NBER WORKING PAPER SERIES

ENDOGENOUS COST-EFFECTIVENESS ANALYSIS IN HEALTH CARE TECHNOLOGY ADOPTION

Anupam Jena
Tomas Philipson

Working Paper 15032
http://www.nber.org/papers/w15032

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2009

We thank James Raftery for his generous provision of the NICE data and Kris Hult for excellent research assistance. We are also thankful to seminar participants at the NBER Summer Institute Meetings, The Wharton School, The University of Chicago, and Duke University for their helpful comments. This project was partly supported by the National Institute of General Medical Sciences through Medical Scientist National Research Service Award 5 T32 GM07281 (Jena), and the Agency for Healthcare Research and Quality through UCLA/RAND Training Grant T32 HS 000046 (Jena) and The George Stigler Center at The University of Chicago (Philipson). The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Endogenous Cost-Effectiveness Analysis in Health Care Technology Adoption
Anupam Jena and Tomas Philipson
NBER Working Paper No. 15032
June 2009, Revised October 2010
JEL No. I0,I1,I11,I18

ABSTRACT

Increased health care spending has been argued to be largely due to technological change. Cost-effectiveness analysis is the main tool used by private and public third-party payers to prioritize adoption of the new technologies responsible for this growth. However, such analysis by payers invariably reflects prices set by producers rather than resources used to produce treatments. This implies that the "costs" in cost-effectiveness assessments depend on endogenous markups which are, in turn, influenced by demand factors of patients, doctors, and payers. Reimbursement policy based on endogenous cost-effectiveness levels may therefore bear little relationship to efficient use of scarce medical resources. Using data on technology appraisals in the United Kingdom, we test for conditions under which adoption based on endogenous cost-effectiveness may lead to adoption of more inefficient treatments in terms of resource use.

Anupam Jena
Department of Medicine
Massachusetts General Hospital
55 Fruit Street
Boston, MA 02114
jena.anupam@mgh.harvard.edu

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

1. Introduction

New medical technologies are often argued to be a leading force behind the growth in health care spending.[1] In order to manage the costs imposed by such technologies and to prioritize health care dollars, both public and private payers have increasingly demanded evidence on combined measures of the benefits and costs of new technologies. These measures include cost-effectiveness, cost-utility, and cost-benefit analysis, hereafter referred to collectively as CE analysis.[2] It is self-evident that payers should attempt to maximize the returns in health they obtain from the limited resources available for health spending. CE analysis is the main method used for this purpose.[3] For example, CE thresholds, which dictate that a given technology will be reimbursed only if costs per quality-adjusted life year (QALY) are below a given threshold, is one way in which CE-based adoption is implemented in practice. The most prominent examples are the United Kingdom (UK) National Institute for Clinical Excellence (NICE) and Australia's Pharmaceutical Benefits Advisory Committee.[4] As a consequence of the extensive use of CE analysis by payers, an enormous health economics literature has developed and shown the conditions under which CE analysis, when applied under a fixed budget constraint, can lead to gains in efficiency. Indeed, the amount of work done on the CE of medical technologies may perhaps be the largest field within health economics, particularly in European countries where such analysis guides a large share of public technology adoption and reimbursement.

When CE analysis is used to guide reimbursement in practice, however, the costs incorporated into these assessments are the prices charged to payers by producers or innovators, rather than the societal resource costs used in production. This is almost inevitable as producers in any industry are never eager to share their data on production

[1] See e.g. Newhouse (1992) and recent summary by Chandra and Skinner (2008).
[2] The literature on these methods 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 (2000).
[3] Building on the area-variations literature, others have suggested that cost-effectiveness analysis of 'accountable care organizations' rather than of specific treatments may be efficient (Chandra, Fisher, and Skinner; 2010).
[4] Bethan et al. (2001) report on Australia. While prior to 1993 no European countries formally required pharmacoeconomic assessments of new products (Drummond et al., 1993), most of the 13 European countries evaluated in a later analysis (Drummond , 1999) had or were in the process of developing formal agencies responsible for such assessments.

costs. Therefore, prices – marked up over costs – determine the CE levels observed for new innovations, not the production costs which ordinarily determine the efficient use of resources.

This paper examines how using prices rather than costs impacts the usefulness of CE analysis in guiding health care resource allocation. When prices impact the likelihood of technology adoption and are chosen to maximize expected profits given this uncertainty, observed CE levels will be *endogenous*. They will be determined by firms' incentives to mark up technologies above their production costs, which will in turn depend on how CE analysis is used in reimbursement. For example, if a third-party only pays for technologies that are cheaper than a fixed CE threshold, as done in the UK, manufacturers may find it in their best interest to price up to that threshold regardless of production costs. Treatments may appear equally cost-effective due to the adoption policy of the payer, despite varying greatly in the extent to which patients are willing to pay above production costs. More generally, demand-side factors that drive markups also drive observed CE levels. In fact, because producers face two customers, payers adopting the technologies and patients or doctors using them, we show that the price-sensitivity of both parties jointly determines the markup. In short, observed CE levels will depend on how CE assessments are used in technology adoption. The traditional "bang-for-the-buck" rationale for CE policies may therefore fail because demand factors are included in the "buck". The overall argument of this paper, therefore, is that the rationale for using CE assessments for health care adoption is weakened when those affected by such adoption policies act in their own self-interest.

