Surplus Appropriation from R&D and Health Care Technology Assessment Procedures

by

Tomas J. Philipson[1]

and

Anupam B. Jena

The University of Chicago

Abstract:

Given the rapid growth in health care spending that is often attributed to technological change, many private and public institutions are grappling with how to best assess and adopt new health care technologies. The leading technology adoption criteria proposed in theory and used in practice involve so called "cost-effectiveness" measures. However, little is known about the dynamic efficiency implications of such criteria, in particular how they influence the R&D investments that make technologies available in the first place. We argue that such criteria implicitly concern maximizing *consumer surplus*, which many times is consistent with maximizing static efficiency after an innovation has been developed. Dynamic efficiency, however, concerns aligning the social costs and benefits of R&D and is therefore determined by how much of the social surplus from the new technology is appropriated as *producer surplus*. We analyze the relationship between cost-effectiveness measures and the degree of surplus appropriation by innovators driving dynamic efficiency. We illustrate how to estimate the two for the new HIV/AIDS therapies that entered the market after the late 1980's and find that only 5% of the social surplus is appropriated by innovators. We show how this finding can be generalized to other existing cost-effectiveness estimates by deriving how those estimates identify innovator appropriation for a set of studies of over 200 drugs. We find that these studies implicitly support a low degree of appropriation as well. Despite the high annual cost of drugs to patients, very low shares of social surplus may go to innovators, which may imply that cost-effectiveness is too high in a dynamic efficiency sense.

---

[1] Corresponding author: t-philipson@uchicago.edu. We are thankful for comments from seminar participants at The University of Chicago, Princeton, Berkeley, SUNY Buffalo, the Kiel Institute Conference on New Technology and National Health Systems, and the NIH-Director conference *Biomedical Innovation and the Economy*. Philipson is thankful for financial support from The Milken Institute, Santa Monica, CA. Jena received fellowship support from the NIH through the University of Chicago Medical Scientist Training Program. We thank Gary Becker, Angus Deaton, Isaach Ehrlich, and Casey Mulligan for comments, Lisa Lee and Michael Campsmith (CDC) for providing data, and Jennifer Kates (Kaiser Family Foundation) and Ruigang Song (CDC) for helpful advice.

I. Introduction

Technological change is often argued to be a central force behind the growth in health care spending.[2] Given this rapid growth, criteria used by private and public institutions to value the increase in health care spending therefore requires a methodology to measure the value of new health care technologies brought about by R&D investments. There is a long-standing and vast health economics literature that attempts to assess the value of new technologies by use of so called cost-effectiveness, cost-utility, or cost-benefit analysis, hereafter referred to collectively as CE analysis.[3] This type of CE analysis has been the major method proposed to evaluate new inventions and has been argued to be central in managing new technologies, their adoptions, and their impact on long term health care spending.

Although not explicitly stated as such, we argue that CE criteria are implicitly concerned with estimating the observed market level of *consumer surplus* associated with a given technology. In particular, many technology assessments attempt to quantify the health impacts of new technologies for patients or health plans by comparing patient benefits with spending at observed market prices. Examples include cost-effectiveness using spending per quality- or disability adjusted life years, as is common by public buyers outside the US, or cost-benefit analysis monetizing mortality reductions through value-of-life estimates, as is common in studies assessing the gains of increased health care spending. The central theme of such standard CE assessments performed in practice seems to be to measure consumer surplus or net consumer benefits. As is the norm in CE practice, technologies are deemed more valuable the larger is the patient- or health plan benefits above what is spent on them.

[2] See e.g. Newhouse (1992).

[3] The literature is vast, but for examples, see Weinstein and Stason (1977), Johanneson and Weinstein (1993), Gold et al. (1996), Meltzer (1997), Drummond et al. (1997), Garber and Phelps (1997), Garber (2000), Cutler and McClellan (2001), and Cutler (2005).

However, when new technologies are brought to life from costly R&D, consumer surplus is a very poor guide to inducing optimal (second-best) R&D investments. Rather, the degree to which producer surplus captures social surplus, often at the expense of consumer surplus, becomes the central issue that determines dynamic efficiency. This, of course, is the rationale for the patent system, which substitutes producer surplus for consumer surplus in order to stimulate more efficient R&D investment. Therefore, we argue that for the same reason that patents are preferred even though they lower consumer surplus after technologies are discovered, technology adoption criteria are preferred that do not only focus on consumer surplus. Put differently, even though measured levels of CE would be higher without the patent system, since patients or health plans would spend less to get the same technology, dynamic efficiency would clearly be lowered. An illustrative case of the dangers of CE criteria may be vaccines, which many times have been estimated to be extremely cost-effective but for that and other reasons lack any appreciable R&D investments.[4]

As consumer surplus or cost-effectiveness determines static efficiency and innovator appropriation determines dynamic efficiency, we analyze the relationship between the two. In the case of monopoly R&D, we arrive at the stark implication that many times both dynamic efficiency and patient health are maximized when CE is *minimized.* We show how this implication is altered under public R&D subsidies (such as those by the National Institutes of Health (NIH) in the US) and competition in R&D (leading to patent racing that may duplicate R&D efforts).

As the ability of innovators to appropriate the surplus of their innovations is central to dynamic efficiency, we estimate the degree of appropriation in an illustrative manner for a major breakthrough in medicine—the new drugs to treat HIV/AIDS that entered the market from the

[4] A major concern here has, of course, been product liability issues. See e.g. Manning (1993).

late 1980's onwards.[5] Our major finding is that innovators captured only 5% of the nearly $1.4 trillion worth of social surplus arising from these new technologies.[6] Thus, despite the high prices of many therapies such as the new HIV drugs, the low degree of appropriation raises concerns about whether cost-effectiveness is too high in that it induces dynamic inefficiency.

We generalize this finding by deriving why and how the CE results of over 200 studies on drugs can be implicitly viewed as identifying the degree of innovator appropriation. We derive conditions under which the measured level of CE of a technology may be used to identify the share of social surplus appropriated by producers of that technology; the CE of a given technology *reveals* information about the cost or demand parameters. When such identification is feasible, the existing and vast CE literature informs us about the degree of innovator appropriation. We find that 25% of the interventions considered have estimated levels of appropriations of less than 7%, while 75% have appropriations less than 25%. Our illustrative finding for HIV/AIDS drugs suggests their appropriation of social surplus is at the twentieth percentile of the 200 technologies considered. As complementary evidence, we show that these estimated levels of innovator appropriation are also consistent with alternative methods of calculating appropriation based on price reductions after patent expiration and the identifying assumption of profit-maximization.

The paper may be briefly outlined as follows. Section 2 discusses the relationship between CE measures and dynamic efficiency. Section 3 presents estimates of the share of social

[5] This finding builds on and extends work of Philipson and Jena (2005) who argue that HIV/AIDS is an important case to consider in and of itself, partly because it is perhaps the major disease targeted by public sector R&D in the US. Public R&D on HIV/AIDS was roughly $2 billion in 2000. Health, in general, is among the three leading industries into which the government allocates its R&D, the other two being defense and aero-space. The National Institutes of Health (NIH) is responsible for allocating the vast majority of the public R&D dollar—in 1999, NIH funding accounted for nearly 81% of public spending on health R&D. Of the $13.9 billion that the NIH spent on research in that year, nearly $1.8 billion (13%) was spent on HIV/AIDS (Health, United States, 2002).

[6] Our findings relate to an existing literature on the general inability of innovators to capture the social value of their inventions, see e.g., Mansfield et al. (1977), Mansfield (1985), Levin et al. (1987), Hall (1996), and Nordhaus (2004).

surplus appropriated by producers of the new HIV/AIDS drugs. Section 4 generalizes these findings to traditional CE measures reported in the literature. Lastly, section 5 concludes.

