

**Who Benefits from New Medical Technologies?
Implications for the Science-of-Science Research Agenda**

by

Tomas J. Philipson¹

and

Cindy Yuan

The University of Chicago

November 15, 2010

PRELIMINARY AND INCOMPLTE-PLEASE DO NOT QUOTE

¹ Corresponding Author: Daniel Levin Professor of Public Policy, The 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

Technological change is often argued to be a central force behind the growth in health care spending. The value of a new technology may be divided between the surpluses of consumers, who may value the technology more than the price they pay, and producers or innovators, whose costs may be lower than the price charged. As is well known in economics, consumer surplus is central to *static* efficiency after an innovation has been discovered, while producer surplus, which forms the incentive for firms to engage in costly R&D to bring an innovation to market, is central to *dynamic* efficiency. Therefore, understanding the degree to which innovation in health care benefits consumers versus producers is important for policies aimed at trading off the two forms of efficiency, such as science policy involving both public funded R&D (e.g. through NIH) as well as policies affecting the value of intellectual property (e.g. through FDA and CMS).

In this paper, we discuss some recent evidence on how much innovators appropriate the social value of health care technology and discuss the implications of these findings for the science-of-science research agenda. We first review frameworks for how to assess the social value of new medical technologies. We then discuss findings indicating that the estimated share of social value that innovators appropriate is fair small. In fact, most estimates show that the social returns from R&D are several of orders of magnitudes larger than the private returns to companies. We conclude with a listing of research agendas to further address how these findings could inform national science policy..

Section 2: The Basic Economics of Innovation in Medical Technologies

The most common measure of the value of a medical technology is its cost-effectiveness (CE). It states a higher value for technologies for which the static health benefits to patients outweigh the costs, whether they are actual costs of production or, as more commonly used, the prices paid by consumers and public payers. Estimates of CE may be conducted at several levels and from different perspectives: for example, by health plans choosing technologies to be covered for their members or by nations financing care for their citizens.

In practice, the benefits and costs used in CE measures are incremental—for example, a given procedure improves health by one quality-adjusted life year (QALY) at a price of \$50,000 compared with a baseline therapy. The benefit is the value of the additional QALY, which comes at an additional cost of \$50,000. The most cost-effective technologies are those for which the incremental benefits far outweigh the additional costs to the health care system. Translating this into more traditional economic concepts, technologies are most cost-effective when the associated consumer surplus—the health benefits to consumers net of the price paid—is also large. Consumer surplus concerns the difference between benefits and costs, versus CE criteria, which concern the ratio.

Consumer surplus differs from social surplus by the amount of producer surplus or profits. The social surplus can be illustrated with a simple example of supply and demand for a given drug therapy. The demand curve for a drug reflects society's willingness to pay for a given level of provision. Its magnitude depends on several factors, one being the price (or availability) of other related technologies. For example, if the demand curve

depicted the willingness to pay for loop diuretics, the demand curve's magnitude would depend on the price of substitute treatments—for example, angiotensin-converting enzyme (ACE) inhibitors. The absence of good substitutes would result in a larger willingness to pay for diuretics. One can then interpret the demand curve for loop diuretics as identifying the incremental benefit of these drugs for a given price of alternative treatments.

The area under the demand curve is the gross benefit to patients from consuming the drug. As more patients consume the drug, the gross benefit increases. The amount by which this benefit exceeds the price paid is the consumer surplus. We interpret the cost-effectiveness of this particular drug as the ratio of the gross incremental benefit to the total amount paid. The higher the gross benefit over the total amount paid, the higher its cost-effectiveness. The main implication of our analysis is that a drug's cost-effectiveness and consumer surplus are intimately related: The higher the consumer surplus, the higher the cost-effectiveness.

