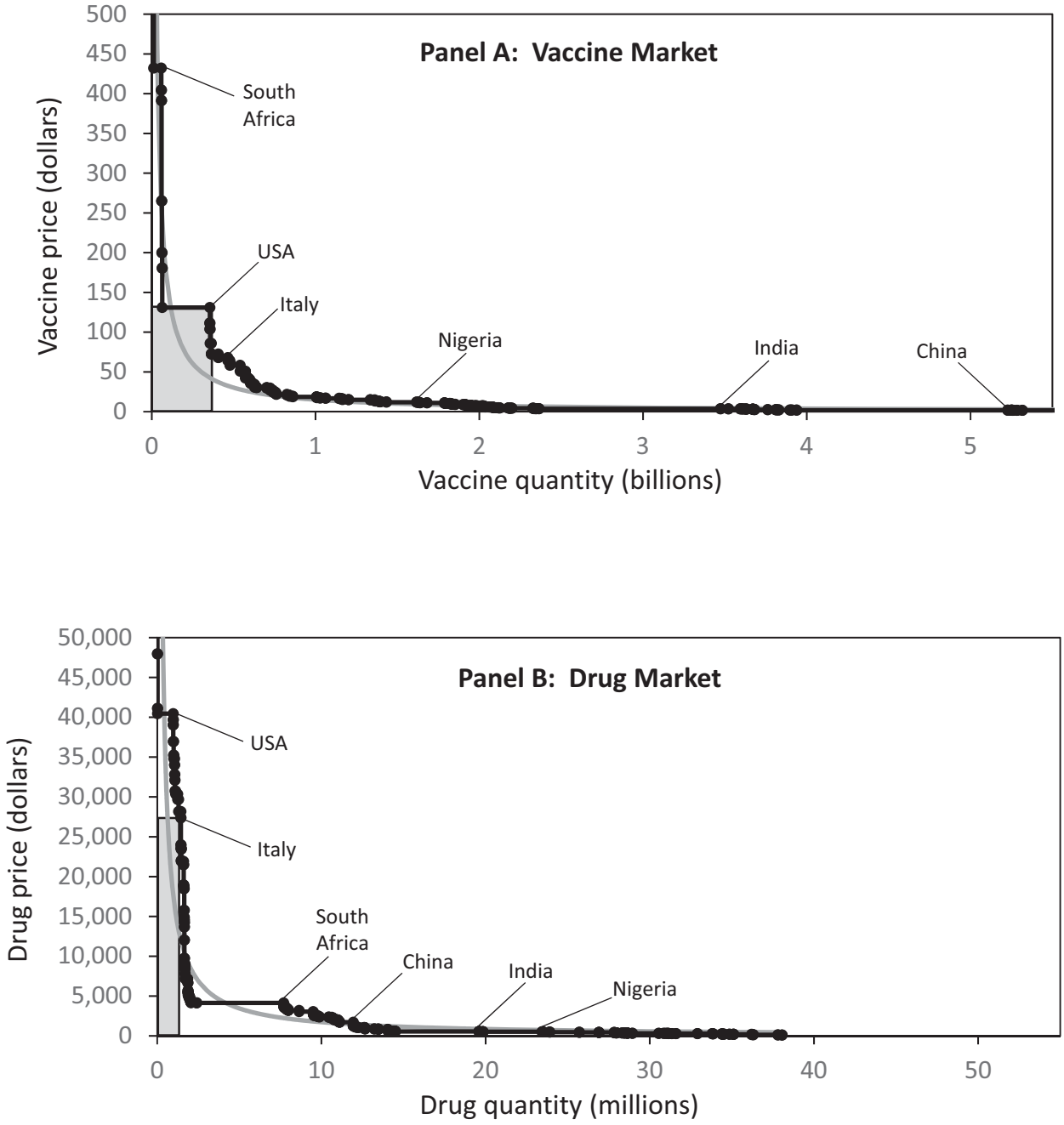
Panel A: HIV Prevalence in 2003 (X_i)



Panel B: GDP Per Capita in 2003 (Y_i)

Sources: See Table 2.