II. Technology Assessment and Dynamic versus Static Efficiency

In order to discuss how CE analysis relates to static and dynamic efficiency, for a given output level q denote the ex-post social surplus of a new technology by w(q). This social surplus can be divided into a consumer surplus, z(q), and producer surplus (variable profits), π(q), as in:

$$w(q) = z(q) + \pi(q) \qquad (1)$$

For example, a commonly analyzed case is when price-discrimination is infeasible, in which case a given output level q induces both profits and consumer surplus according to:

$$\pi(q) = p(q) \cdot q - c(q) \qquad (2)$$

$$z(q) = \int_0^q [p(y) - p(q)]dy = g(q) - p(q)q \qquad (3)$$

where p(q) is the inverse demand function, c(q) is the variable cost function which excludes the fixed cost of R&D, and g(q) is the gross consumer benefit. In the discussion that follows, we refer to the observed surplus as the surplus which results at the market quantity q. The potential surplus obtains at the quantity that maximizes w(q)—in the absence of ex-post market imperfections, this is true at the competitive output.

A. Cost-Effectiveness Criteria

In this standard framework, we argue that typical CE technology evaluation has implicitly centered on consumer surplus, by focusing on how much patients benefit beyond what is spent on the technology after it has been developed. Despite the many forms of such criteria developed to date, their basic goal seems to be to determine whether increased health care spending on new technologies is justified by "societal", "health plan", or "patient" benefits in terms of improved health. Absent from the discussion has been the effect of such criteria on the behavior of innovators who make the technologies available in the first place. Although static efficiency is

often enhanced with increases in CE, as it implicitly concerns consumer surplus, these criteria are less understood in terms of how they relate to dynamic efficiency when the observed level of CE is the result of rational behavior by market participants.

Common measures of CE ratios relate the (here monetized) patient benefits to observed spending levels. In the traditional framework, this can be expressed by:

$$z_R = \frac{g}{p \cdot q} = 1 + \frac{z}{pq} \qquad (4)$$

This measure (z_R) expresses consumer benefits as a *ratio* to spending, similar to the standard consumer surplus measure (z) that expresses it as a *difference* between the two. Ratios are often estimated through spending per quality- or disability-adjusted life years or through monetized versions of health benefits, in which the value of life is compared to observed spending levels. These attempts, however, are implicitly related to the size of consumer surplus, since they compare consumer benefits to observed spending levels.[7] In particular, static technology assessments in health care commonly rely upon the use of "cost-benefit", "cost-utility", or "cost-effectiveness" criteria to determine under what circumstances the value (whose units depend on the measure) of a given technology exceeds what is spent on it. Although it is true that CE analysis concerns the ratio of gross benefit and spending, while consumer surplus concerns their difference, both change in the same direction with unilateral changes in costs and benefits.

Regarding the estimated CE magnitudes, many empirical studies estimate and document z_R ratios above unity for employed technologies (see e.g. references in Introduction). Yet, it would be extremely surprising if correctly measured z_R ratios were found to be below unity, at least in a standard market economy. As an illustration, consider a private market for health care

[7] The implicit consumer surplus estimation of CE analysis differs from traditional economic analysis—the latter typically attempts to assess consumer surplus by estimation of demand schedules, by observing changes in demand during supply-induced price changes. Importantly, the demand curve for a good summarizes the value to consumers of both its observed and unobserved attributes. On the contrary, estimates of consumer surplus based on cost-effectiveness or cost-benefit analysis are typically formed indirectly by monetizing *observable* consumer benefits, e.g. by use of value of life estimates to estimate the gross consumer benefit from mortality reductions.

without public or private insurance, as might exist for certain elective surgeries in the US, such as e.g. plastic surgery. A new plastic surgery technology would have a z_R ratio above unity (if estimated correctly) if individuals bought the product only when their valuation of it exceeded the price. This, of course, would always be predicted under standard demand analysis. Although this expected and basic 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-effectiveness of recent innovations.[8]

More importantly, the fact that a technology is cost-effective in this way only reveals that there is a positive consumer surplus. However, a positive consumer surplus is consistent with *any* output level, regardless of how high or low it is. Consequently, being cost-effective in this sense bears no relationship to either static or dynamic efficiency! Even if output is not at the competitive or monopoly level, consumer surplus is still positive and hence the technology deemed cost-effective. The problem is that being cost-effective is only necessary but not sufficient for static or dynamic efficiency.

B. Cost-Effectiveness and Dynamic Efficiency

To consider the dynamic efficiency induced by common health care assessment criteria, one must consider how such criteria affect efficiency in the presence of technological change driven by endogenous R&D. Let technological change be characterized by x(r), an increasing, differentiable, and strictly concave function representing the probability of discovery for a given level of R&D undertaken, r. The optimal level of R&D that maximizes expected payoffs for any hypothetical ex-post prize, k, is denoted r(k) and is defined by:

$$r(k) = \arg\max_{r}\left[x(r)k - r\right] \qquad (5)$$

Our assumptions about x(r) imply that r(k) is an increasing function so that R&D rises with the ex-post reward.

[8] See, for example, Cutler (2004).

Monopoly R&D Investments

First consider the case of a single monopolist investing in R&D who receives a share, *a,* of the social surplus w, where $0 \leq a \leq 1$. Then, r(a·w) represents the R&D undertaken when those investing in R&D maximize expected profits. If profits drive R&D investments, the expected social surplus is:

$$E(a, w) = x[r(a \cdot w)] \cdot w - r(a \cdot w) \qquad (6)$$

where $w = z + \pi$ is the social surplus ex-post. This expression directly highlights the well-known implication that dynamic efficiency only occurs when those undertaking the costs of R&D have incentives that are properly aligned with society, which is true when social surplus is entirely appropriated as profits, i.e. $a = 1$ (see e.g. Arrow (1961) and Tirole (1988)). In other words, the key factor driving dynamic inefficiency is that profits (π) are less than social surplus (w). More importantly, the size of the consumer surplus, focused on by CE criteria, is what drives a wedge between profits and social surplus and hence leads to under-investment in R&D. Indeed, in this setting, the dynamically efficient R&D investment is r(w), which is obtained when the entire social surplus is appropriated as profits.

More generally, for any technology and preferences, the observed profits associated with a given level of social surplus can be written a·w. The main issue, then, is that $a < 1$. For example, when production is characterized by constant returns to scale, it can be shown that monopolists facing either linear or constant-elasticity demand earn profits that are proportional to the potential social surplus. Specifically, $a = 1/2$ in the case of linear demand and $a = [(\varepsilon - 1)/\varepsilon)]^{\varepsilon}$ under constant elasticity of demand.[9,10] In general, if the total social surplus associated with a technology is w, the size of the under-investment in R&D is $r(w)-r(\pi) = r(w)-r(w-z)$, which, since

[9] The social surplus implicit in these results is the *potential* social surplus available to innovators, i.e. the social surplus that obtains when price is set at its competitive level. This differs from the *observed* social surplus available to the monopolist, which obtains when price and quantity are determined by the monopolist.

[10] Interestingly, profits may even exceed the *private* social surplus (i.e. the gross benefit to consumers net of costs of production) when there are external effects in consumption. See, for e.g., Philipson, Mechoulan, and Jena (2006) who discuss R&D under altruism in health care.

r(.) is an increasing function, rises with the consumer surplus focused on by CE criteria. The fact that dynamic efficiency is driven by the appropriation of social surplus to innovators implies that substituting producer surplus for consumer surplus often raises dynamic welfare. This is analogous to the argument that patents hurt static efficiency but raise dynamic efficiency by engaging in similar substitution.