On the supply side, it is well known that the marginal costs of drug production are quite low, often on the order of cents per pill. If markets are competitive (price equals marginal cost), the net gains to consumers may be large. Both the consumer surplus and cost-effectiveness of the drug are high, primarily because the benefit to patients far outweighs the price they pay. Moreover, the manufacturers of the drug make zero economic profits since the price of the drug equals its marginal cost. When markets are not competitive—as is commonly true of markets for new medical products that are patented—and

producers charge prices that exceed marginal costs, the consumer surplus is lower, and producers earn variable profits to make up for the fixed costs of bring the technology to market. The extent to which these profits compare to consumer surplus defines how the gains from innovation are divided.

Static vs. Dynamic Efficiency

After a drug has been discovered, static efficiency implies society is best off if the price of the drug equals its marginal cost. The total quantity of drug supplied and consumed is at its highest, and drugs are only consumed by those whose benefit exceeds the cost of production. In this case, both consumer surplus and cost-effectiveness are high, and increases in price above marginal cost lower the welfare of current consumers.

Dynamic efficiency takes into account that drug discovery is an expensive ordeal, plagued by uncertainty in both the process of discovery and the ultimate effectiveness of the final product. When R&D is costly, companies require incentives to innovate, whether these incentives take the form of higher profits, subsidies for R&D, or some combination of the two. Higher profits come at a cost to current patients, health plans, and governments that pay higher prices. Higher profits, however, also stimulate innovation and are therefore beneficial to future patients.

This has the key implication that high levels of cost-effectiveness are often inconsistent with dynamic efficiency defined by the highest level of access to therapy by not only current but also future consumers. This can be illustrated with a simple, although extreme, example of perfect price discrimination. In this case, dynamic efficiency holds,

health is maximized because all consumers buy the product, but cost-effectiveness is minimized. In fact, dynamic efficiency is the justification for the patent system, which reduces static efficiency by creating monopolies in the name of innovation. As a side note, the common distinction between consumers and producers are somewhat artificial when patients themselves hold large stakes in companies and therefore benefit from increased profits, either directly as employees or indirectly through their pension plans, mutual funds, and other investments in these very same companies.

Section 2: Empirical Studies of Innovator Appropriation in Health Care

This section discusses empirical studies of the degree of innovator appropriation in health care.

2.1 Appropriation for HIV and AIDS

Philipson and Jena (2005) investigates this issue for a major breakthrough in medicine—the new drugs to treat HIV/AIDS that came on the market in the late 1980's. HIV/AIDS is an important case to consider in and of itself, partly because it is a major disease target of public sector R&D in the US. The benefit to consumers was calculated as the discounted sum of the monetary value of increased survival due to the arrival of antiretrovirals (ARVs). In order to compare the flow of survival gains with the one-time R&D costs, all future HIV-infected individuals must also be included in the consumer surplus calculation. Producer surplus (or “profit”) was calculated based on IMS sales data, existing estimates of markups for branded medications, and the patent-protected lifetime. For the new HIV drugs that came about during this period, their major finding was that innovators captured only 5% of the social surplus arising from these new

technologies. More precisely, consumer and producer surplus from these drugs amounted to roughly \$1.33 trillion and \$63 billion, respectively. In other words, innovators received (appropriated) less than a third of the total social surplus that they generated. They argued that if the new HIV/AIDS therapies are representative of other technologies, the lack of appropriation by innovators has strong policy implications for how to adopt and evaluate new health care technologies. Despite the high prices of many therapies such as the new HIV drugs, patients and health plans are getting too good a deal in the short run which, of course, hurts them in the long run because pharmaceutical companies have insufficient incentive for additional R&D efforts.

The surplus values that they estimated can be understood by some simple back-of-the-envelope calculations. For the size of the consumer surplus, consider the 1.5 million US citizens have been infected by HIV since the start of the epidemic, some of who died before drug therapy became available, some who lived until the advent of ARVs, and others who contracted HIV after the breakthrough drugs entered the market in the mid 1990's. Averaging across all such cohorts, the gain in life-expectancy has been at least 5 years. Using a fairly low estimate of the value of a life-year of \$100,000, the added survival has been worth more than \$500,000 per individual and \$750 billion in aggregate. This figure, of course, does not include the benefit to those individuals who will become infected with HIV in the future but can benefit from drugs that have already been introduced to the market—doing so, while assuming current incidence rates persist in the future, raises the total consumer value of these drugs above \$1 trillion.