Section 2 of this paper begins by deriving a specific condition for when CE rankings of treatments based on endogenous measures will deliver different rankings than when based on exogenous production costs – we term this a 'reversal.' The possibility of such reversals is central to understanding whether the use of CE analysis by payers will lead to efficient adoption (or not) of technologies with the largest health impacts and the cheapest production costs. Reversals occur when markups are negatively related with production costs so that treatments with the lowest resource use are also those marked up the most. We characterize what drives markups in a non-standard environment of dual demand by patients/doctors and payers. In particular, we show how public CE-based

reimbursement policies drive markups in conjunction with patient demand, and hence determine the level of endogenous cost-effectiveness chosen by firms seeking reimbursement. We argue that a central issue affecting reversals is whether there is heterogeneity across treatments in the adoption behavior of payers.

Section 3 proposes a simple test whether existing adoption policies may lead to CE reversals based on a simple heterogeneity test. Regressing adoption decisions on observed endogenous CE levels and a set of covariates, in our case disease-class dummies, we argue that a simple F-test indicates the possibility of reversals. Using data on treatment adoption decisions by the National Institute for Health and Clinical Excellence (NICE) in the United Kingdom from 1999 to 2005, we find evidence suggestive of the possibility of such reversals.

Lastly, we conclude in Section 4 by discussing several shortcomings of the paper and some of the future research areas they suggest.

2. Exogenous versus Endogenous Cost-Effectiveness

This section derives the relationship between exogenous cost-effectiveness, which depends on exogenous resource costs of production, and endogenous cost-effectiveness, which relies on prices faced by payers.

2.1 Basic Framework

We interpret cost-effectiveness determination in a standard quantity-quality framework of a monopolist. Consider a single treatment that provides an exogenous, homogenous incremental benefit in health q over a baseline treatment. The treatment is assumed to be produced by a monopolist who charges an incremental price p. The health benefit q may be the incremental extension in quality-adjusted life years due to treatment (as perhaps revealed by data from clinical trials) and can generally be interpreted in standard economic formulations as the quality of the product. Compared to a baseline treatment, we assume there is a constant marginal cost $c(q)$ of producing a treatment of a given quality level.

In this framework, we define the *exogenous* incremental cost-effectiveness ratio (ICER) to be the cost per unit of quality as in:

$$CE = \frac{c}{q} \tag{1}$$

The numerator is the exogenously determined incremental resource cost to society per person utilizing treatment, and the denominator is the incremental health benefit among those utilizing treatment. In the Appendix, we discuss the conditions under which this is consistent with traditional static efficiency criteria.[5]

The *endogenous* cost-effectiveness ratio uses the price faced by payers as the relevant cost, rather than the cost of resources utilized for production, and is given by:

$$CE_N = \frac{p}{q} \tag{2}$$

If m denotes the markup above costs, it is defined as $p = m \cdot c$. It follows immediately that the two forms of cost-effectiveness are related by:

$$CE_N = m \cdot CE \tag{3}$$

This implies that resource allocation decisions based on endogenous cost-effectiveness may be inefficient even when exogenous cost-effectiveness analysis would deliver efficient resource allocation. In particular, treatment adoption based on endogenous CE may lead to the selection of less cost-effective treatments in terms of resource costs. We discuss conditions under which the two norms of cost-effectiveness involve "reversals" in the sense that

$$CE \leq CE' \qquad \& \qquad CE_N > CE'_N \tag{4}$$

Such CE-reversals, or equivalently in the other direction, amount to changes in the ranking of CE levels from best to worst. These reversals occur when the offsetting markup differences are larger than the exogenous cost-effectiveness differences;

$$\frac{m}{m'} > \frac{CE'}{CE} \tag{5}$$

[5] Jena and Philipson (2009) discuss the inefficiencies implied by using a static efficiency criterion such as cost-effectiveness to guide technology adoption. Static criteria measure consumer benefits relative to costs and take innovation as given. To maximize dynamic efficiency, it may be desirable to lower cost-effectiveness levels that raise producer surplus and thus incentives to innovate.

This takes place when markups are negatively related to production costs so that low-cost treatments are marked up relatively more. For example, compared to medical devices, drugs may have smaller marginal costs of production yet face larger markups due to inelastic demand or patent protection.

2.2 Expected profit maximization and cost-effectiveness levels

As the markup of prices over costs is the key determinant of the concordance between endogenous and exogenous CE, it is important to understand what drives markups. When there are third-party payers, such as private- or public insurers, markups are non-standard as producers face two demand sides; the payer adopting the treatment and the patients or doctors using it. The price-sensitivity of both sets of customers will affect pricing.

More precisely, if the demand price (patient cost-sharing) is given by *s(p)* for a given price, then the quantity demanded *D(s(p),q)* falls in the demand price and rises with quality. If the technology is adopted by the payer, the producer collects the variable profit induced by the pricing, but if it is not adopted he earns no profits. Let the probability of adoption be denoted *A(p,q)* given the price and quality. We assume it is differentiable and decreasing function of price but an increasing function of quality. A special case would be when the chance of coverage is decreasing in the endogenous cost-effectiveness ratio of the technology, i.e. *A(p,q)≡ A(p/q)*. Thus, both forms of demand depend negatively on price ("cost") and positively on quality ("effectiveness").[6]

Given the two components of demand, the expected profits are the post-approval profits discounted by the probability of treatment adoption:

$$E[\Pi] = \max_{p} A(p,q) \cdot D(s(p),q)[p - c(q)] = A(p) \cdot \pi(p) \qquad (6)$$