Figure 4: Demand Curves for HIV Pharmaceuticals in Baseline Calibration



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

Figure 5: Effect of Reference Pricing Pegged to the United States on Producer-Surplus Ratio

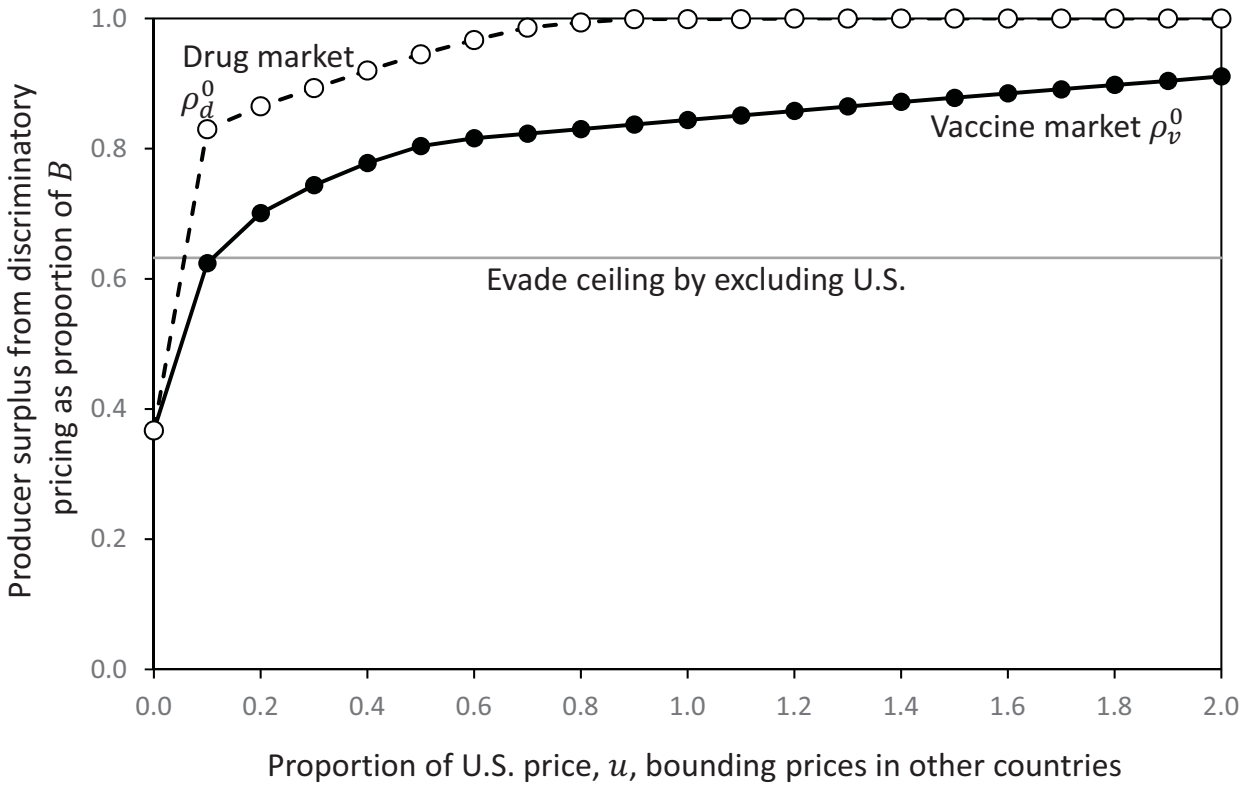
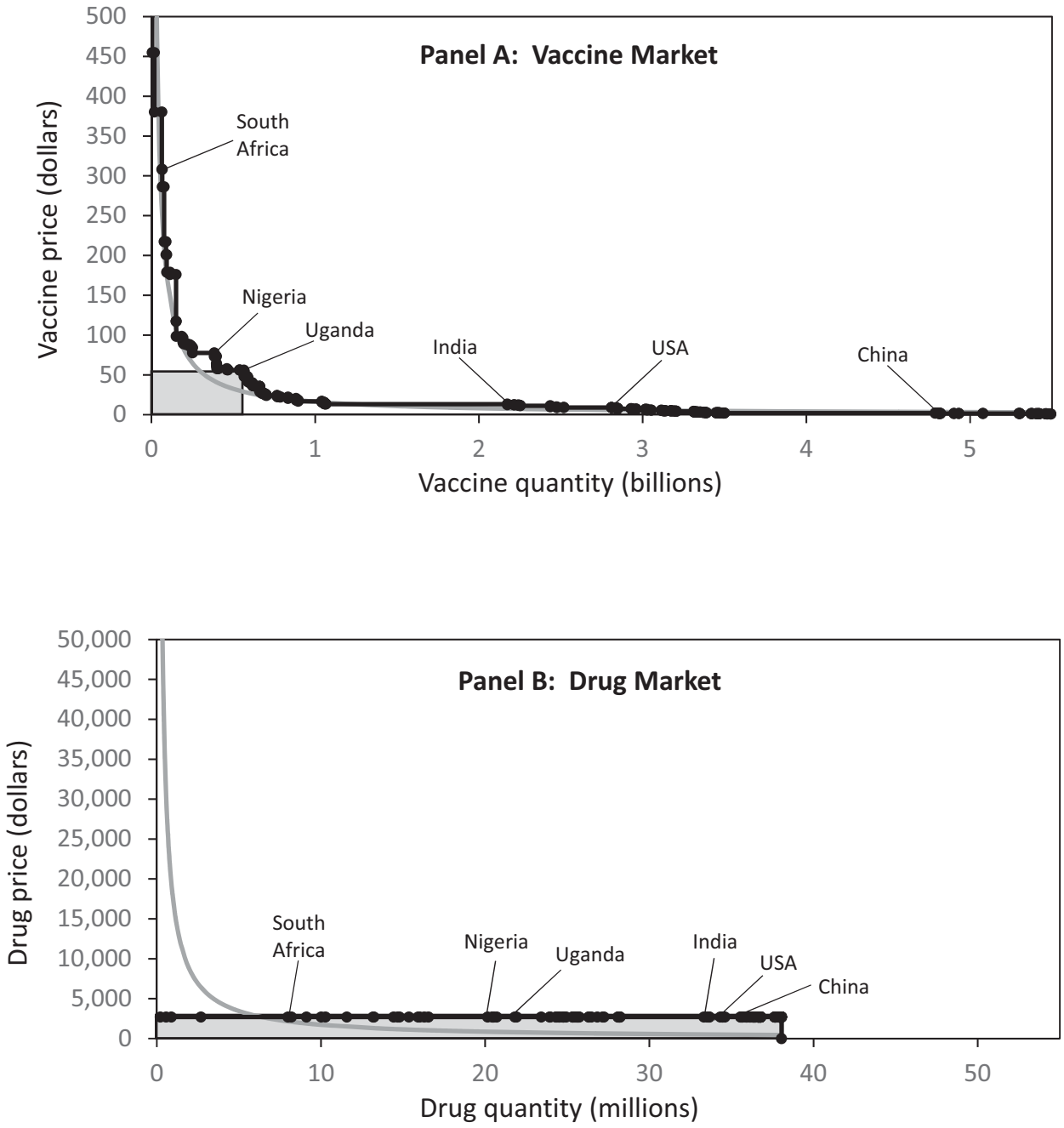


Figure 6: Demand Curves for HIV Pharmaceuticals in Calibration with Heterogeneity in Disease Risk Alone



Notes: Calibration with heterogeneity in disease risk (X_i) alone. As in previous figure, income elasticity is $\epsilon = 1$. Income is 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 the same scales as previous figure; as there, a unit of area represents the same producer surplus in both panels. Both axes are truncated to aid visualization. Shaded rectangle represents the equilibrium pricing decision; its area equals producer surplus. Grey line is the STRZ demand curve with same area underneath (and thus same total surplus) as the calibrated demand.

Figure 7: Effect of Income Elasticity on Producer-Surplus Ratio