

NBER WORKING PAPER SERIES

IP & EXTERNAL CONSUMPTION EFFECTS:
GENERALIZATIONS FROM HEALTH CARE MARKETS

Tomas J. Philipson
Stéphane Mechoulan
Anupam B. Jena

Working Paper 11930
<http://www.nber.org/papers/w11930>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2006

This paper is a revision of earlier versions including NBER Working Paper No 9598, April 2003. We are thankful for comments from seminar participants at the University of Chicago, George Mason University, the University of Maryland, the ALEA Meetings, the AEA Meetings, the Conference on Markets for Pharmaceuticals and the Health of Developing Nations, as well as from Michael Baker, Gary Becker, John Grana, Robert McMillan, Kevin Murphy, and Gilles Saint-Paul. Philipson benefitted from discussions of many related topics with colleagues at FDA, CNS, and the Pull Mechanism Working Group of the Center for Global Development, Washington DC whom, however, should not be held responsible for any views contained. Philipson acknowledges financial support from Pfizer Inc as well as the George Stigler Center for the Study of Economy and the State, and the John M. Olin Law & Economics Program, both at the University of Chicago. Jena acknowledges fellowship support from the NIH through the University of Chicago Medical Scientist Training Program. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

©2006 by Tomas J. Philipson, Stéphane Mechoulan and Anupam B. Jena. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

IP & External Consumption Effects: Generalizations from Health Care Markets
Tomas J. Philipson, Stéphane Mechoulan and Anupam B. Jena
NBER Working Paper No. 11930
January 2006
JEL No. I1

ABSTRACT

There is a long-standing literature that recognizes that an efficient solution in correcting a consumption externality is applying subsidies or taxes that align private with social incentives. An equally long-standing literature tackles the appropriate methods of generating the efficient amount of R&D into goods without external effects in consumption, e.g., the analysis of the welfare effects of patent regulations. This paper addresses the joint determination of intellectual property (IP) and externality remedies. We discuss the impact that IP has on remedies for externalities as well as the reverse problem of the impact externalities have on the design of IP. The results are discussed in the context of health care markets in general, and pharmaceutical markets in particular, the latter being one of the most R&D-intensive industries, and at the same time often being faced with altruistic access issues. A central but non-recognized tradeoff in health care concerns the correct R&D incentives when altruistic motives dictate that lives will be saved whenever feasible technologies exist. Understanding this tradeoff is central to understanding the efficiency of the observed growth in health care spending often attributed to technological change. We calibrate the model for the US health care sector and find that altruistic gains amount to 27 percent of consumer gains and that this implies R&D is under-provided by 61 percent in face of such altruistic motives.

Tomas Philipson
The Irving B. Harris Graduate School
of Public Policy Studies
The University of Chicago
1155 East 60th Street
Chicago, IL 60637
and NBER
t-philipson@uchicago.edu

Stéphane Mechoulan
Department of Economics
The University of Toronto
150 St. George Street
Toronto, Ontario M5S3G7
Canada
s.mechoulan@utoronto.ca

Anupam B. Jena
Department of Economics
The University of Chicago
1126 E 59th Street
Chicago, IL 60637
ajena@uchicago.edu

I. Introduction

A long-standing literature discusses efficient methods of correcting consumption externalities through applying subsidies and taxes that align private with social incentives, as first recognized by Pigou (1932). However, this classic problem assumes that there is no technological change in the good that confers the external effects.

An equally long-standing literature tackles the appropriate methods of stimulating innovation for goods that only have private consumption effects, *e.g.*, the analysis of the welfare effects of intellectual property (IP) regulations.³ However, this literature traditionally assumes that there are no external effects in the consumption of the good for which there is technological change. Although these two issues are well analyzed, the problem of dealing with *both* technological change and external consumption effects remains less explored.⁴

The lack of a framework for understanding this joint allocation problem seems to have led to confusion and disagreement about appropriate policy solutions for many important issues implicitly involving such stakes. This has been particularly true for many policy issues facing the pharmaceutical industry, which is one of the most R&D-intensive industries and is also confronted with altruistic or human rights-based access issues. For example, consider the case of antibiotics in which there has been great pressure to limit usage in order to slow down the rising threat of drug resistance. Such negative external effects, induced by current consumption lowering the value of future

³ Of course, there is a vast literature on the external effects of the R&D-process itself rather than on the external consumption effects of the final good, see *e.g.*, Jones and Williams (2000).

⁴ See Parry (1995) for an analysis of the optimal pollution tax when the state of technology is endogenous.

consumption,⁵ have prompted what may be interpreted as Pigouvian-like measures of taxing or limiting the demand for antibiotics. However, such classic remedies discourage R&D into new antibiotics that will replace those to which bacteria have become resistant. Therefore, the costs of limiting the use of antibiotics may dominate the benefits even though such limits are the appropriate policy in the absence of technological change.

As another example, consider the pressing problem of providing drugs to third world nations for diseases such as AIDS, malaria, or tuberculosis. These nations have the greatest numbers of people infected but cannot afford the costs of new drugs. As it appears that richer developed countries care about expanding the access to drugs in poorer countries this problem appears to be one of efficiently providing both technological change and consumption under positive external effects.

This tradeoff between externalities and R&D is of more general importance to health care markets worldwide. As many observers have argued that R&D is the key to the continued expansion of the health care sector in the economy,⁶ one may also claim that the joint allocation problem studied here is perhaps *the* central one to understand whether the growth in health care spending observed is efficient.⁷ Since many developed nations implicitly have decided that it is intolerable to let people die or suffer when existing medical technologies can prevent it, public financing often covers such technologies. Yet, such altruistic adoption of new technologies once they have been developed needs to be

⁵ For the purpose of this paper, we will assume that these negative external effects dominate the classic positive external effects of treatments for communicable diseases (see *e.g.*, Philipson (2000)).

⁶ See *e.g.*, Newhouse (1992).

⁷ Many other industries, such as “research tools” industries, industries with network-, peer-group-, or herd-effects, and industries in which production induces pollution, seem to involve similar issues of balancing externalities *ex-post* with R&D incentives *ex ante*.

balanced against the technological change it induces. However, little is known about how this is done appropriately and thus whether the observed changes over time in health care spending are efficient.

This paper attempts to remedy this lack of understanding by analyzing and deriving the optimal treatment of externalities and R&D incentives when they co-exist. First, we discuss the impact that IP has on remedies aimed at solving externality problems. We argue that classic Pigouvian solutions are inappropriate under technological change; for goods with external effects, just as for those without, *ex-post* static efficiency is often inconsistent with *ex-ante* dynamic efficiency.⁸ Therefore, Pigouvian solutions are dynamically inefficient for the same reasons that competitive markets for new inventions in standard markets are; both support *ex-post* efficiency but induce dynamic inefficiency.

Second, we discuss the reverse problem of the impact that externalities have on the design of IP. External effects influence whether patents or prizes best reward innovation. When there are positive external effects, such as a result of altruism in health care delivery, we stress that rewards to innovation should not be guided by the potential consumer and producer surplus as it is with patents, but by the entire social surplus that includes the benefits to those externally affected. In the extreme case of third-world diseases, consumers often cannot pay above variable cost, which means there are no *ex-post* profits and hence patents are of little value to innovators. However, if prizes replace patents, efficient *production* is non-trivial under external effects because unrestricted licensing after the prize has been awarded does not generate the efficient output. Given

⁸ Without externalities, this is of course the rationale for patents.

these shortcomings of the traditional forms of prizes under external effects, we analyze alternative prize contracts that may induce better production and distribution incentives.

To assess the importance of altruistic motives for dynamic efficiency, we calibrate the model for US health care markets in general and, and the pharmaceutical market for HIV drugs, in particular. This is a particularly relevant example, as the majority of consumption of HIV drugs is financed by Medicaid, and treatment has undergone tremendous recent technological change. For the case of HIV drugs, we find that the altruistic surplus may be as high as 25 percent of consumer surplus, on the order of \$99 billion since the start of the epidemic. For health care consumption as a whole, we calibrate the altruistic surplus to be nearly 27 percent of consumer surplus, implying estimates of just over \$1.1 trillion annually. We calibrate that these levels of altruism imply an under-investment in R&D of 23 percent for HIV research and 61 percent for R&D into the general health care sector as a whole.

The paper is related to several literatures. It is of course related to the voluminous research on the appropriate methods of treating externalities (see Laffont, 1987 for a survey) without technological change. The paper also extends the classic work on the tradeoffs between R&D stimuli (push) and patents and prizes (pull) (see Nordhaus, 1969 and 1972; Wright, 1983, Scotchmer, 2005) as well as the more recent literature discussing prizes for third-world disease R&D (Kremer and Glennester, 2004).

The paper may be briefly outlined as follows. Section II sets up the allocation problem involving externalities under technological change. Section III discusses the optimal way of intervening to solve externality problems given a certain form of IP. Section IV discusses the reverse problem of how to best design IP given the existence of

externalities. Section V presents our calibration results for HIV/AIDS and the overall health care economy in the US. Finally, Section VI provides concluding remarks.