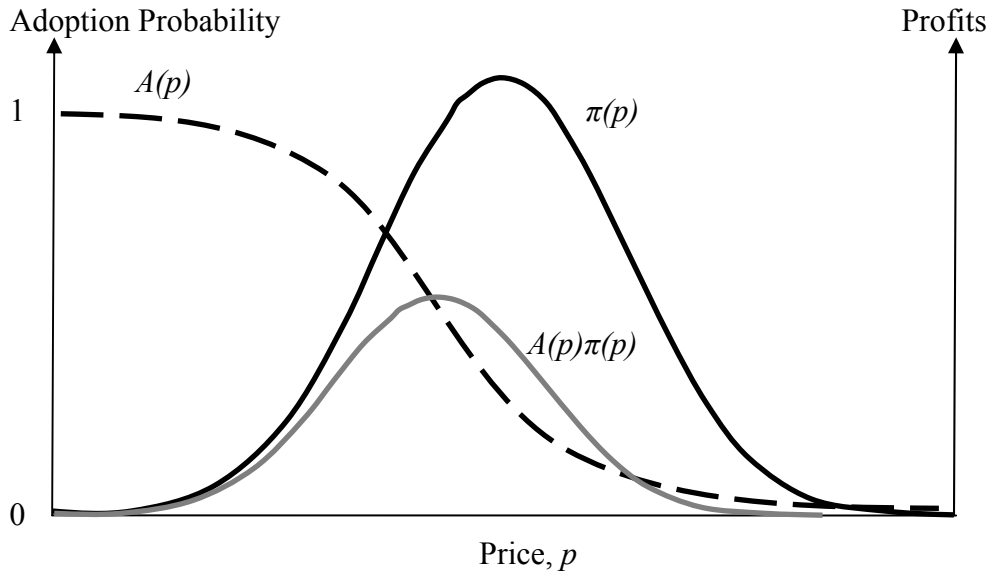
Throughout, we treat quality as exogenous and previously determined before the pricing decision and often suppress it to limit notation. We assume that the probability of acceptance and ex-post profits are well-behaved in the sense that A(p) is differentiable

[6] For example, when the binary demand of the payer is determined by an unknown reservation price R ("CE-threshold") distributed according to the cdf denoted G(r;q), the acceptance function is given by A(p,q)=Pr(r>p)=1-G(p;q).

and strictly decreasing and $\pi(p)$ is differentiable, concave, and has a unique profit-maximizing price.

The probability of technology adoption, the variable profits conditional on reimbursement, and the expected profits are illustrated in the figure below.

FIGURE 1—Probability of treatment adoption, variable and expected profits



These expected profits imply that in raising price, the producer must take into account two types of buyers—the third-party payer making the treatment adoption decision and the patients or providers using the product once adopted. The optimal price balances the profit impacts of the two demand sides and interior price satisfies the necessary first-order condition:

$$A'\pi + A\pi' = 0 \qquad (7)$$

The gain in profits conditional on adoption at a higher price must be balanced with the larger chance of not being adopted. Throughout we assume there is a unique optimal price.[7]

This optimal balance between adoption and ex-post profits has two direct implications. First, producers will not set price low enough to guarantee acceptance since

[7] A sufficient condition for a unique optimum is that A'(0)=0 and that -A'/A is increasing in price. Then there is a unique price at which the increasing function –A'/A crosses the function π'/π which is decreasing in the relevant range of prices below the optimal ex-post price (at which it is zero).

the probability of rejection, *1-A*, is strictly positive at the optimal positive price. Producers take the risk of rejection in exchange for the larger profits obtained when the submitted treatment is adopted. Second, the price that maximizes expected profits ($A\pi$) is lower than the price that maximizes standard profits conditional on reimbursement (π). This follows from the fact that the first-order condition can be restated as $\pi' = (-A'/A)\pi > 0$ so that the ex-ante optimal price is on the upward sloping part of the inverted U-shaped profit function π in Figure 1. Intuitively, pricing above the ex-post optimal price lowers both ex-post profits and the chance of approval and so can never be optimal. Producers do not necessarily raise price enough to maximize ex-post profits for fear of not getting the technology approved at such a high price.

Rearranging the necessary first order condition, it can be written as a modified Lerner condition for the optimal percentage markup as in:[8]

$$(p-c)/p = 1/[\varepsilon_D + \varepsilon_A] \qquad (8)$$

where $\varepsilon_A = -A'p/A$ and $\varepsilon_D = -D_p p/D$ are the absolute values of the price-elasticities of the two forms of demand. This generalized markup condition has several implications. First, the markup falls with the price-sensitivity of both forms of demand. Under one extreme, if there is no impact of raising price on approval then $\varepsilon_A = 0$ and the markup satisfies the standard Lerner condition of ex-post profits. Hence, the price sensitivity of the approval decision lowers markups below the standard level. On the other extreme, if there is no impact of price on ex-post demand, as may be the case when demand prices or co-pays do not vary the supply price, then $\varepsilon_D = 0$ and the price is governed by adoption behavior alone. For example, public payers may differentially adopt treatments based on the severity or prevalence of a disease. Alternatively, certain diseases may have more politically influential interest groups (e.g. HIV/AIDS or breast cancer in the US) and therefore face easier approval.

A second implication of this generalized Lerner condition is that pricing may occur where ex-post demand by patients/doctors is inelastic. That would never occur under standard profit-maximization, since raising price would raise ex-post profits by

[8] The FOC may written as $\varepsilon_A = (\pi'*p)/\pi$ and $(\pi'*p)/\pi = ([D_p(p-c) + D]*p)/(D*(p-c)) = -\varepsilon_D + p/(p-c)$.

increasing revenue and reducing total costs. However, such a rise in ex-post profits on the margin may be optimal when it traded off against the lower chance of reimbursement.