The important implication of this is that the CE associated with the ex-post market for a technology is not clearly and monotonically related to measures of static or dynamic efficiency. Indeed, in a private market with perfect price discrimination, dynamically efficient R&D occurs because the innovator captures the entire social surplus. Therefore, the dynamically optimal allocation of surpluses implies that the consumer surplus should be *minimized*, as opposed to maximized under a CE criteria, to enhance dynamic efficiency. In this case, dynamic efficiency dictates that a technology should just break even ex-post (i.e., $z_R = 1$) and that empirical studies citing more cost-effective technologies are, in fact, documenting a dynamic inefficiency! Indeed, as discussed, the underinvestment in R&D from its socially optimal level, r(w) - r(w-z), rises with how "cost-effective" a technology is assessed to be according to traditional CE analysis. In this case, the dynamically efficient minimization of CE is a direct implication of the classic problem of non-appropriation by innovators leading to under-investment in R&D. Importantly, note that minimization of CE in this context still maximizes patient health (as full demand for the health care product obtains) though not consumer surplus.

Competitive R&D Investments and Appropriation

There are important instances in which full appropriation of social surplus by producers may not be dynamically optimal, a primary one being competitive R&D through so-called patent racing. Since competitive R&D leads to an equilibrium level of R&D that is determined by the average (rather than marginal) profit associated with entry, non-appropriation may enhance efficiency by taxing the over-provision of R&D. This may be particularly relevant to the debate over excessive R&D into so-called "me-too" drugs in the pharmaceutical area. If the total fraction of surplus

appropriated by successful R&D efforts is *a*, where $0 \leq a \leq 1$, the equilibrium level of R&D r is determined by the zero-profit condition:

$$x[r] \cdot aw = r \qquad (7)$$

where x(.) is the probability of discovery given the total investment in R&D *by all firms*, r. Given this particular form of the R&D process, it is straightforward to show that the optimal level of appropriation is characterized by[11]:

$$a = \frac{x'[r(a \cdot w)] \cdot r(a \cdot w)}{x[r(a \cdot w)]} \qquad (8)$$

At the optimum, surplus appropriation equals the elasticity of R&D productivity (i.e. the percentage increase in the probability of discovery given a 1 percent increase in the level of R&D). Our assumptions on x(.), namely concavity and zero probability of success in the absence of R&D, imply that the optimal surplus appropriation is less than one—the extent to which this occurs depends on the nature of R&D productivity, x(.).[12]

If such patent races lead R&D to be over-provided, our conclusions emphasizing under-provision of R&D under monopoly R&D may be altered. However, there appears to be an almost universal policy towards subsidizing (as opposed to taxing) R&D, such that most nations have decided that the forces operating towards over-provision are dominated by those operating towards under-provision. In light of this, although incentives favoring over-provision may change the quantitative conclusions of our analysis, the qualitative conclusion that CE criteria limit already under-provided R&D seems generally applicable to most research areas and countries.

---

[11] This follows from the zero-profit condition and the FOC for the expected welfare, $x'[r(aw)]w = 1$.

[12] An interesting case to consider is when R&D productivity is characterized by $x(r) = 1 - p^r$, where $0 \leq p \leq 1$. In this case, one can show that the optimal level of R&D is $r^* = \ln[-1/(w \cdot \ln(p))]/\ln(p)$ and the optimal appropriation is $a^* = -[r^* \cdot (p^{r^*}) \cdot (\ln(p))]/(1-p^{r^*})$. For an ex-post surplus of $1 trillion (which we argue is roughly the case for HIV/AIDS), the optimal appropriation may be as low as 5% (when R&D is very productive and the optimal R&D is $40 billion) or perhaps as high as 80% (when the optimal R&D is $270 billion). The extent to which appropriation deviates from one depends on the productivity of R&D, as a lower p implies more productive R&D.

Public R&D Subsidies and Appropriation

Another important case in which non-full appropriation may be optimal is when publicly funded R&D comprises a significant portion of total R&D, as is common in US health care through NIH. Since the dynamically optimal level of *total* R&D is still r = r(w), the presence of publicly funded R&D implies that the optimal private R&D (and hence, appropriation) should be lowered accordingly. More precisely, consider the expected social surplus in the presence of publicly funded R&D, s, and surplus appropriation, *a*, where $0 \leq a \leq 1$:

$$E(s,a,w) = x[s + r(a \cdot w)] \cdot w - s - r(a \cdot w) \qquad (9)$$

The probability of discovery is determined by the sum of public and private R&D, the latter being driven by the appropriation of ex-post surplus by producers. For a given level of subsidization the optimal appropriation satisfies:

$$w \cdot x'[s + r(a \cdot w)] = 1 \qquad (10)$$

which implies $s + r(a \cdot w) = r(w)$. Applying the implicit function theorem, this implies that an increase in the subsidy affects the optimal degree of appropriation according to:

$$\frac{da}{ds} = \frac{-1}{w \cdot r'(a \cdot w)} < 0 \qquad (11)$$

Hence, appropriation falls with subsidization. In particular, appropriation is unity in the absence of subsidization and less than unity under a positive subsidy. Put differently, since the marginal product of private R&D is decreasing in the level of subsidized R&D, private R&D (and hence appropriation) optimally falls as its public counterpart increases.

III. Surplus Appropriation for the New HIV/AIDS Drugs

The previous discussion highlighted the importance of surplus appropriation by innovators (and hence low levels of CE) to dynamic efficiency, even if that level of appropriation was not full. As the ability of innovators to appropriate the potential surplus of their innovations is central to dynamic efficiency, we illustrate the degree of appropriation for the new drugs to treat HIV/AIDS that entered the market from the late 1980s onwards. This analysis will then be used to illustrate

how more generally levels of innovator appropriation may be inferred from existing CE estimates in the literature.

A. Estimates of Gross Consumer Benefits

The value of life induced by new drug therapies is the value of increased survival for all individuals who choose treatment, relative to a benchmark in which no (or worse) therapies exist. In a related work, we develop a methodology to value the increases in survival attributable to the now standard treatments for HIV/AIDS. The thought experiment behind the analysis is the following. For a hypothetical individual infected in a given year t, we examine how that individual's survival under treatment (S_t) compares to a counterfactual, baseline survival in which no drugs are available (S_o).[13] We then attach a monetary value to that increased survival and sum across all infected individuals in that cohort. This process is repeated for each set of cases, cohort by cohort, since the start of the epidemic and aggregated up.

This approach delivers the *potential* aggregate value of life induced by treatment, i.e. the value of life obtained when all infected individuals receive treatment. This implicitly assumes that at the competitive output, all individuals infected with HIV (i.e. the full incidence) consume drug therapy. At the competitive output, the ex-post social surplus arising from a technology is highest and therefore represents the potential surplus available for appropriation. This differs from the *observed* gross value of life induced by treatment, which depends on the number of individuals who actually receive treatment.

The aggregate *potential* gross consumer benefit, g, induced by the new drug consumption is calculated by multiplying the size (or incidence) of cohort t, n_t, by the monetary value of increased survival and summing over all calendar years. Moreover, the aggregate *observed* consumer benefit can be calculated by replacing n_t with the number of individuals in that cohort

[13] The ideal counterfactual survival in the absence of treatment (S_o) is the *cross-sectional* survival of individuals infected at the start of the epidemic, here taken to be 1980. S_t is the *longitudinal* (i.e., lifetime) survival of individuals infected in year t. The use of longitudinal survival captures the benefit to individuals infected with HIV prior to 1987 (when drug therapy first became available) who survive until then and consequently face improved life-expectancy due to treatment.

who actually receive treatment. Formally, the aggregate *potential* gross consumer benefit is written as:

$$g = \sum_{t=1980}^{2000} \beta^{t-1980} n_t \cdot g_t \qquad (12)$$

where $g_t = g(S_o, S_t)$, or the monetary value of increasing survival from the baseline survival S_o to the higher future survival faced by cohort t, S_t. The gain in survival, g_t, is calculated using the infra-marginal valuation formula of Becker et al. (2005).[14]

To empirically implement the calculation of aggregate potential consumer benefits, we apply these formulas to published levels of HIV incidence and estimated changes in survival induced by HIV/AIDS drugs (see Philipson and Jena (2005) for a detailed discussion of the various data sources and methods used to estimate improvements in survival).[15] We estimate that the average individual infected with HIV experienced an increase in life-expectancy of roughly 15 years since the start of the epidemic, from 19 to 34 years.[16] These improvements in survival are due to increases in both the time to onset of AIDS (after being infected with HIV) and the period of time alive after a diagnosis of AIDS.

