

Philipson, Tomas J.; Mechoulan, Stéphane

Working Paper

Intellectual Property & External Consumption Effects: Generalizations From Health Care Markets

Working Paper, No. 198

Provided in Cooperation with:

George J. Stigler Center for the Study of the Economy and the State, The University of Chicago Booth School of Business

Suggested Citation: Philipson, Tomas J.; Mechoulan, Stéphane (2005) : Intellectual Property & External Consumption Effects: Generalizations From Health Care Markets, Working Paper, No. 198, The University of Chicago, George J. Stigler Center for the Study of the Economy and the State, Chicago, IL

This Version is available at:

<https://hdl.handle.net/10419/262600>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

INTELLECTUAL PROPERTY & EXTERNAL CONSUMPTION EFFECTS: GENERALIZATIONS FROM HEALTH CARE MARKETS ¹

Tomas J. Philipson²

The University of Chicago

and

Stéphane Mechoulan

The University of Toronto

Date: April 4, 2005

¹ 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 2003, Conference on Antibiotic Resistance at Harvard 2000, as well as from Michael Baker, Gary Becker, John Grana, Robert McMillan, Kevin Murphy, and Gilles Saint-Paul. The paper has also benefited from discussions of many related topics with colleagues at FDA, CMS, and The Pull Mechanism Working Group of the Center for Global Development, Washington DC who, however, should not be held responsible for any views contained. Philipson is thank for financial support from Pfizer Inc as well as the George Stigler Center for the Study of the Economy and the State, and the John M. Olin Law & Economics Program, both at the University of Chicago.

² Corresponding Author: Tomas Philipson, Irving B. Harris Graduate School of Public Policy Studies, The University of Chicago, 1155 E. 60th Street, Chicago, IL 60637, Email: t-philipson@uchicago.edu.

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 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 external effects through human rights-based access issues.

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 technological change for goods that only have private consumption effects, *e.g.*, the analysis of the welfare effects of patent regulations.³ However, this classic problem assumes that there are no external effects in the consumption of the good for which there is technological change. Although both these problems are well analyzed, the problem of dealing with both technological change and external consumption effects remains less explored.

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 the problem. This has been particularly true for many policy issues facing the pharmaceutical industry, which is one of the most R&D-intensive industries and confronted with altruistic or human rights-based access issues. For example, consider the case of antibiotic resistance in which there has been great pressure to limit usage in order to slow down the rising threat of many antibiotics becoming useless against many life-threatening diseases. Such negative external effects,

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).

induced by current consumption lowering the value of future consumption,⁴ has prompted what may be interpreted as Pigouvian-like measures of taxing or limiting the demand for antibiotics. However, such classic remedies for externality problems discourage R&D into new antibiotics that will replace those to which bacteria have become resistant. Therefore, the dynamic costs of limiting the use of antibiotics may dominate the static 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 with these diseases in the world but cannot afford the costs of new drugs. As it appears that richer, developed countries, whether for selfish or altruistic reasons, care about expanding the access to drugs for third-world countries for diseases such as AIDS, this problem appears to be one of efficiently providing both technological change and consumption under positive external effects.

More generally, as many observers have argued that technological change is the key to the continued expansion of the health care sector in the economy,⁵ one may also argue that the joint allocation problem studied here is perhaps *the* central allocation problem for health care industries more broadly.⁶ Because it appears that many developed nations have decided that it is intolerable to let people die or suffer when existing medical

⁴ For the purpose of this paper, we will assume that these negative external effects dominate the classic positive external effects of treatments for infectious diseases, see *e.g.*, Philipson (2000).

⁵ See *e.g.*, Newhouse (1992).

⁶ Many other industries appear to have similar issues. For example, industries for which output is used as inputs to externality-generating R&D, “research tools” industries, industries with network-, peer-group-, or herd-effects, industries in which production induces pollution (see Perry et al (2000)), all seem to involve similar issues of balancing externalities *ex-post* with R&D incentives *ex ante*.

technologies can prevent it, public financing often covers such technologies. Such altruistic adoption of new technologies *ex-post* needs to be balanced against the technological change it induces. Many Pigouvian-style measures suggested, *e.g.*, through so called “cost-effectiveness” criteria in public adoption of new technologies, do not take into account the effects on technological change.