The final and most important implication of the Lerner condition is that if both demand sides affect markups, differences between observed endogenous cost-effectiveness rankings and unobserved exogenous rankings may occur. In particular, the two cost-effectiveness levels are related according to:

$$CE_N=CE/[1-1/(\varepsilon_D + \varepsilon_A)] \quad (9)$$

As a consequence, when demand elasticities counteract cost differences, selecting treatments based on endogenous cost-effectiveness will lead to adoption behavior that selects treatments that do not deliver the largest health benefits for a given amount of resources used to produce them.

Section 3: Causes of reversals in cost-effectiveness

In this section, we discuss factors that lead to the selection of inefficient treatments when adoption is based on endogenous cost-effectiveness levels rather than exogenous ones. Given that demand behavior reflects the difference between the two forms of CE as indicated in (9), of central importance is demand heterogeneity across treatments with similar exogenous cost-effectiveness levels.

Generally, for a given demand or supply factor represented by the scalar z, the price for a given quality level may depend on both demand and supply conditions as in:[9]

$$p(q,z) = m(q,z)c(q,z) \quad (10)$$

If the two cost-effectiveness measures are denoted by $CE(q,z) = c(q,z)/q$ and $CE_N(q,z) = p(q,z)/q$, then a sufficient condition for reversals to occur is that an increase in the factor has opposing effects on CE measures:

$$(\frac{dCE}{dz})(\frac{dCE_N}{dz}) < 0 \quad (11)$$

This condition can be rewritten as:

$$[\mu_m + \mu_c][\mu_c] < 0 \quad (12)$$

[9] For example, under expected profit maximization the parameters (z_1,z_2,z_3) may reflect acceptance-, patient demand-, and cost parameters in the expected profits $V=A(p;z_1)[D(p,q;z_2)(p-c(q;z_3))]$.

where $\mu_m = m_z*z/m$ and $\mu_c = c_z*z/c$ are the markup and cost elasticities with respect to the factor z. If the factor raises costs, $\mu_c > 0$, then it raises exogenous cost effectiveness CE. But if the factor lowers markups more than it raises costs, $\mu_m + \mu_c < 0$, then the factor lowers endogenous cost-effectiveness CE_N as it is the product of the two.

3.1 Reversals due to differential demand behavior

If two treatments have the same production costs and quality they have the same exogenous cost-effectiveness, CE. However, if the two treatments differ in either forms of demand, i.e. approval by a third-party or ex-post demand by patients and providers, then they may have different levels of endogenous cost-effectiveness.

3.1.1 Reversals due to differential patient and doctor demand behavior

First consider the extreme case when all treatments are adopted by the payer, *A=1*, regardless of their cost-effectiveness. This is arguably the case in the US where FDA approval of a technology is sufficient for coverage by the public payers Medicare and Medicaid. In this setting, expected profits $A\pi$ reduce to post-approval profits and optimal pricing satisfies the standard condition; $\pi'=0$. In this standard case of markup determination, the elasticity of patient or doctor demand yields the markup:

$$m = 1/(1-1/\varepsilon_D) \qquad (13)$$

Therefore, reversals may occur when the lower costs treatments are also the most inelastically demanded by patients or doctors. In particular, a negative relationship between markups and costs of production may occur when low cost treatments are produced in less competitive markets or when low cost treatments such as pharmaceuticals are more marked up than devices or surgery.

3.1.2 Reversals due to differential adoption behavior

As an illustration, consider when public payers follow a *reservation price policy*, adopting only treatments priced below the reservation price. In the UK, this is often described as a "CE-threshold" policy in which only technologies whose CE levels are below a given threshold *T* are adopted; $A(p/q) = 0$ if $p/q > T$ and $A(p/q) = 1$ if $p/q \leq T$. Furthermore, suppose cost-sharing has no variable component, as might be true if there is

a fixed payment for filling a given prescription, $s(p) = s$. In this case, as both demand elasticities are zero below the threshold ($\varepsilon_D = \varepsilon_A = 0$), optimal pricing would lead to the endogenous CE set to the threshold. This would, in turn, induce markups that are inversely related to the exogenous CE levels:

$$CE_N = T \rightarrow m \cdot CE = T \rightarrow m = T/CE$$

The adoption rule therefore induces markups that are negatively related to resource costs. Because of this negative relationship, changes in real resource use – as reflected by exogenous CE levels – would have no impact on endogenous levels of CE used for payment purposes.[10]

3.1.3 Reversals due to differential adoption and production costs

If two treatments differ only in costs but are of equal quality then there is a reversal whenever price is lower for the more expensive treatment[11]