As described earlier, the potential value of improved survival for a given cohort is computed by multiplying that cohort's incidence of HIV by the value of increased survival experienced by a single individual receiving treatment in that cohort. Table 1 presents this value for several cohorts.

---

[14] The value of an infra-marginal change in survival from S_o to S_t under a yearly income y_t is determined by $V[y_t + e_t, S_o] = V[y_t, S_t]$, where V is the indirect *lifetime* utility function and e_t is the yearly compensation required to make the hypothetical individual indifferent between the two survival frontiers. The lifetime value for the gain in survival (g_t) is calculated by summing the yearly compensation (e_t) over time, discounting by the rate of interest and the new survival probability.

[15] Importantly, our estimated survival curves are weighted averages across individuals receiving and not receiving treatment. Thus, for each cohort, the reported survival is lower than survival among only those receiving treatment, S_t. We assume our reported curves to be empirical analogs for S_t and therefore underestimate the aggregate observed and potential benefits of treatments introduced to date.

[16] These figures are consistent with those in the literature (see e.g. Lichtenberg, 2005). To further examine the robustness of these estimates, we *predict* the number of individuals alive with HIV/AIDS in 2003, based on the annual reported incidence of HIV and our estimated survival curves. We then compare this to the *reported* number of individuals living with HIV/AIDS in 2003. The predicted and reported figures differ by only 12,000 people (out of nearly 1 million alive with HIV/AIDS).

Table 1: Value of Gains in Survival for HIV Infected Individuals

		Value of Survival Gains ($)	
Year of HIV Infection	HIV Incidence	Individual ($)	Aggregate ($ Billion)
1980	20,000	17,655	0.35
1985	160,000	146,874	23.50
1990	40,000	322,311	12.89
1995	40,000	613,839	24.55
2000	40,000	740,515	29.62
Total Discounted Value			398

Notes: All figures are discounted to 1980 and are in year 2000 dollars. The total discounted value includes all years from 1980 to 2000.

These results demonstrate that the aggregate *potential* value of improved survival experienced by all individuals infected with HIV to date has been nearly $400 billion. This, of course, ignores the value of increasing survival for all individuals who have not contracted HIV yet. To add this component, we forecast the value to future cohorts of HIV infected individuals by assuming that all cohorts experience the same aggregate gain in survival g_t as the last cohort, 2000. Assuming that the future incidence of HIV is equivalent to the last period, we calculate the discounted sum of future gains for individuals infected with HIV in the future. We then add this amount to the value to date shown above, namely $398 billion. This leads to an aggregate *potential* value of increased survival for all past and future cohorts of nearly $1.4 trillion.

The aggregate *observed* value of improved survival can be calculated from data on the annual number of individuals in a given cohort who later receive HIV/AIDS drug therapy. Specifically, if S_t characterizes the survival of someone infected with HIV in year t who receives treatment, the aggregate *observed* consumer benefit for that cohort equals the number of individuals in that cohort who receive treatment multiplied by the value of increased survival, $g(S_o, S_t)$. We use published data from the HIV Cost and Services Utilization Study (HCSUS) to approximate the number of individuals in a given cohort receiving treatment. According to the study, between 60 and 85 percent of HIV infected individuals report some form of anti-retroviral

therapy (Shapiro, et al., 1999).[17] Assuming a mid-range estimate of 70%, this implies an incidence of anti-retroviral consumers equal to 70% of HIV incidence and a corresponding aggregate observed benefit to all past and future cohorts of $980 billion (i.e., $1.4 trillion*0.7).

B. Producer Surplus

The overall producer surplus obtained from R&D is determined by the present value of producer surplus to firms producing HIV/AIDS drugs. We apply existing estimates of markups for brand-name drugs (as estimated from patent expirations) to approximate variable costs as 15% of sales. Using estimates of national spending on HIV/AIDS drugs obtained from IMS Health and reported by Lichtenberg (2005), this implies lifetime sales of HIV/AIDS drugs of roughly $74 billion with corresponding profits (variable costs) of $63 billion ($11 billion). These estimates assume future sales are equivalent to year 2000's patent-protected sales. If the current level of output is 70% of the competitive level, the variable cost of production is $16 billion when all infected individuals are treated.

Using the above figures, we can decompose the total lifetime value of HIV/AIDS drugs into consumer surplus, producer surplus (profits), and production costs. Recall that we estimated the total potential value, g, to be nearly $1.4 trillion, discounted to 1980 and in year 2000 dollars. This is the value that accrues when all infected individuals receive treatment and implies a potential social surplus of nearly $1.38 trillion ($1.4 trillion - $16 billion). With lifetime profits of $62.9 billion, producers appropriate only 5% of the potential social surplus available from their inventions[18]. We can make similar calculations regarding the share of observed surplus captured

[17] One could use annual data on aggregate sales and market prices to estimate the number of individuals receiving treatment. However, since the annual market quantity is composed of users from all infected cohorts, additional assumptions must be made to infer the share of individuals in a given infection-cohort who ultimately receive drug therapy.

[18] It is interesting to note that the small estimated share of social surplus appropriated by investors sheds important light on the recent growth of alternative funding mechanisms to stimulate HIV/AIDS research, e.g. through advance purchasing contracts of governments or private foundations. Given that there is a social surplus above a trillion dollars that is not appropriated by R&D investors, a few billion dollars added to stimulate innovation, as these public or private contracts seem to provide, seems to pale in comparison to interventions that would better allow innovators to capture the value of their innovations. Moreover, since both spending and markups are higher in the US than in the rest of the world (drug sales in the US account

by producers. Recall that the observed surplus is driven by the observed market quantity rather than the competitive quantity which results if all infected individuals receive treatment. With observed gross benefits of $980 billion, profits of $62.9 billion, and variable costs of $11.1 billion, consumers capture 7% of the social surplus available at the observed market quantities (consumer surplus = $906 billion, social surplus = $969 billion).[19]

IV. Extending the Analysis to Traditional Cost-Effectiveness Studies

Given the low estimated share of social surplus appropriated by producers of HIV/AIDS drugs, this raises the question of whether producers of similarly CE technologies appropriate comparable amounts of social surplus. And if so, can these results be generalized to obtain appropriation estimates that vary with a technology's observed level of cost-effectiveness? We begin this section by discussing conditions under which the often estimated CE of a given technology may, in fact, be used to infer the share of social surplus appropriated by producers of that technology.

Recall from our earlier discussion that the observed ratio of gross benefit to spending, z_R, can be written as g(q)/[p(q)q]. Similarly, the degree of observed surplus appropriation can be written as π(q)/w(q), where w(q) is the *observed* social surplus associated with a level of output q, and π(q) is the level of profit induced by that quantity. If m(q)=p(q)/[c(q)/q] is the markup above average costs, it is straightforward to show that for a given level of output, appropriation may be written as a function of the CE level as in[20]:

$$\frac{\pi(q)}{w(q)} = \frac{m(q)-1}{m(q)\cdot z_R - 1} \qquad (13)$$

---

for more than half of worldwide spending) and price controls dominate foreign markets, estimates of appropriation based on US markets alone *over-estimate* worldwide appropriation.