II. External Effects and Intellectual Property

Let y denote the quantity of output, $p(y)$ denote the private inverse demand curve, $e(y)$ the monetary value of the external consumption effects to non-consumers, and $c(y)$ the total cost function. Let the producer surplus (profits) of a monopolist be denoted:

$$\pi(y) = p(y)y - c(y) \quad (1)$$

and let y_π denote the assumed unique output that maximizes profits π . The surplus of the consumers engaged in consumption is denoted:

$$s(y) = \int_0^y [p(q) - p(y)]dq \quad (2)$$

The social welfare $W(y)$ is then defined by consumer and producer surplus together with the surplus $e(y)$ of those affected externally by consumption:

$$W(y) = s(y) + \pi(y) + e(y) \quad (3)$$

Let y_W denote the assumed unique output that maximizes W .

Let $x(\cdot)$ be an increasing, differentiable, and strictly concave function representing the probability of discovering an invention as a function of the level of R&D, r , undertaken. The optimal level of R&D that maximizes expected payoffs for any hypothetical *ex-post* prize z is denoted $r(z)$ and is defined by:

$$r(z) = \operatorname{argmax}_r x(r)z - r \quad (4)$$

Our assumptions on $x(r)$ imply that $r(z)$ is an increasing function.

An allocation (r, y) is defined as an R&D level, r , together with a series of dated outputs, $y = (y_1, y_2, \dots)$ after the invention has been discovered. The expected social welfare given R&D r and output y is:

$$D(r, y) = x(r)W(y) - r \quad (5)$$

The *first-best* R&D and output (r^*, y^*) maximizes this social welfare and implies the necessary first-order conditions:

$$D_y = xW_y = 0 \quad (6)$$

$$D_r = x_r W - 1 = 0$$

Clearly, the first best and *ex-post* optimal output coincide: $y^* = y_w$. The corresponding first-best R&D takes into account the highest level of *ex-post* welfare $r^* = r(W(y_w))$.

We will mainly discuss two forms of IP, patents of length τ and prizes of size θ . The allocation induced by a patent is a monopoly output y_π for τ years, and the competitive output y_c thereafter. *Ex-post* welfare is:

$$W(\tau) = v(\tau)W(y_\pi) + [v(\infty) - v(\tau)]W(y_c) \quad (7)$$

where $v(\tau) = (1 - \beta^\tau)/(1 - \beta)$ is the present value of a claim that pays one dollar a year for τ years under a discount factor $\beta \in (0, 1)$. The R&D induced by a patent is $r(\tau) = r(v(\tau)\pi(y_\pi))$. A traditional prize with free, unrestricted licensing after its award induces an allocation of R&D $r(\theta)$ with the competitive output y_c every year after the invention.

III. The Impact of Intellectual Property on Externality Remedies

Consider traditional interventions designed to solve the externality consumption problem that aim to maximize *ex-post* welfare. The output y_W that maximizes annual *ex-post* welfare W satisfies the necessary first-order condition:

$$W_y = 0 \quad \text{if and only if} \quad p = c_y - e_y \quad (8)$$

This simply says that the output level y_W may be obtained through Pigouvian corrections that align private incentives with social ones.

Conditional on a given size of a prize, the R&D undertaken is unaffected by *ex-post* efficiency. Therefore, if awards are used as methods to stimulate R&D, they impose no alterations for classic *ex-post* measures to handle externalities.

Conditional on a given patent length, here for easy exposition assumed infinite⁹, and the output y , the expected dynamic welfare is:

$$D = x(r(y))aW(y) - r(y) \quad (9)$$

where $a = v(\infty)$ is the present value of an infinite annuity and where $r(y) = r(a\pi(y))$ is the R&D induced by the patented profits.

The Pigouvian output generally does not induce the first-best dynamic allocation because the R&D induced by this output equals the first-best R&D level only when:

$$r(a\pi(y_W)) = r(aW(y_W)) = r^* \quad (10)$$

which implies $s(y_W) + e(y_W) = 0$. This never holds under a positive externality and, generally, never holds under a negative externality.

The Pigouvian output not only fails to induce the first-best allocation, but also fails to induce the second-best allocation given that patents are used as a method to stimulate

⁹ Similar arguments apply to any finite patent length.

R&D. The output y_D that maximizes second-best dynamic welfare satisfies the necessary first-order condition:

$$r_y [x_r a W - 1] + x(a W_y) = 0 \quad (11)$$

The first term represents the R&D effects of expanding output: the impact of the output on R&D times the net social value of that increase in R&D. This R&D effect must be balanced against the *ex-post* welfare effects of expansion. The dynamically optimal output y_D only corresponds to the *ex-post* efficient solution to the externality consumption problem, y_W , when the first term is zero. This is true when there is no under or over-investment in R&D socially. Such under or over-investment in R&D occurs when private rewards to R&D do not reflect social ones, in which case Pigouvian solutions are not optimal; $W \neq \pi$ implies $W_y \neq 0$.

This implication for the case of external effects, *i.e.*, that *ex-post* (static) efficiency through Pigouvian measures is inconsistent with *ex-ante* (dynamic) efficiency, is analogous to the case of goods with only private consumption effects. Without externalities, it is well understood that efficient competition *ex-post* leads to insufficient R&D incentives *ex-ante*, which is of course the common rationale for patents. With externalities, this has the important and unrecognized implication that Pigouvian corrections may often be inefficient under technological change. In most cases, arguing for Pigouvian solutions in the presence of technological change is analogous to arguing for competitive markets for new inventions (!) because both support *ex-post* efficiency without regards to R&D incentives.

Interestingly, the failure of Pigouvian solutions is not necessarily due to the fact that patents are second-best methods of stimulating R&D. To illustrate, consider when full-

price discrimination among consumers is feasible so that in the absence of externalities, the patent above would induce a first-best allocation. However, even in that case, patents are never first-best when there is an externality. This is because price discrimination only allows the firm to capture a *consumer* surplus, but not a surplus derived from external effects. This implies that under a positive externality, the monopolist always under-invests in R&D. Conversely, when the externality is negative, the producer may over-invest in R&D.

A. Dynamic Profit Corrections to Static Pigouvian Measures

The analysis above implies that previous remedies of externalities that consider only the static effects on welfare, W , may be incomplete. This especially occurs if dynamic welfare, D , changes in the opposite direction of static welfare as the remedy affects R&D incentives. Remedies may lower welfare *ex-post* but raise dynamic welfare when profits rise to encourage R&D. In general, the static analysis only concerns total *ex-post* welfare, as opposed to dynamic welfare that depends on the *incidence* of Pigouvian taxation, *i.e.*, on how the distributional impact of taxation affects producers and consumers separately.

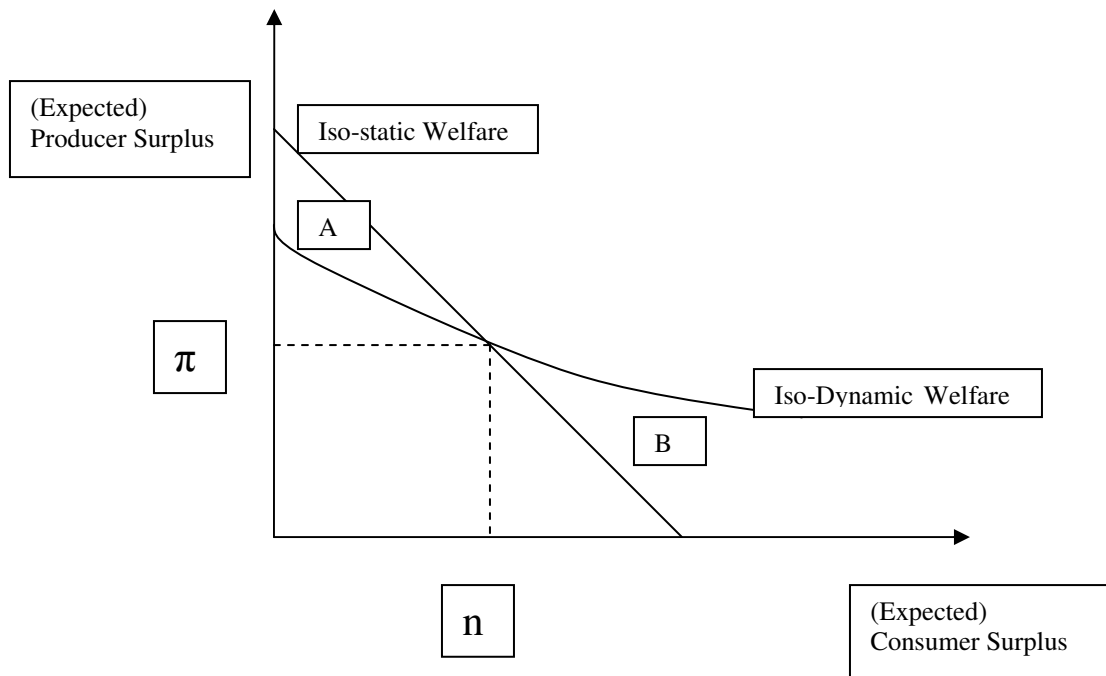
Consider changing the surplus levels π of the producer and $n = s + e$ of the non-producer surplus from an initial level (n, π) to the alternative levels (n', π') with a remedy that aims to correct an externality. As static welfare consists of the sum of the two surpluses, $W = n + \pi$, there is a one-to-one tradeoff between producer surplus and non-producer surplus in affecting static welfare. However, the tradeoff between the two in keeping dynamic welfare constant satisfies

$$dn/d\pi = -D_\pi/D_n = -1 - (1/x) \{ r_\tau [x_r W - 1] \}. \quad (12)$$

The first term, -1 , is the tradeoff between consumer and producer surplus keeping *ex-post* welfare constant. However, this tradeoff is tilted towards producer surplus when R&D is affected by the incidence of the welfare change i.e., by a factor representing the marginal social value of the R&D generated by the profits, $r_\pi[x_r W - 1]$, discounted by the chance of it occurring, $(1/x)$. Even with other models of R&D, the general point remains that the tradeoff will be influenced by how much profits affect R&D.