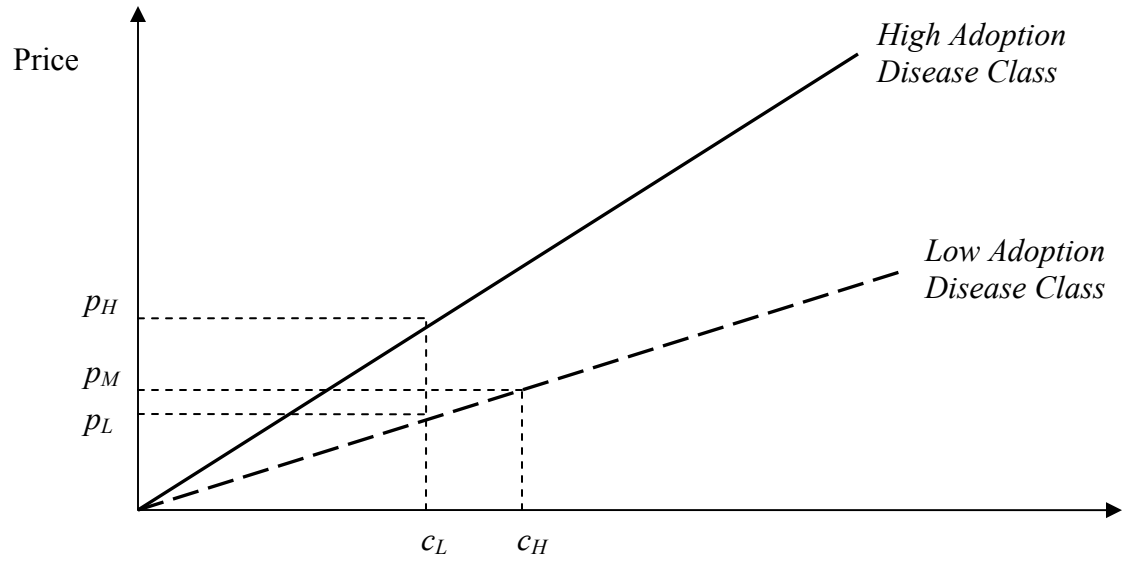
$$c < c' \ \& \ p > p'$$

In this case, the factor z shifts both the cost c(q;z) but also the markup m(q;z) because markups cannot be constant (constant elasticity) when A is bounded. The effect on cost differences is illustrated in Figure 3. Among treatments of the same quality, this figure shows two curves that map out the optimal price of a treatment as an increasing function of the treatment's exogenous CE level where the slope is affected by the endogenous markup.

FIGURE 3—Reversals due to adoption and cost-differences across treatments

[10] This is, of course, true as long as exogenous CE levels are not higher than the threshold, in which case the technology would presumably not be presented to the payer for adoption in the first place.

[11] It may appear that the two forms of cost-effectiveness yields the same ranking of treatments differing in costs when both forms of demand are of constant elasticity. However, the bounded function A cannot be of constant elasticity throughout its support.



Costs and Exogenous CE

Holding quality constant, the *x*-axis corresponds both to different cost levels and exogenous CE levels. The top line traces out the profit-maximizing price charged by a firm whose product treats a disease that has a high probability of public adoption at any given level of submitted cost-effectiveness. Similarly, the lower line characterizes the optimal price for treatments in disease classes that are less favorably adopted. Now, consider two treatments of differing costs of production, $c_L < c_H$. For a given cost level, it is clear that the price will be higher in the higher approval class; that is, $p_L < p_H$ for the same cost level c_L. This directly implies a change in rankings from treatments having the same exogenous CE to having strictly different endogenous CE levels. In addition, full reversals may occur when politically motivated adoption behavior is negatively related to resource costs. For example, suppose a treatment is more expensive to produce and is in the low acceptance disease class; the profit-maximizing price is p_M. If the lower-cost treatment is in the high acceptance class, then its profit-maximizing price is p_H which is higher than p_M. In this case, a full reversal will occur when the lower-cost treatment – in a resource sense – is in a class that is less politically favored by the public payer.

3.2 Reversals due to differential quality

When treatments vary in quality, reversals occur when quality raises one form of cost-effectiveness while lowering the other:

$$(\frac{dCE}{dq})(\frac{dCE_N}{dq}) < 0 \qquad (14)$$

This condition can be rewritten as

$$[\mu_m + \mu_c - 1][\mu_c - 1] < 0 \qquad (15)$$

where μ_m and μ_c are the elasticities of markups and costs with respect to quality. If costs alone do not rise as fast as quality, $\mu_c<1$, then exogenous cost-effectiveness CE falls with quality. But if markups and costs together rise as fast as quality, $\mu_m + \mu_c > 1$, then endogenous cost-effectiveness CE_N rises with quality. Thus, markups and costs may both rise in the same direction with quality, yet still lead to reversals in endogenous and exogenous cost-effectiveness. Similarly, costs may rise faster than quality when quality rises, $\mu_c >1$, while markups fall in quality enough so that prices do not rise as fast as quality, $\mu_m + \mu_c < 1$.

3.3 Generalizing to other payer and demand environments

Our analysis can be generalized to broader objective functions of producers in environments with many different payers and patient/doctor demand structures. In reality and more generally, treatment adoption decisions involve multiple payers, disease indications, patient groups, and doctor groups that are linked to each other by price.

Suppose profits to producers are represented by the general function V(p,q;z) where z is a set of parameters that describes the payer and demand environment. When there is a single payer, the expected profits discussed earlier may be written as:

$$V(p,q,z) = A(p;z_1)[D(p,q;z_2)(p-c(q;z_3))] \qquad (15)$$

Now suppose that there are multiple payers involved, denoted by k = 1, 2,.., K, each with different approval behavior and possibly ex-post profits to the producer. In this case, the general profit function may be written:

$$V(p,q;z) = F[A_1(p;z),\pi_1(p;z),\ldots .,A_K(p;z),\pi_K(p;z)] \qquad (16)$$

For a general profit function V, reversals driven by demand conditions will occur as long as the endogenously chosen price is affected by demand conditions, holding costs of production constant. Because this is nearly always true, reversals will be likely under

those general conditions as well. More precisely, for the general profit-function the necessary FOC determining the optimal price satisfies:

$$\frac{dV}{dp}(p,q;z) = 0 \qquad (17)$$

If z is a scalar demand parameter, then the implicit function theorem directly implies it will affect pricing whenever:

$$\frac{dp}{dz} = \frac{-d^2V/dpdz}{d^2V/dp} \neq 0 \qquad (18)$$

If the SOC is satisfied then the denominator is non-zero and *dp/dz* will be non-zero whenever $d^2V/dpdz$ is non-zero. In other words, if the FOC is shifted by the demand parameter z, so that $d^2V/dpdz$ is non-zero, then *dp/dz* will be non-zero as well. This implies that in order to check for the potential possibility of cost-effectiveness reversals, one need only check for heterogeneity in a demand parameter that has a non-zero cross partial for the profit function V. For example, when there is a single payer which induces the profit function (15) and heterogeneity in factors affecting the probability of adoption (e.g. z_1 reflecting preference towards treatments for diseases with larger influence on the political process), reversals will occur because the partial $dV/dpdz = (A_{pz1})(D^*(p-c)) + (A_{z1})[D_p^*(p-c) + D)]$ is generally non-zero.