This paper derives the optimal treatment of externalities and IP 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. Without externalities, this is of course the rationale for patents with their associated *ex-post* inefficiencies. A simple, but unrecognized, analogous implication under consumption externalities is that Pigouvian solutions may often be inefficient under technological change. In many cases, arguing for Pigouvian solutions in the presence of technological change is analogous to arguing for competitive markets for new inventions (!) as both argue for *ex-post* efficiency without regard to dynamic R&D incentives.

Second, we discuss the reverse problem of the impact that externalities have on the design of IP. Regarding the optimal form of IP, external effects influence whether patents, potentially capturing consumer surplus, or prizes, potentially capturing those externally affected by consumption, best reward innovation. When there are positive external effects, such as a result of altruism in health care delivery, rewards to innovation should not be guided by the potential consumer surplus as it is with patents, but by the entire social surplus that includes the benefits to those externally affected by

consumption. In the extreme case of third-world diseases, consumers often cannot pay above variable cost, which means patents are of little value to innovators. However, if prizes are used, we argue that the way that efficient *production and distribution* takes place under a prize is non-trivial under external effects because free and unrestricted licensing does not induce *ex-post* efficiency. The competition induced by free and unrestricted licensing after the prize has been awarded does not generate the efficient output. Given these shortcomings of the traditional forms of prizes under external effects, we analyze alternative prize contracts that may induce better production and distribution incentives *ex-post*.

We discuss how externalities not only affect the type of IP selected but also how a given individual type of IP, prizes or patents, is efficiently designed in the presence of externalities. Optimal patents may be infinite under external effects, and optimal prizes will reward the *average*, as opposed to the *marginal*, output produced. The discussion has strong efficiency implications for using so called “cost-effectiveness” criteria used for public technology adoption in health care, as is common in other rich countries and often advocated for the U.S.

The paper is related to several literatures. It is of course related to the voluminous literature on the appropriate methods of treating externalities which is, in fact, too voluminous to usefully discuss here (see Laffont, 1987 for a survey). The paper also extends past and recent work on the optimal forms of push-and-pull stimuli including patents, prizes and research contracts (see Nordhaus, 1969 and 1972; Wright, 1983; Kremer 2002, Kremer and Glennester, 2004; Scotchmer, 2005). However, previous work does not appear to have explicitly addressed the impact of external consumption effects.

The paper may be briefly outlined as follows. Section II sets up the allocation problem of 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. We first consider the overall choice of prizes versus patents and then their individual design. Finally, Section V summarizes our conclusions.

II. External Effects and IP

Let y denote the quantity of output, $p(y)$ denote the private inverse demand curve of consumers, $e(y)$ 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 .

The feasible level of technological change in the good consumed is represented by letting $x(r)$ 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:

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

We assume that the chance of discovery $x(r)$ is a differentiable, increasing, and concave function, which implies 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 and output 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 that highest level of *ex-post* welfare $r^* = r(W(y_w))$. Naturally, as the first-best allocation selects from the unrestricted feasible set $\{(r, y): r \geq 0, y \geq 0\}$ that contains any restricted sets selected from under a second-best allocation, the expected welfare is larger in the former case.

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 that of the patented profits $r(\tau) = r(v(\tau)\pi(y_\tau))$. A traditional prize with free and 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 IP on Externality Interventions

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 efficient output involves prices above or below marginal costs depending on whether the externality is negative or positive. 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, Pigouvian measures that maximize *ex-post* welfare also maximize dynamic welfare as 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 infinite in length (similar arguments apply to any finite patent length), 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 R&D. The output y_D that maximizes second-best dynamic welfare satisfies the necessary first-order condition:

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

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 our 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 many cases, arguing for Pigouvian solutions in the presence of technological change is analogous to arguing for competitive markets for new inventions (!) because both argue for *ex-post* efficiency without regards to R&D incentives.