Figure 1 illustrates what determines the change in dynamic welfare resulting from a change in producer and non-producer surplus due to an externality remedy away from initial levels (n, π) . The straight line represents combinations of surpluses that keep classic *ex-post* welfare constant and thus has a unit slope. The curve depicts the combinations that keep dynamic welfare constant when profit changes affect R&D.

Figure 1: Static vs Dynamic Impact of a Pigouvian Remedy



When non-producer surplus falls by more than profits rise through the remedy, the situation is deemed inefficient by classic analysis, but may be dynamically efficient (region A in the figure). Conversely, when *ex-post* non-producer surplus rises more than profits fall by the remedy, this would be considered efficient by classical analysis, but it may be dynamically inefficient (region B in the figure). The figure shows not only that classic Pigouvian welfare calculations will produce quantitatively incorrect results, but also that their *qualitative* conclusions may be inaccurate; the *sign* of the static and dynamic welfare impacts may differ.

The previous discussion considered the optimal choice among two potential alternatives. To characterize the choice among all feasible alternatives, consider selecting surplus levels from a convex and well-behaved set C induced by a remedy w as in

$$C = \{ (n, \pi) \mid (n, \pi) = (n(w), \pi(w)), \quad w \in W \}. \quad (13)$$

This set would make up a utility-possibility frontier in Figure 1 from which surpluses could be selected. Static and dynamic welfare displayed a difference in tradeoffs between profits and non-producer surplus. This has the direct implication that, as long as the choice set concerns positive utility imputations, and C is a subset of $R_+ \times R_+$, *profitability will be higher under dynamic welfare than under static welfare*. More precisely, it can be shown¹⁰ that if (n_W, π_W) maximizes W and (n_D, π_D) maximizes D then it must be that $\pi_D \geq \pi_W \geq 0$ and $n_W \geq n_D \geq 0$.

¹⁰ To show this, assume the contrary, that is $\pi_W > \pi_D$. By definition, we have that $n_W + \pi_W \geq n_D + \pi_D$. Those two inequalities imply that both R&D and *ex-post* welfare are higher under (n_W, π_W) than under (n_D, π_D) . However, since we assume $n_W, n_D \geq 0$, we must have $W_D \geq \pi_D$ so that R&D is under-provided. Thus, there is a contradiction to the dynamic optimality of (n_D, π_D) .

Because expected profits equal $x(r(\pi))\pi - r(\pi)$, the envelope theorem directly implies that expected profits rise with *ex-post* profits. This in turn implies that maximizing dynamic welfare does not only induce higher *ex-post* profits but also higher expected profits than when maximizing static welfare. In sum, dynamic welfare involves a profit correction that reflects the impact of the remedies on technological change¹¹.

An example of the difference in *ex-post* and *ex-ante* efficiency concerns the temptation of governments to force R&D-returns down after an important innovation has been discovered and altruism dictates full adoption. For example, many observers have argued that a major barrier to R&D investments in an AIDS vaccine is that developers realize that if they are successful, governments will mandate full distribution of their products at below monopoly markups because it would be viewed inhumane not to.¹² Such policies would be efficient *ex-post* as the developer would lose less than consumers and altruists gained *ex-post*. However, such government theft of R&D would of course not be dynamically efficient as no vaccine would be developed anticipating the response. In fact, because rich altruists, rather than poor consumers, make up the majority of the surplus from such an innovation, the foregone R&D would be larger than under no externalities.

¹¹ An open question is whether the political process favors reduced profits and short-term Pigouvian solutions over dynamically efficient ones under technological change.

¹² A similar example is related to the recent increase in avian flu. Roche Pharmaceuticals, maker of Tamiflu (a recommended treatment for avian influenza), is facing significant pressure from several governments to allow generic distribution of its drug. While Tamiflu is still under patent, a number of Asian governments have threatened to bypass the patent and proceed with generic manufacturing if negotiated licensing fees are too high (Kanter, 2005).

IV. The Impact of External Effects on Intellectual Property

The previous section discussed the optimal way of intervening to solve externality problems given a certain form of IP. This section discusses the reverse problem of how to best design IP given the existence of externalities. We first consider the overall choice of IP in terms of prizes versus patents and then their individual design.

A. The Optimal Form of IP

The dynamic welfare under a patent of length τ is:

$$D(\tau) = x(r(\tau))W(\tau) - r(\tau) \quad (14)$$

Here, R&D is induced by patented profits, and *ex-post* welfare is induced by the output over time generated by the patent:

$$r(\tau) = r(v(\tau)\pi(y_\pi)) \quad (15)$$

$$W(\tau) = v(\tau)W(y_\pi) + [a - v(\tau)]W(y_c) \quad (16)$$

The dynamic welfare under a prize of size θ is:

$$D(\theta) = x(r(\theta))W(\theta) - r(\theta) \quad (17)$$

Here, R&D is simply that induced by the prize $r(\theta)$ and *ex-post* welfare as the present value of welfare induced by the constant output $y(\theta)$ after the prize has been awarded $W(\theta) = aW(y(\theta))$.

Prizes vs. Patents with no Externalities

Under no externalities, the optimal prize is the present value of the social surplus $\theta^* = aW(y_c)$ and thus implements the first-best allocation (r^*, y^*) . Without externalities, the

optimal prize always dominates the optimal patent, $D(\tau^*) \leq D(\theta^*)$.¹³ This is sometimes interpreted to mean that prizes dominate patents when there are no externalities, with the implicit assumption that the organizations selecting the prizes can set them correctly to represent social surplus. This is an assumption that many times may be unwarranted.

Although not previously recognized, this dominance of prizes under no external effects depends crucially on how *production and distribution* take place after the prize has been awarded. The implicit assumption of the method of production and distribution under a prize is that of *free and unrestricted licensing* of the patent after the discovery, hence generating the competitive output level: $y(\theta) = y_c$. Thus, prizes induce *ex-post* efficiency without externalities, $y(\theta) = y_w$, which is the major reason for their superiority over patents. However, under external effects, prizes with free licensing still induce a competitive level of output, but now this level is inefficient: $y(\theta) = y_c \neq y_w$.

Prizes vs. Patents with Externalities

The inefficient *ex-post* production under a prize with free licensing implies patents may dominate prizes under external effects. More precisely, we have the asymmetric result that when marginal cost is below the inverse demand:

$$\text{If } e \geq 0 \text{ then } D(\tau^*) < D(\theta^*) \quad (18)$$

$$\text{If } e < 0 \text{ then } D(\tau^*) \leq D(\theta^*)$$

The result may be interpreted to state that prizes tend to be more favored over patents the more positive the external effects are. The intuition behind this result is that the markup of a patent holder acts as a Pigouvian tax on *ex-post* output. This “patent-taxation” of

¹³ The exception is when the patent monopolists fully capture social surplus through price discrimination, in which case the optimal prize and optimal patent (infinite in length) yield the same dynamic welfare.

externalities implies that the traditional welfare loss associated with patents is reduced under negative externalities; in that case, the relative size of the elasticity of demand and the harm induced by the externality determine whether the patent monopolist under- or over-prices his output. Of course, under positive external effects, *ex-post* efficiency under the prize is higher than for patents so that prizes always dominate.