[19] Our estimated result on appropriation is consistent with observed expenditures on HIV/AIDS drugs. For example, Lichtenberg (2005) estimates annual drug spending per HIV infected individual in 2001 to be $9,751 in 2001 dollars, implying a net profit of approximately $8,300 for that individual (assuming variable costs are 15% of spending, i.e. $1,400). A value of a life year of $100,000 is consistent with existing estimates and implies an *observed* surplus appropriation of roughly 8% (e.g. $8,300/($100,000 - $1,400)).

[20] To see this, note that π(q) = p(q)q – c(q) and w(q) = g(q) – c(q). Substituting z_R = g(q)/[p(q)q] into the expression for w(q) and simplifying the appropriation share (π/w) yields the above result.

This expression demonstrates that highly cost-effective technologies (those with high z_R) implicitly support low levels of observed surplus appropriation. Moreover, when free-entry is possible and firms earn zero profits (price = average cost), surplus appropriation is zero. The general point, then, is that with information on the degree of market power in an industry, one can use commonly reported CE estimates to infer the degree of appropriation by producers of the relevant technology.

For the case of HIV/AIDS, calculating the appropriation ratio based on our estimates is straightforward. First, recall that we estimated observed gross benefits to consumers to be nearly $980 billion with spending levels of $74 billion. This implies a CE or z_R ratio of roughly 13. Estimates of the average markup can, in turn, be obtained from information on price reductions after patent expiration, which suggests that average costs are as low as 15% of patented prices. Put together, the average markup and estimated CE of HIV/AIDS drugs imply a producer appropriation of *observed* social surplus of 7%, identical to our directly estimated level of appropriation.

With more restrictive cost and demand assumptions, even less information is needed to infer the level of appropriation from CE estimates. Under constant returns to scale and constant elasticity demand, it can be shown that a technology's CE *alone* identifies its elasticity of demand, which in turns identifies the share of surplus appropriated by the producers of that technology. These assumptions also allow us to distinguish between appropriation of two types of surpluses discussed earlier, observed versus potential. To be sure, the *observed* surplus (presented in equation 13) is the surplus which obtains at the market quantity. For example, for a monopoly quantity q_m, the appropriation of observed surplus is simply $\pi(q_m)/w(q_m)$. Alternatively, the *potential* surplus is that which results if the market quantity is determined competitively ($q = q_c$) and hence relates to the total potential surplus available to an innovator. Importantly, the size of profits relative to the *potential* social surplus is most relevant to dynamic policy. For a monopoly quantity q_m, the appropriation of potential surplus is $\pi(q_m)/w(q_c)$. Since

there is a deadweight loss associated with monopoly pricing, the potential surplus from an innovation exceeds the observed surplus. Consequently, estimates of 'surplus' appropriation based on observed surplus will underestimate the deficiency in appropriation by producers of a given technology.

More precisely, consider the common model where variable costs exhibit constant returns, c(y) = cy, and there is a constant elasticity demand curve $p(q) = x/q^{1/\varepsilon}$, where $\varepsilon > 0$ is the elasticity of demand with respect to price and x is a scale factor that shifts demand outward. If q_c and q_m denote the competitive and monopoly output, respectively, the Appendix shows that the ratio of gross benefit to spending (i.e., z_R) under monopoly pricing satisfies:

$$z_R = \frac{g(q_m)}{p(q_m)q_m} = \frac{\varepsilon}{\varepsilon - 1} = \frac{p(q_m)}{c} \qquad (14)$$

In other words, a technology's CE, as described by the ratio of gross benefit to spending, is directly related to the familiar percentage markup of price over marginal cost. In addition, the share of *potential* surplus appropriated as profits under optimal monopoly pricing equals the output expansion due to competition.[21] That is,

$$\frac{\pi(q_m)}{g(q_c) - q_c \cdot c} = \frac{q_m}{q_c} = \left(\frac{1}{z_R}\right)^{\varepsilon} \qquad (15)$$

This interesting result states that, counter-intuitively, the more a monopolist restricts output, as perhaps estimated by patent expirations, the *less* of the surplus it is appropriated.[22] Note that as the elasticity approaches unity (below which profits are infinite) from above, the profits, themselves, rise but as a share of social surplus go to zero.[23] This occurs because the non-

[21] It is straightforward to show that the share of observed surplus appropriated by producers is $(\varepsilon-1)/(2\varepsilon-1)$, which is greater than the potential surplus appropriated.

[22] This result may not be unique to this particular demand structure. For a linear demand curve, it is well known that monopoly output is half the competitive output and that a monopolist always appropriates half the surplus, so that the surplus condition above holds.

[23] It may even be that demand and cost parameters do not affect the share of surplus appropriated by the producer. This is the case when demand is linear (as often estimated) and there are constant returns to scale in production, in which case the share appropriated by producers is always two thirds.

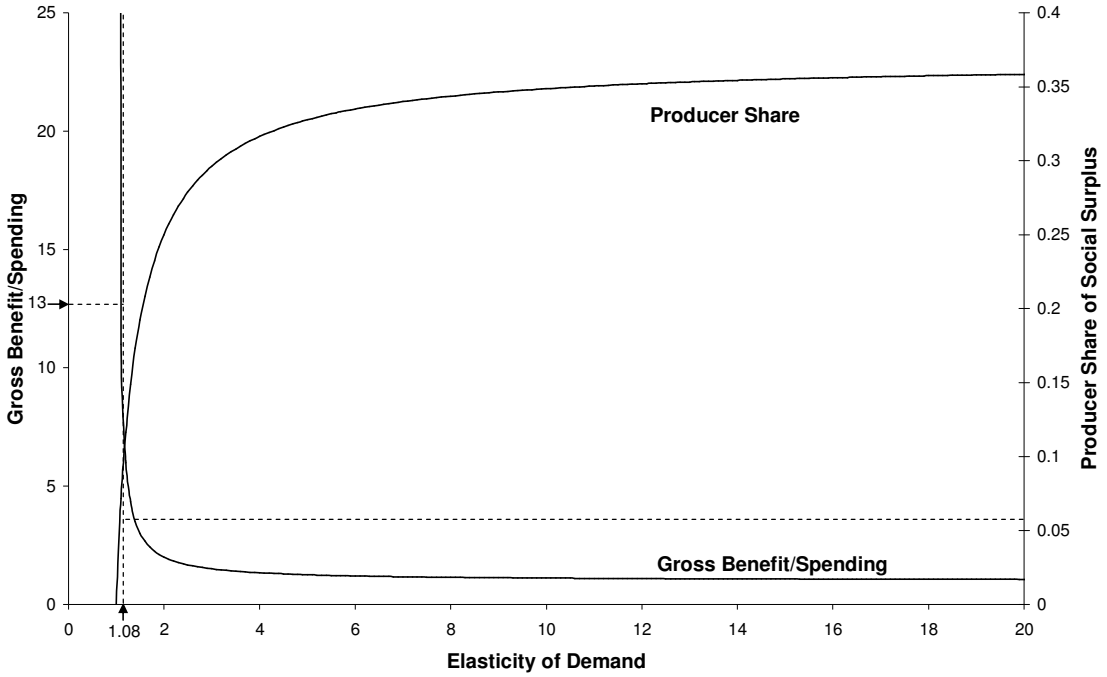
appropriated consumer surplus rises faster than profits as the elasticity falls. Moreover, as market power declines and elasticity approaches infinity, the share of social surplus appropriated as profits tends to roughly 37%.[24] Finally, there is a direct negative relationship between cost-effectiveness and innovator appropriation.