It is important to recognize that 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 by patent holders so that in the absence of externalities, the patent above would induce a first-best allocation since social surplus would coincide with producer surplus under an optimally infinite patent in that case. However, even under first-degree price discrimination, patents are never first-best when there is an externality. This is because price discrimination among the consumers of the product sold 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 because the consumer surplus captured, even though completely captured, is less than the social surplus. Conversely, when the externality is negative, the producer may over-invest in R&D. The problem is that although the producer captures the entire consumer surplus, consumers are not the only ones affected.

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 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 distributional impact affects producers and consumers separately.

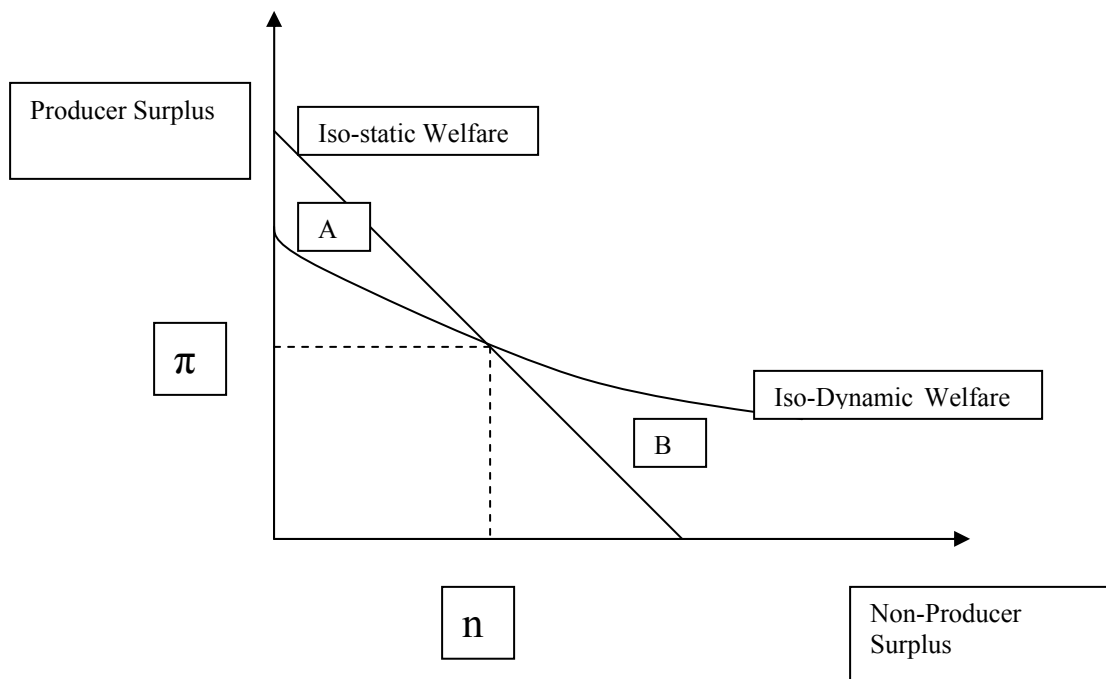
Consider changing the surplus levels π of the producer and $n = s + e$ of the non-producer surplus from an initial level (n, π) to the alternative levels (n', π') due to a remedy that aims to correct an externality. As static welfare consists of the sum of the two surpluses, $W = n + \pi$, there is a 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. Intuitively, the tradeoff is tilted towards profitability by a factor representing the marginal social value of the R&D generated by the profits, $r_\tau[x_r W - 1]$, discounted by the chance of their occurring, $(1/x)$. Even with other models of the R&D process, the general point remains that the tradeoff will be influenced by how much profits affect R&D.