The case of third-world disease R&D is a useful illustration of how patents compare to both standard and non-standard prizes under positive external effects. For diseases present only in poor countries, consider when social surplus mainly consists of the external altruistic effects of developed nations. More precisely, consider when the inverse demand curve is below marginal costs, $p < c_y$, because consumers are too poor to be able to pay variable costs of production, let alone the fixed costs of R&D. This implies that the patented monopoly power does not confer any profits and thus implies no R&D spending and zero dynamic welfare: $\pi = r = D = 0$ for any patent length τ . Moreover, when the inverse demand lies below marginal cost, the difference between patents and prizes depends heavily on the method of production and distribution associated with the given prize. Under the *standard* mechanism for generating output associated with a prize, *i.e.*, free licensing, patents would, in fact, dominate any *positive* prize. This is because the R&D would be undertaken without distribution for any positive prize, while under a patent, the R&D would not occur.¹⁴ In this case, the optimal *standard* prize would be zero, consistent with zero *ex-post* welfare from the absence of output. Importantly, however, certain *non-standard* prizes guaranteeing a positive *ex-post* output would dominate patents here. This is because the patent holder can only capture consumer

¹⁴ This is a relevant issue for third-world disease R&D where a lack of profit motive for the distribution of existing innovations often seems of equal importance for the discovery of new ones. In general, excessive government funding for innovations that do not pass a market test also falls into this case.

surplus (which is zero when consumers cannot pay variable costs), while the prize holder can capture the non-consumer surplus contributing to welfare. Note that this dominance has little to do with the second-best nature of patents because even when they are first-best, as when the monopolist is allowed to fully price discriminate, they may still be dominated by prizes. Rather, the problem with patents under altruism is that the output is not sold to those willing to pay for it.

B. Optimal Prizes under External Effects

The way that production and distribution take place under a prize is non-trivial under external effects because free, unrestricted licensing does not induce *ex-post* efficiency. Given this shortcoming of the traditional form of prizes under external effects, we now analyze alternative prize contracts that may induce better production incentives *ex-post*.

Public Price Guarantees

Consider a public *price-guarantee* contract represented by a price level at which the public sector promises to purchase a given product if it is discovered. This makes the monopolist face a price that does not change with output. If we let $y(p)$ be the supply at a given fixed price p , then the profits under a given price-guarantee contract will be:

$$\pi(p) = py(p) - c(y(p)) \quad (19)$$

The price guarantee contract implements the first-best allocation if:

$$y(p) = y^* \text{ and } r(\pi(p)) = r^* \quad (20)$$

As the monopolist is faced with a “competitive” price that does not change with output, the optimal price that implements the first best allocation is the Pigouvian price

$$p = c_y - e_y. \quad (21)$$

It follows that the R&D investment is first-best only if $\pi(p)=W(y(p))$ which does not hold generically. A guaranteed price will not implement the first-best allocation because two conditions on the single optimal purchasing price are required, one implied by the cost structure or supply curve of the firm and the other by the size of the surplus generated by external effects and consumption. An illustrative special case is when there are constant returns to scale in production, in which case output will be infinite when the price is above marginal costs, but there will be no R&D if prices are at or below marginal costs. The two conditions imposed by a price guarantee cannot be met simultaneously for generic cases of preferences and technology so that a single measure is unlikely to solve the two aspects of the allocation problem.

Public Demand Contracts

Consider an award that consists of a committed *public demand contract* represented by the quantity and price (y,p) , *e.g.*, 100 million doses of a vaccine at the price of \$10 per shot. What does such a first-best purchasing contract (y^*, p^*) look like?

Clearly, the output level of the contract must equal the first-best level:

$$y^* = y_w \quad (22)$$

To examine what price induces the first-best level of R&D, let the *ex-post* profits obtained under the contract be denoted:

$$\pi(y,p) = py - c(y) \quad (23)$$

The contract yields the correct amount of R&D whenever:

$$r(\pi(y_w,p)) = r(W(y_w)) \quad (24)$$

This implies that the optimal contract price is:

$$p^* = c(y_W)/y_W + W(y_W)/y_W = [s_o(y_W) + e(y_W)] / y_W \quad (25)$$

where $s_o = \int p(q) dq$ is the aggregate consumer surplus when the good is given out free of charge under the public program. The optimal contract price is determined by the *average* social value of output which differs from the *ex-post* efficient Pigouvian price determined by the *marginal* social value of output: $p = c_y - e_y$. As the monopolist faces the social costs but not the social benefit of the production, only the revenue side of the producer tradeoffs has to be adjusted to have the R&D reflect the social benefit of the activity.¹⁵

The optimal purchasing contract illustrates the more general point that optimal pricing of external effects *ex-post* is not appropriate for optimal R&D incentives *ex-ante*.¹⁶ This may be exemplified by the case of constant returns: when the product is free, the consumer surplus satisfies $s_o(y) = ay - by^2/2$ so that the optimal purchasing contract satisfies:

$$y^* = y_W = (a - c + e) / b \quad (26)$$

$$p^* = [s_o(y_W) + e(y_W)] / y_W = a - (b/2)y_W + e = (a + c + e) / 2.$$

Note that the optimal Pigouvian price $p = c - e$ falls with the externality; as there are more benefits to non-consumers on the *margin*, more consumption should take place.

¹⁵ In the special case of no externalities, the optimal price is simply the average consumer surplus under the Pigouvian output level equating profits with *ex-post* welfare.

¹⁶ This simple result contrasts many alternative discussions of what sufficient R&D incentives should be for drugs in developing world nations; see, *e.g.*, Lanjouw (2002) or Sachs (2001). Many discussions argue they should be comparable with returns on drugs demanded by developed countries while the preferences or technology implying this claim is left unspecified. However, if such alternative investments reflect the share of consumer surplus captured by inventors in rich markets, they have no impact on optimal contract design as discussed here.

This is in contrast to the optimal contract price that rises with the externality because the price needs to reflect *average* consumer surplus to encourage innovation correctly.

Note that if the revenue received by the innovator under such a contract was simply awarded as a prize of size $\theta = p^* y^* = [s_o(y_w) + e(y_w)]$, then the first-best allocation would not be obtained because after the reward was received there would be no incentive for production and distribution. Public price guarantees or demand contracts generate an incentive for production and distribution.¹⁷

C. Optimal Patents under External Effects

When patents are optimal to use, how do external effects affect their design? The patent length that maximizes dynamic welfare D solves:

$$\text{Max } D(\tau) = x(r(\tau))W(\tau) - r(\tau) \quad (27)$$

This yields the necessary first-order condition for the optimal patent length:

$$r_\tau[x_\tau W - 1] = x(-W_\tau) \quad (28)$$

The left-hand side is the marginal benefit of extending a patent by one year. It is comprised of the marginal impact on R&D the extension has times the net social value of the increase in R&D. The marginal benefit is positive whenever there is under-investment in R&D. The right hand side is the marginal cost of extending the patent, which is made up of the expected increase in the welfare loss of a patent monopoly.

¹⁷ The efficient separation between R&D and production, as commonly solved by licensing, is not discussed here but needs to be better understood. For third-world drugs, innovative companies may be rewarded for their R&D after which generic manufacturers may produce and deliver.

The impact of external effects on patents is thus a result of their changes on the marginal costs and benefits of patent extension. Consider first the marginal cost of the patent extension related to its impact on *ex-post* welfare:

$$W_{\tau} = v_{\tau}[W(y_{\pi}) - W(y_c)] \quad (29)$$

When there are no externalities, this derivative is of course negative.

How do externalities affect the *ex-post* welfare change induced by a patent, and thus the marginal cost of patent extension? The markup of a patent holder acts as a Pigouvian tax and, therefore, a patent may be beneficial for *ex-post* efficiency under a negative externality, such as the antibiotic case. However, it is harmful for *ex-post* efficiency under a positive externality, such as the AIDS drug case.¹⁸

How is the marginal benefit of patent extension affected by externalities? The amount of R&D induced by a given patent length is $r(\tau) = r(v(\tau)\pi(y_{\pi}))$. Naturally, this implies R&D rises in the length of protection: $r_{\tau} > 0$. External effects do not have an impact on this effect; a patent extension raises R&D equally much regardless of the externality. The second factor in the marginal benefit of patent extension is the net social benefit of the additional R&D the patent extension induces, $x_{\tau}W - 1$. External effects have an indeterminate effect on this net gain in R&D.

If $\tau(e)$ denotes the optimal patent length given the externality, the implicit function theorem applied to the first-order condition of the optimal patent length $F(\tau, e) = dD/d\tau = 0$ yields:

$$d\tau/de = F_e / (-F_{\tau}) = [r_{\tau}x_{\tau}W_e + x_{\tau}W_{\tau e}] / (-F_{\tau}) \quad (30)$$

¹⁸ In other words, the traditional welfare loss associated with patents may not be present under negative externalities but is exaggerated under positive externalities. Indeed, in the case of negative external effects, the *ex-post* welfare function may well *rise* in patent length, $W_{\tau} > 0$, which would imply the corner solution of an optimally infinite patent.

Here, the denominator is necessarily positive as long as the second-order condition holds. This expression was obtained by using the fact that the optimal R&D level does not depend on the size of the externality: $r_e = 0$. As a consequence, the optimal chance of discovery does not depend on the externality: $x_e = x_{re} = 0$.