4. Testing for heterogeneity in adoption behavior: an analysis of NICE

Heterogeneity in adoption policies across types of treatments (e.g. pharmaceuticals vs. devices vs. procedures) or disease classes is a sufficient condition for reversals to occur. Using data on treatment adoption decisions by the National Institute for Health and Clinical Excellence (NICE) in the United Kingdom, this section discusses tests for such heterogeneity in adoption behavior. Co-pays in the UK's National Health Service (NHS) do not depend on producer prices so that the relationship between the two forms of cost effectiveness becomes is driven by approval behavior $CE_N = CE/[1-1/\varepsilon_A]$. Thus, our tests for approval heterogeneity may be interpreted as sufficient conditions for reversals in the context of NICE.

3.1 Background on NICE

Introduced in 1999 as a special health authority for England and Wales, the initial purview of NICE was to make recommendations to the British National Health Service (NHS) on the coverage of selected new and existing medical technologies and to develop clinical guidelines (Buxton, 2001). Although other countries have developed similar organizations, NICE was the first national agency with the power to guide technology adoption for all new health technologies including pharmaceuticals, procedures, and devices (Schulper et al., 2001). While NHS authorities were initially mandated to take into account but not necessarily follow NICE's advice, in 2002 they became legally obligated to fund treatments recommended by NICE. The initial spectrum of assessments by NICE included pharmaceuticals, medical devices, procedures, diagnostic and screening technologies, and health promotion programs, but most referrals to date have concerned either pharmaceuticals or devices.

Following the selection of technologies to be assessed, NICE commissions or accepts reports from several sources, including the supply side (manufacturers), the demand side (professional and patients' groups), and independent sources (academics). The evidence typically gathered for a given technology includes its clinical effectiveness, cost per quality-adjusted-life-year (QALY) gained, and impact on costs borne by the NHS (Raftery, 2001). After gathering this information, NICE first issues a provisional appraisal, which is reviewed by the parties involved, followed by a final appraisal to the NHS. According to guidelines set forth by the Secretary of State for Health, the final guidance rendered by NICE should account for the clinical priorities of the NHS, the need of patients under consideration, the cost-effectiveness of the treatment, and the strength of clinical evidence and cost-effectiveness estimates (Buxton, 2001).

The final guidance issued by NICE summarizes whether a treatment is recommended to the NHS and the reasoning behind the decision. The appraisal committee makes one of four recommendations: the technology can be recommended with no restrictions, recommended with minor restrictions, recommended with major restrictions, or not recommended. If a manufacturer is unsatisfied with the recommendation, it can appeal the decision.

3.2 Data on technology approvals by NICE

Since its inception in 1999, NICE has published 141 guidances. Our data analyzes the 86 guidances submitted to NICE between the years 1999 and 2005—the dates of guidance publication range from 2001 to 2007.[12] We define a particular treatment as each combination of a drug or technology and the disease it addresses. Since the same drug or technology may be used to treat multiple diseases or the drug or technology may have different parts that must be recommended separately, a single guidance may contain multiple treatments. Our database, therefore, has 145 treatments in the 86 guidances we examine, and the unit of observation is a treatment. Table 1 provides descriptive statistics on these guidances in terms of endogenous CE levels (p/q) as well as acceptance behavior (A).

TABLE 1—Descriptive statistics of NICE guidance data

Total no. of guidances	145		
Treatments recommended by NICE	Percent		
Yes	30		
Yes, with minor restrictions	32		
Yes, with major restrictions	22		
No	16		
No. of guidances with published CE	76		
Endogenous CE (Cost per *QALY* (£)) by range of estimate	No. of treatments	Mean	Std. Dev.
Low estimate	35	12,297	11,704
High estimate	37	43,673	35,701
Mean estimate	51	28,132	18,798
Endogenous CE (Cost per *LYG* (£)) by range of estimate			
Low estimate	20	8,276	6,304
High estimate	22	19,506	13,744
Mean estimate	26	17,397	11,404
Avg. of est. mean cost per *QALY* or *LYG* (£)	76	24,710	17,380
Range of est. mean cost per *QALY* or *LYG*	Percent		
Less than £10,000	22		
Between £10,000 and £20,000	25		
Between £20,000 and £30,000	18		

[12] We are thankful to James Raftery for providing us with his detailed collection of these guidances. We do not have data on guidances submitted after 2005, which compromise the remainder of the 141 guidances noted above.

Between £30,000 and £40,000	16
More than £40,000	18

Source: NICE published treatment guidances, 1999 – 2005.

Of the 145 NICE guidances present in our data, 23 (16%) were not recommended by NICE, 32 (22%) were recommended with major restrictions, 46 (32%) were recommended with minor restrictions, and 44 (30%) were recommended with no restrictions. A "no" recommendation is given for either poor cost-effectiveness or insufficient evidence to warrant the use of the treatment. While treatments with major restrictions are still recommended by NICE, such treatments are only recommended for either second-line use by those refractory to alternative treatments or by targeted subgroups with severe disease. Recommendations with minor restrictions limit use in one of several ways; e.g. recommendations may require the particular treatment to be the least costly option, may require specialist supervision, or may require treatment monitoring. The treatments that are recommended as "yes" without any restrictions can be used routinely and as the primary treatment for a disease. Overall, 84% of treatments included in our data were recommended with no, minor, or major restrictions.