Under these assumptions, a given estimated CE or z_R ratio implies a specific elasticity of demand, which in turn implies the degree to which a firm appropriates social surplus. In the case of HIV/AIDS, for which z_R is roughly 13, the implied elasticity of demand is around 1.08, which (according to equation 16) implies a producer share of potential social surplus of 6%. This can be compared to our *directly estimated* share of potential surplus appropriation of 5%.[25]

More generally, the above relationship between CE and surplus appropriation can be used to infer the share of potential surplus appropriated by those producers whose technologies are examined in existing CE studies. Figure 1, below, graphs the relationship between surplus appropriation, cost-effectiveness, and market power (interpreted as a reduction in the elasticity of demand). As market power decreases, the producer's share of potential social surplus approaches slightly more than a third, while z_R approaches 1. As described earlier, z_R is bounded from below by unity since individuals only purchase goods for which the benefits exceed the costs.

[24] Note that while $(\varepsilon-1)/\varepsilon$ approaches unity as elasticity becomes infinite, $[(\varepsilon-1)/\varepsilon]^{\varepsilon}$ does not do the same.
[25] Given that the share of *observed* surplus appropriated by producers is $(\varepsilon-1)/(2\varepsilon-1)$, an elasticity of 1.08 (implied by the observed z_R of 13 for HIV/AIDS) suggests an observed level of appropriation of 7%, identical to our directly estimated level.

Figure 1: Elasticity of Demand and Producer Shares Implied by CE Estimates



The above figure illustrates how one can potentially use estimates of cost-effectiveness from the large health economic literature to infer the share of surplus appropriated by producers of the relevant technology. For example, consider the technologies to treat HIV/AIDS. With an estimated ratio of gross benefit to spending of roughly 13, this implies an elasticity of demand of 1.08 and a producer share of potential social surplus of a twentieth.

We exemplify this general identification strategy using estimates of cost-effectiveness from the literature. Neumann et al. (2000) review the cost-effectiveness of more than 200 pharmaceuticals using the established "cost-utility" method which focuses on costs per QALY gained and therefore concern both the prolongation and quality of life. The authors note that while no accepted standards exist for how much benefit a technology must confer to be deemed a "good value," the range between $50,000 and $100,000 per QALY has been a benchmark for the US. In the context of our framework, this value (or range) is the gross benefit to consumers of a technology which leads to an additional quality adjusted year of life. Table 2 presents the

spending required to obtain an additional QALY for several interventions reviewed by the authors. For example, an intervention with a price of $1,000 that leads to an increase in 0.2 QALYs requires the same spending per QALY as an intervention with a price of $5,000 that leads to an additional QALY. While the magnitude of gross benefit differs across the two interventions, the gross benefit per QALY is the same (namely in the range described above). Thus, assuming the gross benefit arising from an additional quality adjusted year of life is between $50,000 and $100,000, we can compute estimates of the ratio of gross benefit to spending per QALY for these interventions, as well as the implied shares of social surplus appropriated by producers.

Table 2: Estimated Producer Share of Potential Social Surplus for Several Cost-Effective Technologies

Intervention	Spending per QALY ($)	z_R (Gross Benefit/Spending)		Producer Share of Surplus	
		$50,000	$100,000	$50,000	$100,000
Captopril Therapy	4,000	12.5	25	0.06	0.03
Hormone Replacement Therapy	12,000	4.2	8.4	0.15	0.09
INH Prophylaxis	18,000	2.8	5.6	0.20	0.12
Hip Fracture Prevention	34,000	1.5	3.0	0.30	0.19
Chemotherapy for Breast Cancer	58,000	.9	1.8	---	0.26

Notes: CE and producer share of surplus are presented for two, separate values of an additional quality adjusted life year. The final intervention has a gross benefit less than cost when gross benefit per QALY equals $50,000.
Description of Interventions—1) Captopril therapy in patients with myocardial infarction, 2) Hormone replacement therapy (HRT), 3) Isoniazid (INH) prophylaxis for tuberculosis, 4) Treatment to prevent hip fracture in patients with osteoporosis, and 5) Chemotherapy for breast cancer. For a more detailed description, see Neumann et al. (2000).

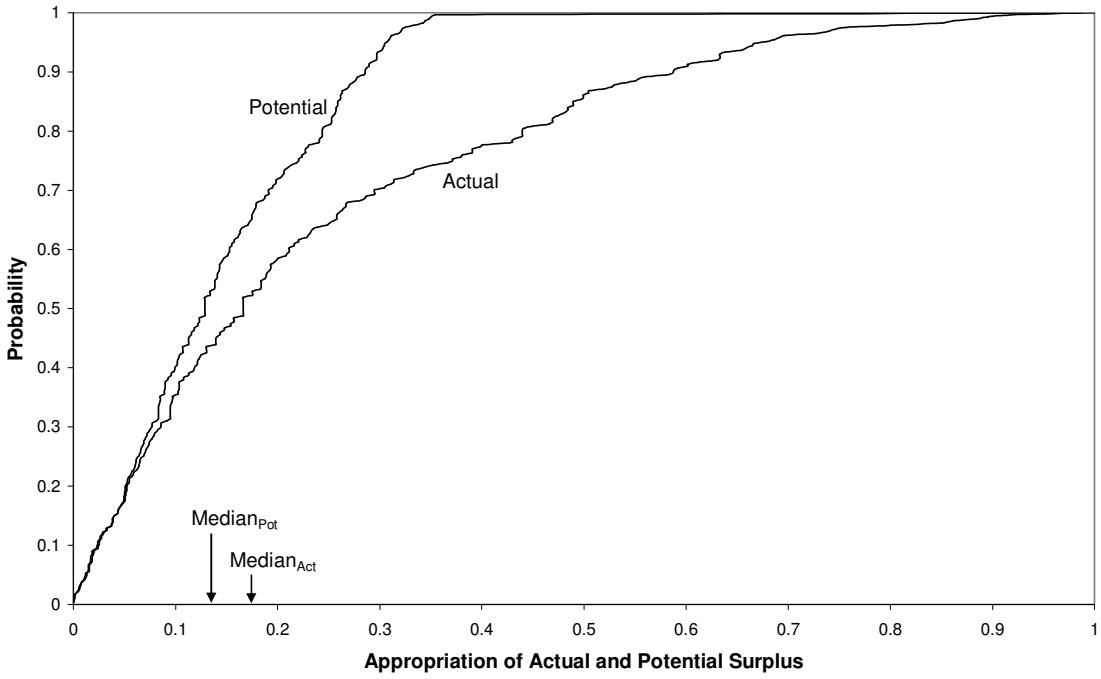
Table 2 demonstrates that, as illustrated by the case of HIV/AIDS, those technologies deemed to be extremely cost effective may also result in low surplus appropriation by producers. For example, the highly cost effective Captopril therapy results in only 3% - 6% of potential social surplus going to producers.

While Table 2 presents estimates of the producer share of social surplus for only five interventions, cost-effectiveness estimates from a large, *random* sample of interventions could be used to estimate the distribution of producer shares. We use data from over 200 published cost-utility analyses contained in the Harvard Cost-Effectiveness Analysis (CEA) Registry to estimate this distribution.[26] Including analyses from 1976 to 2001, the Registry reports the spending per QALY of various interventions compared to benchmark comparator groups. This spending per QALY can in turn be used to estimate the share of potential social surplus appropriated by the producer of that technology, as in Table 2 above.[27] This can be compared to estimates of the producer's actual appropriation, identified by the technology's CE and average mark-up as in expression (13) above. The average mark-up is based on our earlier assumption that average costs are 15% of patented prices. Figure 2 plots the distribution of observed and potential producer shares for the interventions considered. Because the studies included in the Registry may not be a random sample of all technologies, however, we can only estimate the distribution of producer shares conditional on inclusion into the Registry.