Figure 1 depicts what determines the direction of the change in dynamic welfare resulting from a change in non-producer and producer surplus due to an externality remedy away from the initial levels (n, π) . The straight line represents combinations of surpluses that keep classic *ex-post* welfare the same 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 an externality remedy has the same effect on both producer and non-producer surplus (n and π both fall or rise), then clearly dynamic welfare is changed in the same direction as classic *ex-post* welfare. In the figure, both static and dynamic welfare change in the same direction in the first and fourth quadrant around the initial surplus levels (n, π) . However, 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. This would be true for the new levels of surpluses in 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. This would be true for new levels in region B in the figure. The figure illustrates not only that classic Pigouvian welfare calculations can produce quantitatively incorrect results, but also that their *qualitative* conclusions may be inaccurate; the *sign* of the static and dynamic welfare impacts may differ. This occurs only if Pigouvian remedies affect profitability and classic *ex-post* welfare differentially. In that case, *ex-post* profits can be raised to overturn a given reduction in classic welfare, or profits can be reduced to overturn a gain in classic welfare.

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 $\mathbb{R}_+ \times \mathbb{R}_+$, *profitability will be higher under dynamic welfare than under static welfare*. More

precisely, it can be shown⁷ 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 $n_W \geq n_D \geq 0$.

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, whether to static or expected profits, to standard Pigouvian remedies that reflect the impact of the remedies on technological change⁸.

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 when 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. 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,

⁷ 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 $n_W, n_D \geq 0$, it must be that $W_D \geq \pi_D$ so that R&D is under-provided. Thus, a contradiction to the dynamic optimality of (n_D, π_D) .

⁸ An open question is whether the political process favors reduced profits and short-term Pigouvian solutions over dynamically efficient ones under technological change.

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.

IV. The Impact of External Effects on IP Interventions

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

Without externalities, the optimal prize always dominates the optimal patent, $D(\tau^*) \leq D(\theta^*)$. 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 patent is always second-best and is thus dominated by such a prize.⁹ 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, whether private or public, 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: $y(\theta) = y_c \neq y_w$. For example, if antibiotics are to be limited in their use, then competitive markets generate too much usage. Efficient *ex-post* production and distribution under a prize, while trivial through free licensing in the case of no externalities, is a non-trivial issue when externalities exist.

⁹ 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.

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,¹⁰

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

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

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 under negative external effects. More precisely, under negative external effects, *ex-post* efficiency involves a markup of $p = c_y - e_y$ while the patent involves a markup $p = [1/(1+\varepsilon)]c_y$, where ε is the elasticity of demand. This “patent-taxation” of externalities implies that the traditional welfare loss associated with patents, induced by monopoly markups, is reduced under negative externalities but is exaggerated under positive externalities. For a negative externality, 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 useful to consider as an illustration of this general dominance of prizes over patents under positive external effects, even under

¹⁰ This can be shown formally by noting that under positive external effects: $W(y_\pi) < W(y_c) < W(y_w)$. *Ex-post* welfare is thus higher under the prize than under the patent. The optimal prize $\theta^* = aW(y_c)$ induces more of the under-provided R&D than any patent. Under negative external effects, the reverse inequality $W(y_c) < W(y_\pi) < W(y_w)$ may be obtained. The optimal prize size is still $\theta^* = aW(y_c)$, which may induce less of the under-provided R&D whenever patented profits exceed the prize: $v(\tau)\pi(y_\pi) \geq \theta^*$.

conditions when patents are first best under no externalities. For diseases present only in poor countries, consider when social surplus mainly consists of the external altruistic effects of developed nations and, to a lesser degree, of the consumer surplus in developing 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 dynamic welfare: $\pi = r = D = 0$ for all patents of length τ . Prizes dominate patents here because the patent holder can only capture consumer surplus, but when consumers cannot pay variable costs, there is only 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 are still dominated by prizes. Rather, the problem with patents under altruism is that the output is not sold to those willing to pay for it.

Moreover, when the inverse demand curve is below marginal costs, the importance of the method of production and distribution under a prize is quite drastic under such positive external effects. Under the standard mechanism for generating output associated with a prize, *i.e.*, free licensing, patents always dominate any *positive* (and hence non-optimal) prize. This is because the R&D would be undertaken without distribution for any positive prize. However, under a patent, the R&D would not occur. 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.