For the size of the producer surplus, consider that sales of HIV/AIDS drugs have grown from \$1 billion to \$4 billion annually since the breakthrough drugs came on the market in 1996. From these revenues, one can apply appropriate discounting of the future (based on interest rates that account for the growth rate of money's present value) to compute a present value of sales of \$74 billion, assuming that drugs sell at current levels in the future. We can then subtract the variable costs of production, which are approximated to be 15% of revenues based on estimates of markups stemming from differences in drug prices pre- and post patent expiration. Net costs of production, we arrive near their estimate of producer surplus, \$63 billion.

Currently, the NIH spends about three billion dollars annually on HIV/AIDS research. Total federal spending has been growing steadily since 1995, so it was previously much less, but even if we consider \$3 billion per year for 25 years, that spending pales in comparison to the \$1.33 trillion gain in consumer surplus. To do a simple calculation of return on investment for NIH dollars, if we consider that NIH funding is responsible for all of the gain in consumer surplus, this would give a 17.7 fold return. Assuming NIH funding is only responsible for 25% of the gain from HIV/AIDS research, there is still a gross return-on-investment of 4.4.

2.2 Appropriation in Oncology

Looking specifically at cancer, Lakdawalla et al. (2010) considers R&D's ability to prevent disease (vaccines, behavioral changes), increase screening and early detection, and directly improve survival. Their work quantifies the costs and benefits of cancer

R&D, focusing on the benefits of earlier detection of malignant disease and improved cancer therapies. They focus on the value of improved cancer survival from 1988 to 2000, well after R&D initiated in the early 1970s came to fruition. They calculated that the average newly diagnosed cancer patient, earning the average US income, would be willing to pay approximately \$31,000 annually to retain survival prospects facing patients in 2000, as compared to the prospects available in 1988. This is equal to roughly half of full income. Given the absolute magnitude of a gain in longevity, greater value is placed on it when an individual has a shorter life expectancy (Becker et al., 2007). Therefore, the high rates of mortality associated with cancer magnify the value of even small absolute survival gains. Consider the following example: since the value of consumption and holding wealth are significantly lower after death, individuals may be willing to pay nearly their entire end-of-life wealth for as little as a few extra weeks of life. In addition, their results also suggest that wealthier individuals would be willing to pay a larger share of their income for these higher survival prospects. In economic parlance, this implies that cancer treatment behaves like a “luxury good,” in the sense that higher income individuals are willing to spend a greater share of total income on it.

Aggregating over patients, they estimate that these improvements in survival generated approximately 23 million additional life-years for patients, valued at \$1.9 trillion. These numbers imply that cancer patients were willing to pay \$86,000 for the average life-year gained, which is well within the range of conventional estimates for the value of a statistical life-year. When comparing this total benefit of survival gains with the cost of cancer treatment (defined as aggregate spending on cancer treatment, combined with

research and development costs), they counted the total cost of (and profits from) cancer care provided from 1988 to 2000 and all cancer R&D spending from 1970 to 2000, determining that cancer care providers (drug companies, hospitals, doctors, and health professionals) earned at most \$433 billion in profits over this time period, while the net surplus to patients was approximately \$2.5 trillion.