Evaluating the sign of $d\tau/de$, note that *ex-post* welfare rises with the externality simply because the more people enjoy the output, the larger the externality is $W_e > 0$. Thus, the first term is positive. Regarding the remaining second term, which depends on the sign of $W_{\tau e}$, we need to sign the impact the externality has on the marginal effect of raising the patent length. If the externality is positive, we know that extending the patent is harmful, $W_{\tau} < 0$. Furthermore, the larger the size of the positive externality the more harmful it is to extend the patent: $W_{\tau e} = v_{\tau} d[W(y_C) - W(y_{\pi})]/de < 0$.

Under such an externality, it therefore follows that raising the size of the externality has an ambiguous effect on the optimal patent length. A larger positive externality not only raises the social value of the invention, $W_e > 0$, but also increases the harm imposed by restricting its consumption through patents, $W_{\tau e} < 0$, making up two offsetting forces on the optimal patent life. If the externality is negative, an analogous argument applies.

V. Calibrating Altruism and Dynamic Inefficiency in R&D

Given the theoretical importance of missed altruistic surplus for underinvestment into health care R&D, in general, and HIV R&D in particular, this section calibrates the quantitative size of this dynamic inefficiency.

A. Calibrating the External Consumption Effect

The main assumption used for the calibration is that the observed public subsidization of a technology coincides with the ex-post Pigouvian solution to the problem of under-consumption under altruism. Generally, compared to the social optimum, the under-provision of any good may be due to several factors, including externalities in consumption as well as imperfect competition in the market in question. For the case of HIV, for which the majority of therapies remain on patent, we assume that the observed subsidy is the Pigouvian correction to both problems.¹⁹ Recall that ex-post social welfare is given by:

$$W(y) = \pi(y) + s(y) + e(y)$$

Suppose that for each unit sold, firms receive a per-unit subsidy δ in addition to the price consumers pay for that unit, $p(y)$. Social welfare can then be written as:

$$W(y, \delta) = [p(y) + \delta]y - c(y) + s(y) + e(y) - \delta y \quad (31)$$

where $e(y) - \delta y$ is the net altruistic surplus, $s(y)$ is the consumer surplus, and $[p(y) + \delta]y - c(y)$ is the producer surplus under the subsidy. For a patent monopolist, the optimal output in the presence of the subsidy is:

$$y(\delta) = \arg \max \pi(y; \delta) = \arg \max [(p(y) + \delta)y - c(y)] \quad (32)$$

The induced ex-post welfare is $W(y(\delta), \delta)$ and the ex-post optimal (Pigouvian) subsidy maximizes $W(y(\delta), \delta)$ with respect to δ .

¹⁹ Later, we change our assumption of monopoly structure to examine how the altruistic surplus implied by our model changes under perfect competition.

A.1 Parameterizing Altruism and Demand

We specify the external consumption effect $e(y)$ to take the following form:

$$e(y) = N\alpha \cdot s(y) \quad (33)$$

This specification captures the public-good nature of the external consumption effect. That is, each of N individuals in a society is assumed to value a fraction, α , of the consumer surplus. Moreover, altruism is a public good in the sense that each altruists' "consumption" does not preclude that by another. The net surplus enjoyed by altruists is the external consumption effect less the subsidy:

$$N \cdot [\alpha \cdot s(y(\delta)) - \frac{\delta}{N} y(\delta)] \quad (34)$$

Since each altruist pays only an N^{th} of the per-unit subsidy, as the number of altruists increases, the cost to each of subsidizing a given level of output decreases.²⁰

We assume a constant elasticity of demand $q = (\beta/p)^\varepsilon$, where $\varepsilon > 0$ is the elasticity of demand and β is a parameter that shifts demand outwards.

A.2 Optimal Subsidy

The ex-post Pigouvian subsidy is derived by maximizing the parameterized ex-post welfare $W(y(\delta), \delta)$ with respect to δ . The details of the maximization are presented in the Appendix. Under constant returns to scale and a constant elasticity of demand, it is straightforward to show that the optimal subsidy, demand price, and supply price satisfy:

²⁰ The increase in N , through its effect on the subsidy, will increase output. Specifically, note that the quantity demanded by consumers depends on the price they face which, in turn, depends on the subsidy. The lower per-person cost of subsidizing a given level of output will lead to an increase in the per-unit subsidy, δ , and consequently output. While possibly even leading to an overall increase in per-person costs ($\delta y/N$), per-person costs will certainly increase above the level that would prevail if N were to increase without any compensating changes in δ and y .

$$\begin{aligned}
\delta &= \frac{c(1+N\alpha)}{\varepsilon+N\alpha} \\
p_D &= \frac{c\varepsilon}{\varepsilon+N\alpha} \\
p_S &= p_D + \delta = \frac{c(1+\varepsilon+N\alpha)}{\varepsilon+N\alpha}
\end{aligned} \tag{35}$$

The optimal subsidy is increasing in both the degree of altruism, α , and the number of altruists, N . Note that the optimal subsidy in the presence of a monopolist is higher than that in perfect competition as the monopolist restricts output. Finally, note that while the prices paid by consumers and received by firms are decreasing in α and N , firm profits rise with the degree of altruism and the number of altruists.

Under the assumption that the *observed* subsidy is the ex-post Pigouvian solution to the problem of external consumption effects, the level of altruism will be identified through the optimality condition:

$$\frac{\delta}{p_S} = \frac{1+N\alpha}{1+N\alpha+\varepsilon} \tag{36}$$

Note that this condition implies that even in the absence of altruism, there is subsidization to correct the distortion induced by monopoly pricing.²¹ It is straightforward to show that under perfect competition, the analog optimality condition is:

$$\frac{\delta^c}{p_s^c} = \frac{N\alpha}{N\alpha+\varepsilon} \tag{37}$$

Under perfect competition, altruism is necessary for subsidization.

²¹ Moreover, small observed shares are consistent with a *negative* external consumption effect. Since the subsidy is designed to induce a socially optimal output, if output is observed to be below the level that would be socially optimal *in the absence of altruism*, it must be because there is a negative externality.

B. Calibration for HIV/AIDS

Philipson and Jena (2005a) estimate the consumer surplus, s , generated by the new HIV/AIDS technologies to be roughly \$395 billion since the start of the epidemic nearly 25 years ago. In the Appendix, we discuss the methods used to estimate the share of the price that is subsidized ($\delta/p_s=0.5$), the demand elasticity ($\epsilon=1.25$), and the size of the non-consumer pool ($N=190$ million annually). The demand elasticity is the most indirect parameter to be calibrated, for which we use existing patent expiration data to estimate markups of brands relative to generic competition, and hence demand elasticities. These quantities can then be used to identify α , the fraction of the aggregate consumer surplus enjoyed by a single altruist, for either market structure, as well as the aggregate, external value to non-consumers, $N\alpha$. For the case of HIV, the aggregate value to non-consumers is a quarter of the consumer surplus (i.e., $N\alpha=0.25$). From 1980 to 2000, this amounts to roughly \$99 billion under the estimated level of consumer surplus.²² It is important to note that the magnitude of this effect is driven by the public goods nature of the externality. To see this more clearly, note that the aggregate external consumption effect of \$99 billion amortized over 20 years is simply \$5 billion per year. With 190 million altruists enjoying this annually, the value of the externality amounts to \$26 per altruist per year! With an estimated \$3.25 billion spent on subsidies from 1980-2000 (50 percent of the \$6.5 billion total HIV/AIDS drug spending), this amounts to \$163 million spent *annually* by all altruists or 85 cents per altruist per year. Including these costs of

²² An alternative specification of the externality would be $e(y) = N\alpha y$, interpreted as altruists caring about the health of others rather than their welfare (as is true when $e(y) = N\alpha \cdot s(y)$). In this case, the share of the supply price that is subsidized, δ/p_s , is equal to $[cy + N\alpha y \cdot (\epsilon-1)]/[cy \cdot (\epsilon+1) - N\alpha y]$. If variable costs are 20 percent of sales, $cy = \$15$ billion; meanwhile, $\delta/p_s = 0.5$ and $\epsilon = 1.25$. Thus, the gross altruistic benefit ($N\alpha y$) is \$2.5 billion. In light of the \$99 billion predicted above, this result stresses the discrepancy between wrong but commonly accepted measure of welfare, namely health, and actual welfare.

subsidization leads to a *net* external consumption benefit of roughly \$25 (= \$26 - \$.85) per altruist per year.

C. Calibration for the US Health Care Sector

Recent estimates suggest that healthcare spending in the US has been quite valuable, with consumer benefits of four to five dollars for every dollar spent.²³ With nearly \$1.44 trillion spent on healthcare in 2003 alone, this suggests an annual consumer surplus of between \$4.3 and \$5.7 trillion arising from healthcare consumption. Given our earlier results for HIV/AIDS, this raises the question of how altruistic surplus compares to consumer surplus for the health care sector as a whole.