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 and unrestricted licensing does not induce *ex-post* efficiency. Given these shortcomings of the traditional forms of prizes under external effects, this section analyzes 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. A single measure is unlikely to solve the two aspects of the allocation problem, and thus price guarantee contracts are unlikely to implement first-best allocations.

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. In the special case of no externalities, the optimal price is simply the average consumer surplus under the Pigouvian output level, $p^* = s_o(y_w)/y_w$ equating profits with *ex-post* welfare.

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*.¹¹ 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. 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.

¹¹ 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.

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 in the absence of innovation prizes.¹²

Public Technology Adoption Criteria in Health Care

The previous types of contracts specified price or quantity explicitly for a particular type of innovation. Other methods of awarding innovation through public demand would entail specifying the prize *implicitly through a general decision rule* that determines public technology adoption.

For public technology adoption in health care, many developed countries, though not the U.S., make use of some form of “cost-benefit” or “cost-effectiveness” criteria determining whether governments adopt and pay for new health care technology. Although there are many forms of such criteria developed for public purchasing decisions in the absence of private market signals (see, *e.g.*, Weinstein, 1995; Johanneson, 1996; or Gold *et al.*, 1996), the basic goal of such criteria is to determine whether increased-health care spending through new technology adoption is justified by its health benefits to a population that may or may not be able to pay for those benefits. These criteria have not been related to standard economic efficiency generally, and efficiency under external consumption effects in particular.

¹² 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.

One form of technology adoption criteria involves a public purchase only if consumer valuations are higher than spending, that is, if $s > 0$. In a private market under competitive conditions, this implies consumer valuations higher than costs of production: $s(y_c) > 0$. In estimating the cost-benefit or cost-effectiveness of an innovation one thus needs an estimate of demand and spending.¹³ In a private market without public or private insurance, *e.g.*, elective surgery procedures, a new technology being estimated to be cost-effective in this sense would of course be the expected outcome as it is a direct consequence of consumers buying only at a price below their valuations. Although this expected outcome has to be qualified by the presence of private or public insurance, it is supported by a large existing and growing empirical health economics literature on the cost-benefit and cost-effectiveness of recent innovations (see Weinstein, 1995; Cutler and McClellan, 2001).

Now consider the technology adoption criterion induced by cost-effectiveness defined as the public sector adopting the new technology at price p for which the benefits to all patients in the public program, whether they can pay or not in a private market, are larger than spending: $s_o(y) > py$. The public sector could either specify such a price or it could be implied by optimal private monopoly pricing given the existence of the adoption rule. Given the adoption rule, the patent holder would choose a price $p = s_o(y)/y$ given the infinite elasticity of demand when raising the price above that level.

This rule for public technology adoption is unlikely to be *ex-post* efficient or dynamically efficient under external consumption effects. It is not *ex-post* efficient

¹³ However, as opposed to estimating demand directly through price changes, which would incorporate both observable and unobservable dimensions of the surplus s , common forms of cost-effectiveness analysis typically attempt to estimate the consumer surplus indirectly through monetizing observable health benefits in various ways. A common method is to apply value of life- or contingent valuation estimates to health improvements implied by mortality- or morbidity reductions.

because it is not Pigouvian; it does not reflect the marginal social costs and benefits when non-consumer surplus is present. Under altruistic motives for health care provision, the *ex-post* optimal technology adopts when $s + e > c$ rather than when $s > 0$. This is simply the discrete and infra-marginal version of a Pigouvian policy for the decision to adopt/not-adopt a given technology. Although the average consumer surplus determines the price, it is not dynamically efficient because the average non-consumer surplus is not incorporated as we derived would be the case under a dynamically first-best policy.

In sum, if taxed-financed technology adoption reflects non-consumer surplus, then it should not be based on common cost-benefit or cost-effectiveness criteria induced by private markets, but be more lenient towards adoption. The close to universal adoption of new health care technology by many governments in richer nations, as exemplified by CMS in the U.S., may be efficient. In a sense, because we would expect private markets to adopt so called cost-effective treatments by themselves, as in the elective surgery case, consumer surplus comparisons with spending is the correct public decision criterion only when there is no role for public financing!