[26] The Registry is not limited to only pharmaceutical interventions. More detailed information can be found at: http://www.hsph.harvard.edu/cearegistry/

[27] For these calculations, we assume the gross benefit of an additional QALY to be $100,000. Consequently, we limit our attention to those interventions with published costs of less than $100,000 per QALY gained.

Figure 2: Cumulative Distribution of Actual and Potential Surplus Appropriation



Since the constant elasticity of demand assumption predicts a producer appropriation of social surplus of no more than 37%, all interventions considered in Figure 2 have estimated producer shares less than this amount. The median intervention requires a spending per QALY of roughly $19,000, which corresponds to a producer share of potential (actual) social surplus of nearly 13% (17%).[28] Moreover, 25% of the interventions considered have estimated potential appropriations of less than 7%, while 75% have appropriations less than a fourth. Moreover, 75% of the interventions have an actual appropriation of less than 40%. If the estimated distribution of producer shares generalizes to the distribution across all health interventions (i.e., not only those included in the Registry), our empirical finding for producers of HIV/AIDS drugs suggests their appropriation of potential social surplus is at the twentieth percentile.

Our estimates of appropriation can be compared to alternative, theory-based methods of calculating this share. Specifically, given the previously described relationship between the

[28] If the gross benefit of an additional QALY is assumed to be $50,000 (rather than $100,000), the median intervention has an implied producer share of social surplus closer to 20%.

elasticity of demand and the share of potential social surplus appropriated by innovators, one can use information on price reductions after patent expiration to estimate patent-protected markups (Caves et al. (1991).[29] These markups identify the elasticity of demand for the patent-protected drugs and thus the share of surplus allocated to the producer. In particular, the larger is the price reduction upon patent-expiration, the lower is the elasticity and the *smaller* is the share of surplus allocated to the producer. Existing estimates suggest that price reductions are on the magnitude of 85% percent, implying a demand-elasticity around 1.17. This elasticity implies a producer share of social surplus of 10%, which is highly related to our major finding that the share of social surplus appropriated by R&D investors in this area is, in fact, quite low. This is true even though prices for these drugs are high, presumably due to the inelastic nature of demand.

V. Concluding Remarks and Future Research

We argued that popular technology assessment criteria in health care going under the rubric of "cost-effectiveness" are often implicitly concerned with maximizing the observed level of consumer surplus, which is many times consistent with maximizing static efficiency after an innovation has been developed. Dynamic efficiency, however, aligns the social costs and benefits of R&D and is therefore determined by the how much of the social surplus from a new technology is appropriated by innovators. For the case of HIV/AIDS, our earlier estimates suggested that producers appropriated only 5% of the social surplus arising from new drug therapies. Given the low degree of appropriation by producers of the highly cost-effective HIV/AIDS therapies, we showed how other CE estimates in the literature could be related to the standard framework—our main finding was that these CE estimates implicitly support a low degree of surplus appropriation by producers, comparable to our directly measured estimates for HIV/AIDS. Despite the high annual costs of these drugs to patients, the low share of social

[29] Caves, Whinston, and Hurwitz (1991) estimate that with 20 generic competitors, the ratio of prices between generic- and brand-drugs is roughly 17%. We use the price of generic drugs as an upper bound of the marginal costs of production.

surplus going to innovators raises concerns about advocating cost-effectiveness criteria that would further reduce appropriation share, and hence further reduce dynamic efficiency.

In addressing why producer surplus is so small, and why the CE of therapies may be inefficiently high, it seems natural to suggest that this is ultimately due to a lack of profits and market power. This is potentially due to prices being held down by: 1) the threat of public regulation if pharmaceutical companies raise prices, or 2) patents that are weakly enforced or too narrowly defined to allow patent-protected monopolies to raise price appropriately. However, even with free pricing and nearly-inelastic demand, the share of social surplus allocated to producers may be small. This is most easily illustrated by the constant elasticity case, in which a producer share of social surplus of 5% is consistent with monopoly pricing under a demand curve that is almost as inelastic as it can be, $\varepsilon = 1.08$, as an elasticity below unity, of course, leads to infinite profits. In fact, we showed that higher prices (such as those of HIV/AIDS drugs) induced by lower elasticities of demand often lead to less surplus captured by inventors. Put differently, even though *profits*, of course, rise as the elasticity of demand falls, many times the *share* of social surplus appropriated by the monopolist *falls*. This occurs because the non-appropriated consumer surplus rises faster than profits as the elasticity falls.

It is important to stress that arguments about the difference between static and dynamic efficiency are a different matter than whether prices used for calculating spending in CE analysis reflect costs of production in general, and average costs of production (reflecting R&D costs), in particular. Under traditional CE analysis, even if one could measure costs perfectly, and did not need to approximate unobserved costs by observed prices, one would be concerned with the wrong measure, total ex-post surplus. This is because the *division* of the surplus is what matters for dynamic R&D policy, as opposed to only the *total* surplus which is relevant for static policy. In particular, this holds true whether the costs represented are marginal or average costs, the latter potentially including fixed costs such as R&D. In both cases, the division of social surplus is ignored but is what drives optimal R&D policy.

Several issues may be important in generalizing our conclusions and are therefore suitable for future research. The first concerns the interpretation of CE analysis in a non-monopoly context; the field of "industrial organization of technology adoption" needs to be better understood. Another concern is the effect of altruism, which seemingly motivates much of public financing, on optimal technology adoption and the efficient form of surplus appropriation.[30] A third concerns the effect of ex-post inefficiencies such as moral hazard. Fourth, the impact of the joint demand of physicians and patients on observed levels of CE must be examined further. Fifth, the effect of improved treatment on disease prevalence, whether through increased life-expectancies among infected individuals or increased risky behavior (due to lower costs of infection induced by treatment) among non-infected individuals, must be considered (see e.g. Philipson (2000)). Sixth, the role of public funding, comprising almost half of US medical R&D spending, on the optimal degree of appropriation is not well understood. While much basic research in the US is financed by tax-payers (mainly through the NIH), little analysis exists on the implications of that for optimal appropriation.[31]

Our analysis and evidence, if they generalize to other technologies, suggest that interventions aimed at raising innovator appropriation may be desirable even though full appropriation may not be. This would lead to lower observed levels of cost-effectiveness for these technologies as traditionally measured from market prices, but would raise efficiency by not unduly sacrificing the health and well being of future patient populations for the benefit of current ones.

[30] Philipson, Mechoulan, and Jena (2006) discuss optimal technology assessment in the presence of altruism that motivates public health care delivery, in general, and R&D into third-world diseases, in particular.

[31] The discrimination between public and private funding may be mitigated by private expenditures towards the licensing of publicly-funded discoveries.

Appendix

Assume a constant elasticity demand function and constant returns to scale as in:

$$p(q) = \frac{x}{q^{1/\varepsilon}} \qquad\qquad c(q) = cq$$

where $\varepsilon > 0$ is the elasticity of demand with respect to price, and x is a demand shifter. This results in an optimal quantity and price of

$$q_m = \left[\frac{c \cdot \varepsilon}{x \cdot (\varepsilon - 1)}\right]^{-\varepsilon} \qquad\qquad p_m = \frac{c \cdot \varepsilon}{\varepsilon - 1}.$$

The gross consumer benefit $g(q_m)$ can be expressed by the following formula:

$$g = \int_0^{q_m} p(q)dq = \frac{x \cdot \varepsilon}{\varepsilon - 1} \cdot (q_m)^{\frac{\varepsilon - 1}{\varepsilon}}.$$