We can use our framework to inform this question. First, since the overall market for healthcare is more competitive than that for HIV/AIDS, we begin by assuming that firms behave competitively—in this case, the share of the supply price that is publicly subsidized (δ/p_s) equals $N\alpha/(\varepsilon + N\alpha)$. Second, we use the fact that Medicaid is the primary provider of subsidized health care in the US and in 2003, accounted for nearly 18 percent (\$254 billion) of personal health care spending.²⁴ We therefore assume $\delta/p_s = 0.18$, which implies $N\alpha = 0.22 \cdot \varepsilon$. If $\varepsilon = 1.25$, the aggregate value to non-consumers is 27 percent of consumer surplus, which is remarkably similar in magnitude to our estimate for HIV/AIDS. For consumer surpluses ranging from \$4.3 to \$5.7 trillion, this implies an altruistic surplus of \$1.1 to \$1.5 trillion in 2003 alone. This corresponds to a gross benefit

²³ See e.g., Cutler and McClellan (2001) and Philipson and Jena (2005b). Philipson and Jena develop a methodology to link observed estimates of cost-effectiveness to surplus appropriation by producers. In their examination of over 200 health care technologies, the median ratio of gross benefits to spending was nearly 5, in line with published estimates that consumers obtain \$4 - \$5 of benefits for every dollar spent.

²⁴ We exclude Medicare since its benefits presumably reward contributions made by beneficiaries throughout their working lives, rather than reflect purely altruistic motives on the part of the current young. Including Medicare would simply raise the estimated level of altruism.

to each altruist of \$5,800 to \$7,900 annually and a net benefit (gross benefit – cost of subsidy) of \$4,400 to \$6,500.

D. Implications for Underinvestment in R&D

Given the altruistic surplus implied by our model, we present back-of-the-envelope calculations on the degree of underinvestment into HIV R&D due to non-appropriation of this surplus. To do so requires two pieces of information: the amount of R&D to date and the expected increase in R&D if altruistic surplus were fully appropriated. For the former, Philipson and Jena (2005c) report \$16 billion (discounted to 1980 and in year 2000 dollars) worth of private R&D into HIV/AIDS to date. For the latter, we use estimates from Finkelstein (2003) that a one dollar increase in the expected discounted present value of market revenue from a particular vaccine induces 5 to 6 cents worth of investment into that vaccine.

With estimates of the altruistic surplus for HIV/AIDS around \$99 billion, this implies an underinvestment in R&D of \$5 billion. These figures suggest that fuller appropriation of non-consumer surplus would have increased R&D by 33 percent of R&D completed to date. Put differently, our figures suggest an underinvestment in R&D of roughly 23 percent. We can compute similar estimates for health care in general. In 2003, private health care R&D was nearly \$35 billion. With a predicted altruistic surplus of \$1.1 trillion in that year alone, this implies a potential increase in R&D of \$55 billion, suggesting an underinvestment into overall health R&D of nearly 61 percent.

VI. Concluding Remarks

This paper considers how dealing with externalities is affected by common forms of IP as well as the reverse problem of how IP design is affected by externalities, applying the results to health care markets under altruistic consumption. For the first problem of the effect of IP on externality remedies, we stress that although traditional Pigouvian measures are efficient ex-post, they do not generate the correct R&D-incentives ex-ante. For the second problem of optimal IP design in the presence of externalities, we discuss the optimal form of IP in terms of patents or prizes, as well as the design of each particular form. Our analysis is illustrated through health care markets, in which altruism often seems to induce public subsidization of the poor or frail, and in which technological change is so often thought to be a key determinant in the expansion of the relative size of this sector. In particular, we estimate the aggregate value non-consumers place on the consumption of HIV drugs in the U.S. to be nearly 25 percent of the patients' surplus, with similar estimates true for health care consumption generally. For the case of HIV/AIDS, using this surplus to stimulate investment could raise R&D by as much as 33 percent of total R&D to date.

An important area of research suggested but not fully explored by the discussion above points to more elaborate evaluations of proposals to stimulate R&D into many prevalent third-world diseases. Without externalities, it seems efficient that a disproportionate low share of the world R&D spending on drugs is allocated to third-world diseases even though these diseases may be more prevalent and clinically more devastating world-wide. Altruism makes it an externality/R&D problem among rich

nations. However, existing policy proposals²⁵ to deal with this implicit externality problem have been *ad hoc* in the sense that it is not clear which allocation problems are underlying the proposed solutions. Examples include Sachs *et al.* (2001) who advocate cost-based pricing financed by donor countries or Lanjouw (2002) who advocates cost-based pricing through competition rather than regulation, through country- and disease-specific cut-backs in IP rights.²⁶

One may suspect a basic conflict between these policy proposals and an efficient provision of R&D under altruism as they reduce the benefits to innovators when those benefits should be increased rather than decreased to reflect the value to non-consumers.²⁷ In a sense, for exclusively third-world diseases, where demand curves are below variable costs, R&D is done for the rich countries, not for the poor! Therefore, as our calculations did not include the cross-country altruism in the case of HIV, the share of social surplus not accrued by consumers is under-estimated.

Related to this problem, the provision of AIDS drugs in poor countries mimics the problem of providing drugs for rare diseases in the U.S., as well as against agents of bio-

²⁵ Some proposals even demand that shareholders of innovative firms not only fund R&D to discover new treatments, but by reducing prices also cover the bill to satisfy the altruistic desires of the tax base.

²⁶ See also Grossman and Lai (2002) who discuss IP protection across countries.

²⁷ At the root of many proposals to restrict the IP rights of the pharmaceutical industry in developing countries is the argument that PhRMA lobbies governments and the WTO to force poor countries to give up their rights to freely copy molecules patented in the west under TRIPs. Consequently, some argue that pharmaceutical companies should compensate poor countries for this hijack. Such critiques resulted in the 2001 Doha Declaration which makes it legal for a country to produce drugs without the consent of the patent owner in case of public health crises. If we consider drugs that already have a profitable market in rich countries but that poor countries cannot afford, notwithstanding the threat of illegal parallel re-imports, the idea of cost based pricing in poor countries for existing drugs may be defensible on welfare grounds. However, an extreme consequence of that reasoning is the prevalent argument that pharmaceutical companies should not capture the altruistic benefits of people in rich countries for drugs that would be mostly or even exclusively targeted towards the developing world (!)

terror,²⁸ and it seems that international lessons can be learned from this domestic experience. With the purpose of stimulating R&D into disease classes too rare to generate R&D, the U.S. Orphan Drug Act of 1983 both reduced the cost and raised the benefit of R&D for such rare diseases.²⁹ If a society cares or wants to provide insurance for those who are unlucky enough to catch uncommon diseases, the social surplus will in addition to consumer surplus contain non-consumer benefits. The Orphan Drug Act may be interpreted to encourage R&D to reflect altruism, as opposed to international proposals for developing world diseases that discourage R&D in spite of such altruism. The enormous growth in drugs for rare diseases generated by the Orphan Drug Act may contain important lessons for the appropriate international policy.

In addition, the important issue of how world R&D should be financed across countries seems to fall under the discussed allocation problem. Many discussions of whether the U.S. is carrying too large a load of financing world drug R&D centers on the fact that about half of world sales are obtained in the unregulated markets of the U.S., with other price-regulated markets free-riding on the R&D investments this yields.³⁰ The non-exclusivity induced by the free flow of innovations across countries, and the desire to free ride due to that non-exclusivity, entails a classic externality or public goods problem in the consumption ex-post, with the additional feature of involving technological change as analyzed here.

²⁸ In the US, the legislation BioShield authorized \$5.6 billion over 10 years for the government to purchase vaccines and drugs to fight anthrax, smallpox and other potential agents of bio-terror.

²⁹ For a description of the main features of the act, see www.fda.gov/orphan. Also see Grabowski (2003) for a related but independent discussion.

³⁰ Becker *et al.* (2005) discuss the potential impact the sharing of the benefits of medical R&D across rich and poor countries has had on reducing world inequality.

Finally, our analysis assumes a particular structure of the R&D process, one which does not take into account the possibility of excessive R&D due, for example, to so-called “patent races.” Without altruism, these models predict that full appropriation of consumer surplus leads to overinvestment in R&D. Under such circumstances, non-appropriation may enhance efficiency by taxing the over-provision of R&D. Under altruism, however, full appropriation of consumer surplus may still, in fact, lead to underinvestment. This would be the case if non-consumer surplus were large relative to consumer surplus, which seems true for many third-world diseases.

In general, future research may fruitfully address the design of optimal externality and IP measures in health care and other areas. In order to achieve first-best allocation, one needs to break the link between ex-ante R&D and ex-post output provision.³¹ A single instrument is not sufficient to appropriately control both R&D incentives ex-ante and externalities ex-post. Appropriate policy must *simultaneously* solve the externality problem ex-post and the R&D problem ex-ante. A much better understanding of how this is done in practice is needed, particularly to understand whether the growth in health care spending observed worldwide is economically efficient.