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.

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 as it represents the deadweight loss of a patented monopolist; a patent extension reduces *ex-post* welfare by simply extending monopoly protection.

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. In other words, the traditional welfare loss associated with patents may not be present under negative externalities but is exaggerated under positive externalities.¹⁴ 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.

¹⁴ As the patent holder marks up according to $p = c_y / (1 + \varepsilon)$ and the social optimal price markup is $p = c_y - e_y$, the relative size of the elasticity of demand, ε , and the sign of the externality determines whether the marginal cost of patent extension is reduced or raised by external effects.

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, $x_r W - 1$. External effects have an indeterminate effect on this net gain in R&D due to the nature of its under-investment.

As an illustration, consider the case of constant returns in which the cost-function is of the form $c(y) = cy$, the externality of the form $e(y) = ey$, and in which demand is linear as in $p(y) = a - by$. 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:

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

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 as the patented profits do not depend on 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 indeterminate 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. In sum, for a positive externality ($e > 0$) or a private good ($e = 0$), it is ambiguous whether a rise in the externality should involve a shorter or longer patent. If the externality is negative, an analogous argument applies.

V. 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. 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*. Thus, arguing in favor of Pigouvian solutions under technological change is incorrect, just as is arguing in favor of competitive markets for new inventions in the absence of externalities. 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 a related empirical paper, see Philipson and Jena (2005), we estimate the model discussed here and show how economic value of the large public investments into AIDS research in the US critically depend on the type of external effects discussed here.

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 medically harmful world-wide. Altruism by richer nations makes it an externality and a policy problem. However, existing policy proposals 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.¹⁵ 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. For exclusively third-world diseases, where demand curves are below variable costs, efficient R&D is done for the *rich* countries, not for the poor! In a sense, many proposals inefficiently involve demanding firm owners not only to pay for R&D to

¹⁵ In a more general context, Grossman and Lai (2002) discuss the optimality of streamlining IP protection across countries.

discover new drugs, but also to cover the bill for the distribution and consumption to fulfill the altruistic desires of the tax base.

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-terror,¹⁶ 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.¹⁷ If a society cares about those who are unlucky enough to catch uncommon diseases, the social surplus is larger than the consumer surplus. The Orphan Drug Act encourages 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 a better international policy.

Lastly, 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.¹⁸ The non-exclusivity induced by the free flow of innovations across countries, and the desire to free

¹⁶ 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.

¹⁷ 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.

¹⁸ Becker *et al.* (2005) discuss the impact the sharing of the benefits of medical R&D across rich and poor countries has had on reducing world inequality.

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.

In general, future research may fruitfully address the design of optimal externality and IP measures in health care and other areas. The analysis suggests that single measures aimed at solving the external consumption or private R&D problem *alone* often fall short. In order to achieve first-best allocation, one needs to break the link between *ex-ante* R&D and *ex-post* output provision.¹⁹ 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*.

¹⁹ 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).

References

- Becker, G., T. Philipson, and R. Soares 2005. The rise in longevity and world inequality. *American Economic Review*, forthcoming, March 2005.
- Cutler, D. and M. McClellan. 2001. Is technological change in medicine worth it? *Health Affairs* 20(5):11-29.
- Gold, M.R. *et al.* 1996. Report of panel on cost-effectiveness in health and medicine. Department of Health and Human Services, United States.
- 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.
- Johannesson, M. 1996. *Theory and Methods of Economic Evaluation of Health Care*. Dordrecht, the Netherlands: Kluwer Academic Press.
- 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.
- 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 A. Jena (2005) "External Consumption Effects and Medical R&D: The Case of HIV", mimeo, The University of Chicago.
- Philipson, T. and S. Mechoulan. 2003. IP, R&D, and external consumption effects. NBER Working Paper No 9598, Cambridge, Mass.
- 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.