Similarly, the maximized profit can be written as:

$$\pi = p(q_m) \cdot q_m - c \cdot q_m = \frac{c \cdot q_m}{\varepsilon - 1}.$$

We can now determine the share of profits in potential social surplus, i.e. the social surplus that obtains in perfect competition with p = c. Specifically,

$$\frac{\pi(q_m)}{g(q_c) - q_c \cdot c} = \frac{(cq_m)/(\varepsilon - 1)}{\frac{x\varepsilon}{\varepsilon - 1}(q_c)^{\frac{\varepsilon - 1}{\varepsilon}} - q_c \cdot c} = \frac{(cq_m)/(\varepsilon - 1)}{q_c\left(\frac{x\varepsilon}{\varepsilon - 1}(q_c)^{\frac{-1}{\varepsilon}} - c\right)} = \frac{(cq_m)/(\varepsilon - 1)}{q_c\left(\frac{x\varepsilon}{\varepsilon - 1}\left(\left(\frac{x}{c}\right)^{\varepsilon}\right)^{\frac{-1}{\varepsilon}} - c\right)}$$

$$= \frac{(cq_m)/(\varepsilon - 1)}{q_c\left(\frac{c\varepsilon}{\varepsilon - 1} - c\right)} = \frac{(cq_m)/(\varepsilon - 1)}{q_c\left(\frac{c}{\varepsilon - 1}\right)} = \frac{q_m}{q_c}$$

That is, the share of profits in potential social surplus is equal to the ratio of the monopolist output to the competitive output. In terms of the exogenous parameters, this simplifies to:

$$\frac{\pi(q_m)}{g(q_c) - q_c \cdot c} = \left(\frac{\varepsilon - 1}{\varepsilon}\right)^{\varepsilon}$$

Using the above expressions, it is straightforward to derive the ratio of gross benefit to spending, z_R, as well:

$$z_R = \frac{g(q_m)}{p(q_m)q_m} = \frac{\varepsilon}{\varepsilon - 1}$$

References

Arrow, Kenneth, "Economic Welfare and the Allocation of Research for Invention," in The Rate and Direction of Inventive Activity: Economic and Social Factors," E. R. Nelson, Princeton University Press, 1961.

Becker, Gary S., Philipson, Tomas, Soares, Rodrigo, "The Quantity and Quality of Life and the Evolution of World Inequality," American Economic Review, 2005, Vol. 95, pp. 277-291.

Caves, R., Whinston, M., Hurwitz, M., "Patent Expiration, Entry, and Competition in the U.S. Pharmaceutical Industry," Brookings Paper on Microeconomic Activity, Microeconomics, 1991, pp. 1-66.

Cutler, David M., Your Money or Your Life: Strong Medicine for America's Health Care System, Oxford University Press, New York, New York, 2004.

Cutler, David M., and Richardson, Elizabeth, "The Value of Health: 1970-1990," American Economic Review, 1998, Vol. 88, pp. 97-100.

Cutler, David M., and McClellan, Mark, "Is Technological Change in Medicine Worth It?," Health Affairs, 2001, Vol. 20, pp. 11-29.

Duggan, Mark, and Evans, William, "The Impact of HIV Antiviral Treatments: Evidence for California's Medicaid Population," Working Paper, 2005.

Drummond, M.F., O'Brien, B., Stoddart, G.L, and Torrance, G.W., Methods for the Economic Evaluation of Health care Programmes, Oxford University Press, 1997.

Health, United States; National Center for Health Statistics with Chartbook on Trends in the Health of Americans, Hyattsville, MD, 2002.

Garber, Alan M. and Phelps, Charles E., "Economic Foundations of Cost-Effectiveness Analysis," Journal of Health Economics, 1997, Vol. 16, pp. 1-32.

Garber, Alan M., "Advances in Cost-Effectiveness Analysis of Health Interventions," NBER Working Paper 7198, 1999.

Gold, M.R., Siegel, J.E., Russell, L.B., and Weinstein, M.C., Cost-Effectiveness in Health and Medicine, Oxford University Press, 1996.

Griliches, Zvi. R&D and Productivity: The Econometric Evidence, University of Chicago Press, 1998.

Hall, Bronwyn H., "The Private and Social Returns to Research and Development," Technology, R&D, and the Economy, ed. B. Smith and C. Barfield, Brookings Institution/American Enterprise Institute, 1996.

Johannesson, M., and Weinstein, M.C., "On the Decision Rules of Cost-Effectiveness Analysis," Journal of Health Economics, 1993, Vol. 12, pp. 459-467.

Kates, J., and Wilson, A. "Medicaid & HIV/AIDS," Pub. 7172, Henry J. Kaiser Family Foundation, 2004.

Lakdawalla, Darius, and Philipson, Tomas J.," Intellectual Property and Non-Price Discrimination," University of Chicago, Department of Economics Working Paper, 2005.

Lichtenberg, Frank R., "The Impact of Increased Utilization of HIV Drugs on Longevity and Medical Expenditure: An Assessment Based on Aggregate U.S. Time-Series Data," Working Paper, Columbia Business School, 2005.

Levin, Richard C., Klevorick, Alvin K., Nelson, Richard R., Winter, Sidney G., Gilbert, Richard, and Griliches, Zvi, "Appropriating the Returns from Industrial Research and Development," Brookings Papers on Economic Activity, 1987, Vol. 3, pp. 783-831.

Manning, Richard L., "Changing Rules in the Tort Law and the Market for Childhood Vaccines," Journal of Law and Economics, 1994, Vol. 37, pp. 247-275.

Mansfield, Edwin, "How Rapidly Does New Technology Leak Out?," The Journal of Industrial Economics, 1985, Vol. 34, pp. 217-223.

Mansfield, Edwin, Rapoport, John, Romeo, Anthony, Wagner, Samuel, and Beardsley, George, "Social and Private Rates of Return from Industrial Innovation," The Quarterly Journal of Economics, 1977, Vol. 91, pp. 221-240.

Meltzer, David, "Accounting for Future Costs in Medical Cost-Effectiveness Analysis," Journal of Health Economics, 1997, Vol. 16, pp. 33-64.

Nordhaus, William D., "Schumpeterian Profits in the American Economy: Theory and Measurement," NBER Working Paper, 2004.

Phelps, Charles E., and Parente, Stephen T., "Priority Setting in Medical Technology and Medical Practice Assessment," Medical Care, 1990, Vol. 28, pp. 703-723.

Philipson, T.J., "Economic Epidemiology and Infectious Disease," Chapter in Handbook of Health Economics. Edited by J. Newhouse and A. Culyer, Elsevier B.V., North Holland, 2000.

Philipson, T., and Jena, A.B., "Who Benefits from Medical Technologies? Estimates of Consumer and Producer Surpluses for HIV/AIDS Drugs," BE Press Forum for Health Economics and Policy, Forum: Biomedical Research and the Economy, Article 3, 2005. (Also NBER Working Paper #11810)

Philipson, T., Jena, A.B., and Mechoulan, S., "Intellectual Property & External Consumption Effects: Generalizations from Pharmaceutical Markets," NBER Working Paper #11930, 2006.

Pharmaceutical Research and Manufacturers of America, PhRMA Annual Membership Survey, 2003.

Neumann, Peter J., Sandberg, Eileen A., Bell, Chaim M., Stone, Patricia W., and Chapman, Richard H., "Are Pharmaceuticals Cost-Effective? A Review of the Evidence," Health

Affairs, 2000, Vol. 19, pp. 92-109.

Shapiro, Martin F., et al. “Variations in the Care of HIV Infected Adults in the United States,” Journal of the American Medical Association, 1999, Vol. 281, pp. 2305-2316.

Tirole, Jean. The Theory of Industrial Organization, The MIT Press, 1988.

Weinstein, Milton C., and Manning, Willard G., Jr. "Theoretical Issues in Cost-Effectiveness Analysis" Journal of Health Economics, 1997, Vol. 16, pp. 121-128.

Weinstein, M.C., and Stason, W.B., “Foundations of Cost-Effectiveness Analysis for Health and Medical Practices,“ New England Journal of Medicine, 1977, Vol. 296, pp. 716-721.