References

- Anderson, R.M. and R.M. May. 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press.
- Becker, G., T. Philipson, and R. Soares 2005. The rise in longevity and world inequality. *American Economic Review*, March 2005.
- Cutler, D. and M. McClellan. 2001. Is technological change in medicine worth it? *Health Affairs* 20(5):11-29.

³¹ In this paper, we do not discuss whether public versus private production and financing of R&D would come closer to implementing the “ideal” first-best policy, in particular how asymmetric information affects the optimality of such choice (see Wright, 1983).

- Finkelstein, A. 2003. Health Policy and Technological Change: Evidence from the Vaccine Industry. NBER Working Paper No 9460, Cambridge, Mass.
- Grabowski, H. 2003. Increasing R&D incentives for neglected diseases: Lessons from the orphan drug act. Working Paper. Duke University
- Grossman, G. and E. Lai. 2002. International protection of IP. NBER Working Paper # 8704.
- Jones, C. and J. Williams. 2000. Too much of a good thing? The economics of investment in R&D. *Journal of Economic Growth* 5:65-85.
- Kanter, J. Roche Offers to Negotiate on Flu Drug, The New York Times, October 19th, 2005.
- Kremer, M. 2002. Pharmaceuticals and the developing world. *Journal of Economic Perspectives* 16 (4):67-90.
- Kremer, M. and R. Glennester. 2004. *Strong Medicine: Creating Incentives for Pharmaceutical Research on Neglected Diseases*. Princeton, NJ: Princeton University Press.
- Laffont, J.-J. 1987. Externalities. In *The New Palgrave: A Dictionary of Economics*, ed. J. Eatwell, M. Millgate and P. Newman. London and Basingstoke: Macmillan.
- Lanjouw, J. 2002. A patent policy for global diseases: US and international legal issues. *Harvard Journal of Law & Technology* 16(1):85-124.
- Newhouse, J.P. 1992. Medical care costs: How much welfare loss? *The Journal of Economic Perspectives* 6 (3):3-21.
- Nordhaus, W. D. 1969 *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change*, Cambridge: MIT Press.
- Nordhaus, W. D. 1972. The optimum life of a patent: Reply. *American Economic Review* 62 (3):428-431.
- Pigou, A.C. 1932. *Economics of Welfare*, 4th ed. London: Macmillan & Co.
- Parry, I., W. Pizer, and C. Fischer. 2000. How important is technological innovation in protecting the environment? Working Paper, *Resources for the Future*.
- Philipson, T. 2000. Economic epidemiology and infectious disease. In *Handbook of Health Economics*, ed. J. Newhouse and T. Culyer. New York, NY: North-Holland.

- Philipson, T. and S. Mechoulam. 2003. IP, R&D, and external consumption effects. NBER Working Paper No 9598, Cambridge, Mass.
- Philipson, T. and A. B. Jena. 2005a. Who Benefits from New Medical Technologies? Estimates of Producer and Consumer Surpluses for HIV/AIDS Drugs. Forthcoming, *Forum for Health Economics & Policy*. Berkeley Electronic Press.
- Philipson, T. and A. B. Jena. 2005b. Surplus Appropriation from R&D and Health Care Technology Assessment Procedures. Mimeo, University of Chicago.
- Philipson, T. and A. B. Jena. 2005c. Innovation and Technology Adoption in Health Care Markets. Mimeo, University of Chicago.
- Sachs, J. *et al.*, 2001. *Macroeconomics and Health: Investing in Health for Economic Development*, WHO, United Nations.
- Scotchmer, S. 2005. *Innovation and Incentives*. Cambridge, MA: The MIT Press.
- Weinstein M.C. 1995. From cost-effectiveness ratios to resource allocation: Where to draw the line? In *Valuing Health Care*, ed. F.A. Sloan, 77-97. New York: Cambridge University Press.
- Wright, B.D. 1983. The economics of invention incentives: Patents, prizes, and research contracts. *American Economic Review* 73 (4):691-707.

Mathematical Appendix

We assume constant returns to scale (constant marginal cost c) and constant elasticity of demand, $p(q) = \beta/(q^{1/\epsilon})$. The social welfare maximization is:

$$Max_{\delta} W(y) = \int_0^y p(q) dq - c \cdot y + N\alpha \left[\int_0^y p(q) dq - p(y) \cdot y \right] \quad \text{s.t. } y = y(\delta) \quad (A1)$$

where $y(\delta) = \arg \max_y [(p(y) + \delta)y - c \cdot y]$ describes the monopolist's optimal response to a subsidy δ . Note that $p(\cdot)$ is the price paid by the consumer and δ is the per-unit subsidy received by the monopolist above and beyond the price paid by the consumer. Under our assumptions on demand and production, it is straightforward to show that the monopolist-induced demand price and output satisfy:

$$p(\delta) \equiv p(y(\delta)) = \frac{(c - \delta)\epsilon}{\epsilon - 1}, y(\delta) = \left[\frac{\beta(\epsilon - 1)}{(c - \delta)\epsilon} \right]^{\epsilon} \quad (A2)$$

We can rewrite the maximization in A1 as follows:

$$Max_{\delta} W(y(\delta)) = \int_0^{y(\delta)} p(q) dq - c \cdot y(\delta) + N\alpha \left[\int_0^{y(\delta)} p(q) dq - p(y(\delta)) \cdot y(\delta) \right] \quad (A3)$$

Recalling that $p(\delta) \equiv p(y(\delta))$, the first order condition with respect to δ is:

$$[1 + N\alpha] \cdot p(\delta) \cdot \frac{dy(\delta)}{d\delta} - [c + N\alpha \cdot p(\delta)] \cdot \frac{dy(\delta)}{d\delta} = y(\delta) \cdot N\alpha \cdot \frac{dp(\delta)}{d\delta} \quad (A4)$$

which can be simplified to:

$$[p(\delta) - c] \cdot \frac{dy(\delta)}{d\delta} = y(\delta) \cdot N\alpha \cdot \frac{dp(\delta)}{d\delta} \quad (A5)$$

Since by definition, $dp(\delta)/d\delta$ can be rewritten as $dp(y(\delta))/d\delta$, by the Chain Rule, we obtain:

$$\frac{dp(\delta)}{d\delta} \equiv \frac{dp(y(\delta))}{d\delta} = \frac{dp(y(\delta))}{dy} \cdot \frac{dy(\delta)}{d\delta} \quad (A6)$$

Using A6, we can rewrite A5 as follows:

$$[p(\delta) - c] = y(\delta) \cdot N\alpha \cdot \frac{dp(y(\delta))}{dy} \quad (A7)$$

which, under constant elasticity of demand, can be written as:

$$[p(\delta) - c] = -\frac{p(\delta) \cdot N\alpha}{\varepsilon} \quad (A8)$$

Using the expression for $p(\delta)$ in A2, we can solve A8 to obtain the optimal subsidy δ as well as the demand price p_D (recall that this is equal to $p(\cdot)$) and supply price p_s (note, $p_s = p_D + \delta$).

$$\delta = \frac{c(1 + N\alpha)}{\varepsilon + N\alpha}, p_D = \frac{c\varepsilon}{\varepsilon + N\alpha}, p_s = \frac{c(1 + \varepsilon + N\alpha)}{\varepsilon + N\alpha} \quad (A9)$$

Using A9, we obtain an expression relating the share of total expenditure on drugs that is publicly subsidized (δ/p_s) to the level of altruism and the elasticity of demand.

Specifically,

$$\frac{\delta}{p_s} = \frac{1 + N\alpha}{1 + N\alpha + \varepsilon} \quad (A10)$$

Finally, we can calculate the ratio of profits to social welfare as follows:

$$\begin{aligned} \frac{\pi}{W} &= \frac{\left[\frac{c}{\varepsilon + N\alpha} \right] \cdot y(\delta)}{\int_0^{y(\delta)} p(q) dq - c \cdot y(\delta) + N\alpha \left[\int_0^{y(\delta)} p(q) dq - p(\delta) \cdot y(\delta) \right]} \\ &= \frac{\left[\frac{c}{\varepsilon + N\alpha} \right] \cdot y(\delta)}{\frac{\varepsilon\beta}{\varepsilon - 1} \cdot (y(\delta))^{\frac{\varepsilon-1}{\varepsilon}} - c \cdot y(\delta) + N\alpha \left[\frac{\varepsilon\beta}{\varepsilon - 1} \cdot (y(\delta))^{\frac{\varepsilon-1}{\varepsilon}} - \frac{c\varepsilon}{\varepsilon + N\alpha} \cdot y(\delta) \right]} \end{aligned}$$