Their approach is, if anything, deliberately conservative in estimating the rate of return earned by the war on cancer. First, they do not incorporate the value generated by cancer prevention, only counting the benefits that accrue to individuals who acquire cancer. Second, while counting all cancer R&D expenditures from 1970 onwards as the total size of investment, they only count net benefits (gains in survival, less medical costs) from 1988 to 2000. Analysis was restricted partly due to data limitations for earlier years, but also to recognize the lags inherent in the medical R&D cycle. For instance, if we consider an advance that was developed in 1970, it would be inappropriate to start counting benefits immediately as it would take years from discovery to market. Typically, drug development takes 10 to 15 years from inception to launch (DiMasi et al., 2003), so enforcing an 18-year lag thus seems a conservative assumption. Moreover, they include R&D costs from the late 1990s for products that are not yet on the market by the end of the window for tallying health improvements. This further underestimates the gains from innovation since the benefit from this R&D has not yet been realized or measured, though costs have already been included.

2.3 Generalizing by Inferring Appropriation from Empirical CE Studies

In a follow-up paper, Jena and Philipson (2006) apply similar methods towards analyzing a wider class of therapies. Specifically, they look at captopril (MI), ticlopidine (stroke), and mesalamine (Crohn’s disease). They find that these technologies are extremely cost effective and also result in low surplus appropriation by producers. Table 1 presents the spending required to obtain an additional QALY for each intervention, under patent at the time of the original study. For example, an intervention with an incremental price of \$1000 that leads to an increase of 0.2 QALYs requires the same incremental spending per QALY as an intervention with an incremental price of \$5000 that leads to one additional QALY. While the magnitude of gross benefit differs across the interventions, the gross benefit per QALY is the same (both cost \$5000-per-one QALY). Thus, assuming the gross benefit arising from an additional quality adjusted year of life is between \$50,000 and \$100,000, we can compute monetized versions of these cost effectiveness estimates, as well as the implied shares of potential social surplus appropriated by producers.

Table 1: Jena and Philipson (2006)

Intervention	Spending per QALY	Cost-benefit ratio (1 QALY valued at \$50,000 or \$100,000)		Producer share of potential social surplus	
		\$50,000	\$100,000	\$50,000	\$100,000
Captopril	\$4,000	0.08	0.04	0.06	0.03
Ticlopidine	\$48,000	0.96	0.45	0.36	0.24
Mesalamine	\$6,000	0.12	0.06	0.09	0.04

Table 1 demonstrates that 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 roughly 3–6% of potential social surplus going to producers.

While Table 1 presents calculations of the producer share of social surplus for only three interventions, cost-effectiveness estimates from a larger sample of interventions could be used to infer the overall distribution of producer shares of social surplus across all innovations. They illustrate this using data from over 200 published cost-utility analyses contained in the Harvard Cost-Effectiveness Analysis (CEA) Registry. It is important to recognize that the studies included in the CEA Registry are not random and, therefore, cannot be expected to yield a representative distribution of innovator appropriation.

Nonetheless, they do provide an initial benchmark to *illustrate* the levels of appropriation that may already exist in this selected sample and how one can convert cost-effectiveness studies into appropriation assessments. Including analyses from 1976 to 2001, the Registry reports the spending per QALY of various interventions, as compared to benchmark comparator groups. This spending per QALY can, in turn, be used to calculate the share of potential social surplus appropriated by the producer of that technology, as in Table 1 above. This can be compared to calculations of the producer's actual appropriation, identified by the technology's cost-effectiveness and average mark-up. As a simplifying assumption, they apply existing estimates of markups for brand-name drugs (as estimated from patent expirations) to approximate variable costs as 15% of sales (see e.g. Caves et al., 1991).

Figure 1, from Jena and Philipson (2006), plots the distribution of observed and potential producer shares for the interventions considered. 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%). Approximately 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%. These calculations can be compared to directly estimated levels of appropriation for producers of HIV/AIDS drugs presented in Philipson and Jena (2005). Given that the previous paper found that firms appropriated roughly one-twentieth of the social surplus generated by these technologies, HIV/AIDS therapies are at the 20th percentile of appropriation in the CEA Registry.