For those guidances for which cost-effectiveness is reported[13], NICE measures cost-effectiveness in two ways, cost per quality-adjusted life year (QALY) gained and cost per life year (LY) gained, both measured relative to some baseline treatment. Quality-adjusted life years differ from life years gained by incorporating both quality and quantity of life into measures of a treatment's effectiveness. Cost-effectiveness ratios are calculated in the usual manner.[14]

Because measuring effectiveness precisely can be difficult, NICE guidances often report high, mean, and low estimates of cost per quality-adjusted life year or standard life year gained for each treatment. For those treatments for which high and low estimates exist, Table 1 presents the average cost per QALY or LY gained within each range. The

[13] As shown in Table 1, NICE does not always report cost-effectiveness for each treatment so estimates only exist for roughly half (76/145) of the observations in our data.

[14] For example, if a new drug costs £15,000 and the existing treatment costs £5,000, the numerator in the cost per QALY (or LY) gained measurement is £10,000. If the new treatment adds 0.9 QALYs and the previous treatment added 0.4 QALYs, the denominator is 0.5 QALYs. Therefore, the cost per QALY is £10,000/0.5 = £20,000.

within-treatment variability is substantial—the estimates of average cost per QALY or LY gained vary from roughly £12,000 (low-estimate group) to £44,000 (high estimate group) in our data.

In order to have a unified cost-effectiveness measure for our subsequent analysis, we do not distinguish between QALYs and standard LYs and assume that the cost per QALY or LY gained takes on either the mean cost per QALY gained or mean cost per LY gained, depending on which variable exists for a given treatment. Under this measure, the mean cost per QALY or LY gained is approximately £24,710 with a standard deviation of £17,380.

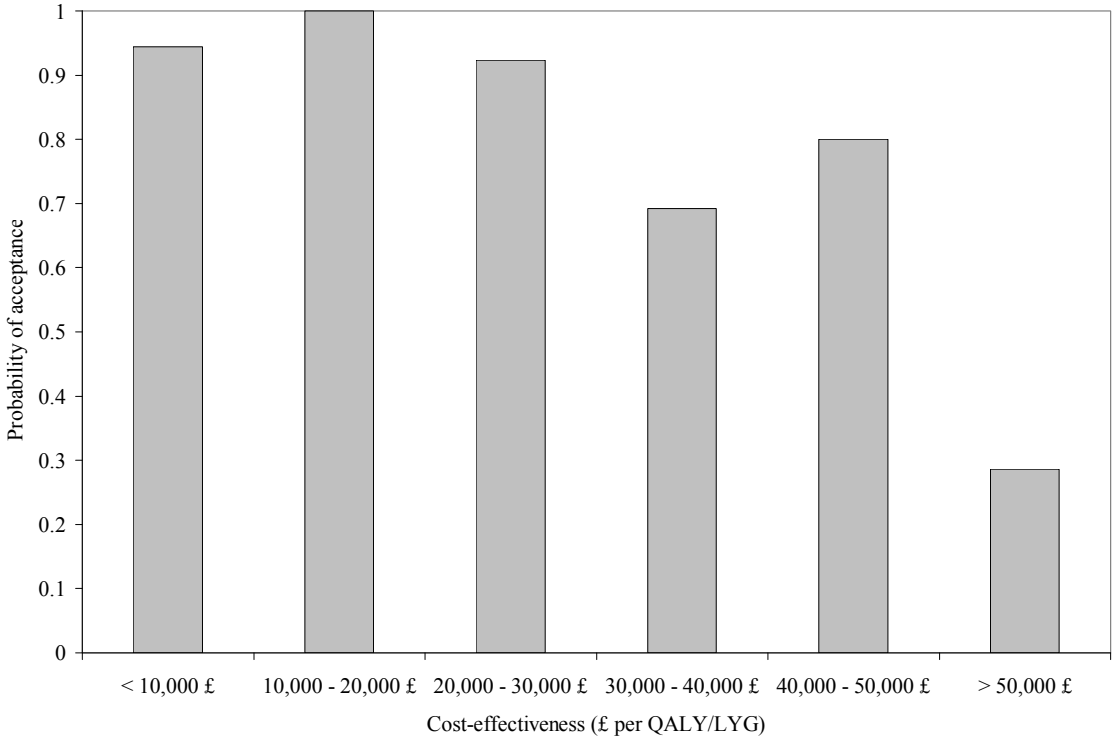
3.3 Heterogeneity and endogeneity bias

Our theoretical analysis implies that heterogeneity in treatment adoption across either treatment types or disease classes is central to reversals. However, estimating the impact of observed cost-effectiveness on the probability of adoption is hampered by the fact that adoption and observed cost-effectiveness levels are endogenously determined.. To illustrate the implications of this endogeneity bias, consider two treatments; the first is adopted with certainty (A=1) and the second is adopted with some probability that depends negatively on the submitted cost-effectiveness level (A'<0). This would lead to an ex-post optimal monopoly price for the first treatment but a lower price for the second one. Acceptance would therefore be high for the first treatment with a high endogenous CE level but low for the second treatment with a low endogenous CE. Heterogeneity may therefore imply an unconditional positive relationship between adoption and endogenous CE even though the relationship is negative conditionally. More precisely, for two treatments with different adoption functions A_1 and A_2 that induce prices $p(A_1) > p(A_2)$, it may well be true that $A_1(p(A_1)) < A_2(p(A_2))$. A sufficient "diagnostic" test for the presence of heterogeneity in the case of NICE would therefore be an increasing relationship between adoption and endogenous CE levels as it would never be observed if the adoption-function $A(.)$ was homogeneous and decreasing in endogenous CE levels

across all treatments. In this homogeneous case, the unconditional relationship between approval and endogenous CE would depict the homogenous function A.