$$\begin{aligned}
&= \frac{\left[\frac{c}{\varepsilon + N\alpha} \right]}{\frac{\varepsilon\beta}{\varepsilon-1} \cdot (y(\delta))^{\frac{-1}{\varepsilon}} - c + N\alpha \left[\frac{\varepsilon\beta}{\varepsilon-1} \cdot (y(\delta))^{\frac{-1}{\varepsilon}} - \frac{c\varepsilon}{\varepsilon + N\alpha} \right]} \\
&= \frac{\left[\frac{c}{\varepsilon + N\alpha} \right]}{\frac{\varepsilon\beta}{\varepsilon-1} \cdot \frac{p_D}{\beta} - c + N\alpha \left[\frac{\varepsilon\beta}{\varepsilon-1} \cdot \frac{p_D}{\beta} - \frac{c\varepsilon}{\varepsilon + N\alpha} \right]} \\
&= \frac{\left[\frac{c}{\varepsilon + N\alpha} \right]}{\frac{\varepsilon}{\varepsilon-1} \cdot \frac{c\varepsilon}{(\varepsilon + N\alpha)} - c + N\alpha \left[\frac{\varepsilon}{\varepsilon-1} \cdot \frac{c\varepsilon}{(\varepsilon + N\alpha)} - \frac{c\varepsilon}{\varepsilon + N\alpha} \right]} \\
&= \frac{1}{\frac{\varepsilon^2}{\varepsilon-1} - (\varepsilon + N\alpha) + N\alpha \left[\frac{\varepsilon^2}{\varepsilon-1} - \varepsilon \right]} \\
&= \frac{(\varepsilon-1)}{\varepsilon^2 - (\varepsilon + N\alpha)(\varepsilon-1) + N\alpha[\varepsilon^2 - \varepsilon(\varepsilon-1)]}
\end{aligned}$$

$$= \frac{(\varepsilon-1)}{\varepsilon + N\alpha} \quad (\text{A11})$$

Note that the share of social surplus appropriated to producers is positive since the monopolist operates in the elastic portion of the demand curve ($\varepsilon > 1$).

Data Appendix

This Data Appendix describes how the following are obtained: 1) the share of the price of HIV/AIDS drugs that is publicly subsidized, 2) the elasticity of demand, and 3) the number of non-consumers (i.e. altruists). These, along with consumer surplus measures obtained from Philipson and Jena (2005a), are used to calibrate our model.

Consumer Surplus from HIV/AIDS Drugs

Using the methodology developed in Becker, Philipson, and Soares (2005), Philipson and Jena (2005a) estimate the value of increased survival attributable to HIV/AIDS drugs. For each cohort infected with HIV, the authors estimate the aggregate value of improved survival relative to a benchmark in which no treatment was available. They repeat this for each new set of cases, cohort by cohort, since the start of the epidemic and aggregate up. This delivers the gross value to consumers of improved survival induced by HIV/AIDS therapies. The consumer surplus is obtained by netting out total spending, which is described below.

Financing of HIV/AIDS Drugs

The majority of public spending on HIV/AIDS drugs is administered through two sources, Medicaid and the AIDS Drug Assistance Program (ADAP). To be eligible for Medicaid, individuals must be low-income and in one of several mandated categories. Many AIDS patients qualify for Medicaid by being recipients of Supplemental Security Income (one of the mandated categories). These individuals are both low-income and disabled (Kates and Wilson, 2004).³²

The AIDS Drug Assistance Program began shortly after the introduction of AZT in 1987. Since 1990, ADAP has been part of the Ryan White CARE Program, the third largest federal source for care of HIV/AIDS patients. Since 1996, Congress has specifically designated funds for ADAP through the CARE program. ADAP is a payer of last resort for prescription medications needed by those without insurance or other means to finance drug treatment. In 2001 alone, an estimated 135,000 individuals received assistance from ADAP.

Figure 1 presents estimates of national spending on HIV/AIDS drugs broken down by public and private payers. The estimates for total spending are from IMS Health.³³ Public spending is approximated by the sum of Medicaid and ADAP expenditures. The Medicaid estimates include both federal and state contributions and were calculated from the Medicaid State Drug Utilization Data using National Drug Codes (NDC) for all antiretrovirals introduced since 1987.³⁴ Medicaid expenditure on HIV/AIDS drugs is unavailable prior to the last quarter of 1991—this is likely because Medicaid began its Prescription Drug Rebate Program (for all drugs, not just anti-

³² Eligibility for SSI requires an income below 74 percent of the Federal Poverty Line (FPL). In 2004, this amounted to an annual income of nearly \$7,000.

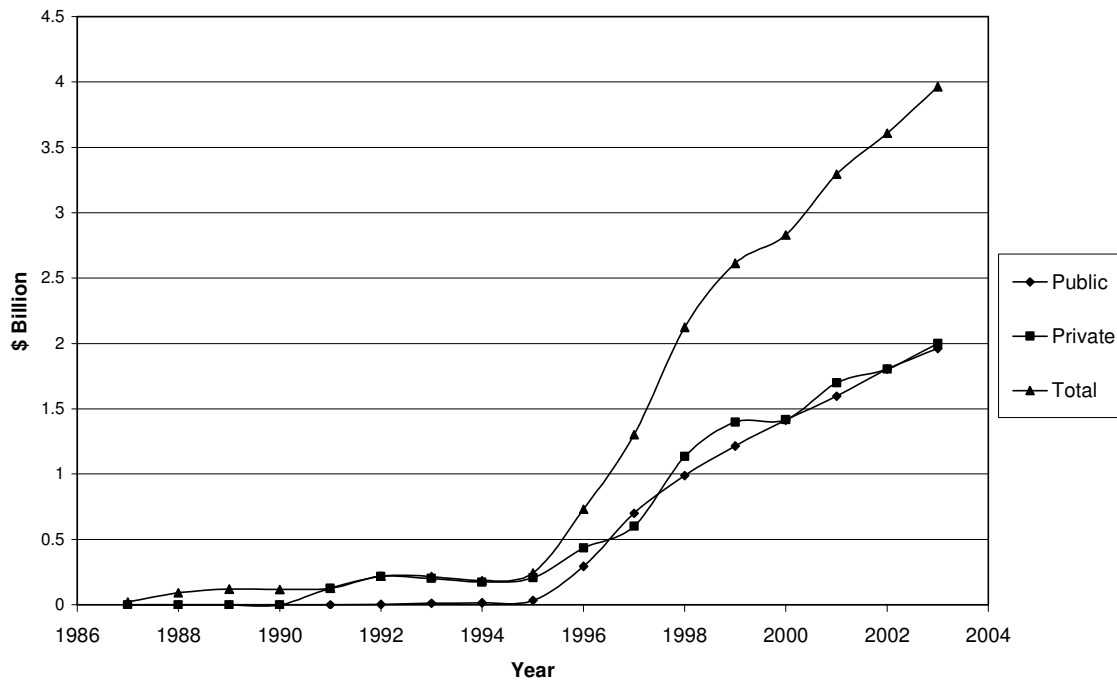
³³ Lichtenberg, 2005.

³⁴ <http://www.cms.hhs.gov/medicaid/drugs/drug5.asp>

retrovirals) only in 1990.³⁵ Data on ADAP expenditures are unavailable prior to 1996, though it was informally covering some individuals through the Ryan White CARE Program prior to that.³⁶

Since 1995, total spending has increased from \$250 million to almost \$4 billion, largely due to increased spending on protease inhibitors and nucleoside reverse transcriptase inhibitors. Figure 1 also demonstrates the large share of total spending on HIV/AIDS drugs financed by public sources, nearly 50 percent from 1996 onwards. Based on the above data, we parameterize δ/p_s to equal 0.5.

Figure 1: National Spending on HIV/AIDS Drugs



Elasticity of Demand and the Number of Altruists

We use the familiar monopolist mark-up condition, $(p-c)/p = 1/\epsilon$ to provide an estimate of the elasticity of demand for HIV/AIDS drugs.³⁷ Using estimates from the literature on the prices of generic drugs relative to their branded counterparts, we assume variable costs to be no more than 20 percent of sales.³⁸ With constant returns to scale in variable costs, marginal cost is constant and equal to variable cost. This suggests, $(p-c)/p = .8$ or alternatively that $\epsilon = 1.25$.

³⁵ Key Milestones in CMS Programs, <http://www.cms.hhs.gov/about/history/milestones.asp>

³⁶ Through communication with Kaiser Family Foundation.

³⁷ Since the monopolist only produces in the elastic portion of the demand curve, ϵ is bounded from below by unity.

³⁸ "Patent Expiration, Entry, and Competition in the U.S. Pharmaceutical Industry," R. Caves, M. Whinston, M. Hurwitz, *Brookings Paper on Economic Activity, Microeconomics* (1991): 1-66. We use the price of generic drugs as an upper bound of the marginal costs of production. The authors estimate that with 20 generic competitors, the ratio of prices between generic and brand drugs is roughly 20 percent.

We assume the number of altruists financing HIV drug consumption, N , to equal 190 million annually. This is the average number of adults alive in the US each year from 1980 to 2000. While this figure does not reflect the annual number of tax-payers in the US, it does partly capture non-working individuals in households who also benefit from the external consumption effect. Note that our choice of N will not alter the *aggregate* value altruists place on consumer surplus—it simply affects our estimates of the per-altruist external consumption benefit.