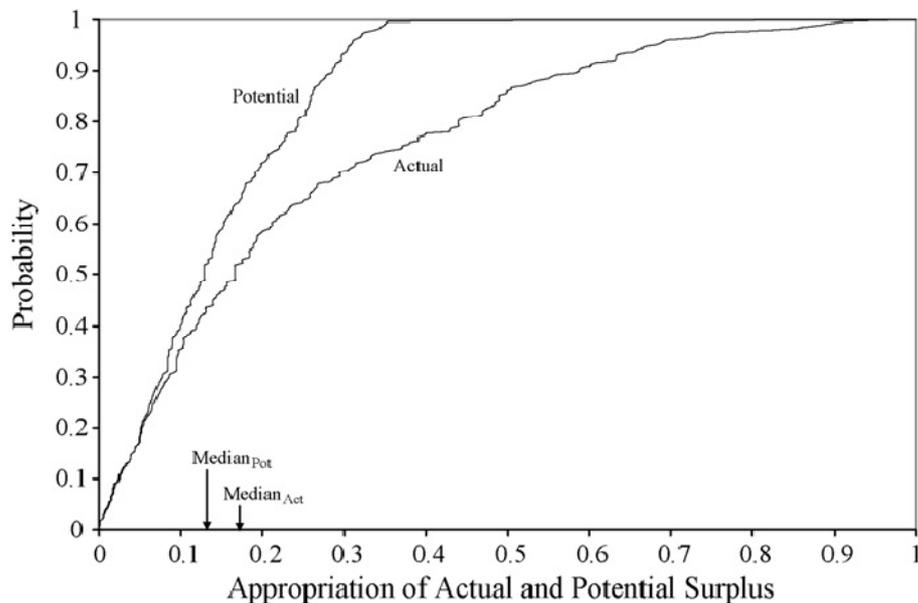


Figure 1: Cumulative distribution of actual and potential surplus appropriation.

Section 3: Future Research in Valuing Public Investments in Health Care R&D

The empirical review of innovator appropriation suggests several areas of importance for a research agenda on the value and effects of public R&D subsidies in health care.

3.1 Better Assessment of Optimal Appropriation under Public R&D Subsidies

Public R&D affects the optimal amount of innovator appropriation by lowering the share of total social surplus that is obtained by the innovator. In a market with firms choosing an optimal level of R&D without public subsidies, economic theory implies that the consumer surplus should be *minimized* to enhance dynamic efficiency, as opposed to maximized under a cost-effectiveness criterion. However, when publicly funded R&D comprises a significant portion of total R&D, like the NIH's influence in the US health care market, the optimal private R&D (and hence, appropriation) should be lowered by the amount of the public subsidy. Since the marginal product of private R&D is decreasing in the level of subsidized R&D, private R&D (and hence appropriation) falls as its public counterpart increases. The less-than-full appropriation that is optimal under public subsidization of R&D implies that lower cost-effectiveness thresholds may be preferred in settings where the total R&D budget relies on a large public contribution. Of course, increases in cost-effectiveness thresholds may not be practically feasible, particularly when fixed public budgets impose a necessary trade-off between static and dynamic efficiency that precludes the raising of cost-effectiveness thresholds. Further research needs to be done on how NIH subsidization of R&D, in conjunction with policy practices by the FDA and CMS, impact the incentives for private innovation.

3.2 Better Assessment of Value of Intellectual Property under Public Pricing

A better understanding is needed of the impact of public reimbursement regulations, primarily through Medicare and Medicaid, on innovative returns, appropriation, and, hence, R&D incentives.

There is a longstanding and vast health economics literature that attempts to assess the value of spending on new technologies by use of cost-effectiveness, cost-utility, or cost-benefit analysis, hereafter referred to collectively as CE analysis

(Johannesson(), Garber(), Weinstein()). There is a growing emphasis on using such analysis to guide new technology adoption and manage its impact on long term health care spending. As the name suggests, CE analysis can offer governments and private payers a quantitative way to allocate often scarce health care resources based on the costs and effectiveness of available medical technologies.

In practice, CE analysis so far has guided policy decisions in the form of thresholds that determine technology adoption rates: a given technology will be reimbursed only if the incremental costs per QALY is below a given threshold. Currently, this type of analysis already plays a role in public reimbursement decisions outside the US. For example, both the UK's National Institute for Clinical Excellence (NICE) and Australia's Pharmaceutical Benefits Advisory Committee have been reported to follow CE thresholds in technology adoption decisions. In Australia, for example, only 2 out of 26 submissions were accepted for reimbursement whose cost per-life year saved exceeded US\$ 57,000—similarly, only 1 out of 26 submissions was rejected whose cost per life-year

saved was less than US\$ 32,000 (Bethan et al., 2001). Similarly, in a review of NICE determinations for which the cost per QALY saved was stated, Raftery (2001) finds that, with the exception of one drug, all recommended technologies had a cost per QALY saved less than £30,000. Such explicit thresholds for adopting medical technologies are not used in public coverage and reimbursement decisions by the Centers for Medicare and Medicaid Services (CMS) in the US. However, their use has been discussed extensively and it seems reasonable to hypothesize that, de facto, technologies that cost more and offer fewer health benefits are more closely scrutinized before they are adopted, if at all.

Given the widespread and growing use of explicit or implicit CE analysis for guiding technology adoption and its creeping influence of US reimbursement policy, more research is needed on how these CE criteria are distorting the incentives for bringing technologies to market in the first place as profitability declines. The central theme of standard CE assessments in practice seems to be a measure *static* consumer surplus or net consumer benefits—technologies are deemed more valuable the larger that the patient health benefits are, above what is spent on them. However, when new technologies are brought to life via costly R&D, consumer surplus, as discussed, may be a poor tool for inducing optimal R&D investments. Rather, the degree to which producers can capture 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, for the same reason that patents are preferred even

though they lower consumer surplus after technologies are discovered, we should prefer technology adoption criteria that focus on more than just consumer surplus, as CE criteria do. Put differently, even though measured levels of cost-effectiveness would be larger without patents, since patients or health plans would spend less to get the same technology, dynamic efficiency would presumably be lowered. An illustrative case of this may be vaccines, which, due to government monopsony power, have often been estimated to be extremely cost-effective yet lack any appreciable R&D investments.

3.3 What factors contribute to the low degree of appropriation in health care?

Current research reviewed raises the question of why the share of surplus appropriated by producers is so small and the cost-effectiveness so seemingly high? One may be tempted to argue that there is a lack of market power by those holding patents on these new technologies due to therapeutic competition by other patented products within the same drug class. However, this certainly does not seem to be the case for the breakthrough drugs in HIV/AIDS. Furthermore, Philipson and Jena (2005) showed that even if substitutes do not exist and patents are very broad, causing demand to be highly inelastic, the share of the social surplus allocated to the producer may still be very small, even though inelastic demand raises profits.

3.4 How do altruistic concerns for poorer nations affect optimal appropriation?

The share of US social surplus appropriated by investors sheds an 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 US social surplus above one trillion dollars that has not been appropriated by R&D investors, giving innovators a few billion dollars extra to stimulate innovation (as these public or private contracts provide), seems to pale in comparison to the consumer surplus they are not appropriating. Adding to this the benefits from enhancing consumption in poor countries implies that current estimates of appropriation have an upward bias. Better valuing world social surplus under altruism or externalities is an important area of future research.

3.4 Third party insurance and appropriation

Finally, the peculiar aspects of healthcare's third-party payer markets may raise some nonstandard issues regarding the efficient form of surplus appropriation. One concern is the altruism inherent in subsidy and social insurance programs such as Medicaid (and perhaps Medicare). Another concern is the tradeoff between risk-sharing and incentives (moral hazard) that makes over-consumption of services an issue at the time of service. A third concern is the impact of the joint demand of physicians and patients. In general, the standard analysis of appropriation needs to be better modified to the peculiar aspects of health care.