Table 1 discussed above provides information about this unconditional relationship in the data. We group the mean cost per QALY or LY gained into five categories: less than £10,000, between £10,000 and £20,000, between £20,000 and £30,000, between £30,000 and £40,000, and more than £40,000. The proportion of treatments within each range is fairly similar, with approximately 34% of treatments having cost-effectiveness levels above commonly reported thresholds of NICE adoption (~£30,000). Using these data, Figure 4 plots the unconditional relationship between a treatment's endogenous CE and the probability of NICE recommendation—the reduced form of the acceptance function *A(p/q)* in our analysis. The figure illustrates our diagnostic test of a non-monotonic relationship between endogenous CE and acceptance and is suggestive that heterogeneity in acceptance across treatments may be present.

FIGURE 4—Endogenous cost-effectiveness and the probability of treatment acceptance, NICE 1999-2005



Source: NICE published treatment guidances, 1999 – 2005.

In addition, Figure 4 is also consistent with the discussed prediction that under uncertain adoption, optimal pricing will result in a strictly positive fraction of treatments being rejected as firms trade off higher ex-post profits due to higher prices with the increased probability of rejection that these higher prices induce.

3.4 Heterogeneity and reversals: NICE adoption behavior by disease class

Our analysis implies that differences in treatment adoption behavior across treatment types (pharmaceuticals vs devices vs procedures) or disease classes will induce reversals in CE rankings. We test for heterogeneity in adoption behavior across disease classes; Table 2 specifies how the probability of acceptance of treatments by NICE has varied by disease class and endogenous cost-effectiveness levels.[15]

TABLE 2—Number of treatments submitted and accepted by disease class and endogenous cost-effectiveness, NICE 1999-2005

	Endogenous Cost-effectiveness (1,000£/QALY)				
Disease Class	< 20	20 - 40	>40	Total	% accepted
Arthritis	5/5	2/2	0/1	8	88
Cancer	14/14	8/9	2/3	26	92
Heart	7/7	4/4	0/0	11	100
Infectious	2/2	2/5	2/5	12	50
Mental	4/5	1/2	0/1	8	63
Prevention	2/2	2/2	0/0	4	100
Other	3/3	2/2	2/2	7	100
Total	38	26	12	76	84
% accepted	97	81	50	84	

Source: NICE published treatment guidances, 1999 – 2005. Each cell reports the number of accepted treatments/submitted treatments for a given disease class and endogenous cost-effectiveness range.

Table 2 first illustrates that the NICE guidances in our data span a relatively large group

[15] Of course, an ideal empirical test of reversals would be to test how well exogenous and endogenous CE measures align by ranking treatments according to their exogenous and endogenous CE levels. This is infeasible as markups are unobservable, a standard and central empirical problem in industrial organization more generally.

of diseases and categories of treatment. Of the 76 guidances considered, 26 (34%) were for cancer, 11 (14%) were for heart disease, 12 (16%) were for infectious disease, 8 (11%) were for mental health, and so on. Next, for a given range of cost-effectiveness (columns), each row of Table 2 displays both the number of treatments accepted by and submitted to NICE for a given disease. For example, out of 14 treatments for cancer with submitted CE levels below 20,000 £/QALY, 14 were accepted by NICE with minor, major, or no restrictions. For cancer treatments with submitted CE levels in the range of 20,000 – 40,000 £/QALY, 8 out of 9 treatments were adopted by NICE, while in the range of 40,000 £/QALY and above, 2 of 3 submitted treatments were accepted. Importantly, however, in the same ranges of 20,000 – 40,000 £/QALY and greater than 40,000 £/QALY, respectively, 2 out of 5 treatments for infectious disease were accepted. In the same respective ranges, 1 out of 2 and 0 out of 1 submitted treatments for mental health were accepted. This table suggests that differential adoption behavior by NICE towards specific diseases may exist. This should, of course, be qualified by the power issues that are present—the data at hand are clearly limited by the number of guidances issued to date and the broad range of diseases covered.

The data in Table 2 suggest a general methodology to test for the potential of CE reversals, namely by testing whether the probability of treatment acceptance depends not only on submitted cost-effectiveness, but on the disease class of the treatment as well. Table 3 specifies such a test and reports the coefficients of a linear probability model of the impact of cost-effectiveness and disease class on the probability of treatment acceptance by NICE. The linear probability model was selected due to well-known problems with logit or probit specifications in fitting the full acceptance levels displayed in the descriptive table. While sample size considerations prohibit a fully interacted model of the differential impact of disease class on the CE-adoption relationship, our model employs indicators for disease classes to determine how and whether specific disease classes affects the probability of adoption, conditional on endogenous CE levels. The excluded disease class was the smallest class, diabetes. A necessary implication of a homogeneous acceptance function A across disease classes is that the dummies have no effect and that variation in adoption is fully explained by endogenous CE alone.

.

TABLE 3—Impact of endogenous cost-effectiveness and disease class on probability of treatment acceptance

Probability of treatment acceptance	
Mean cost-effectiveness (1,000£/QALY)	-0.009*
	(0.002)
Cancer	-0.034
	(0.098)
Heart	-0.031
	(0.122)
Infectious	-0.322*
	(0.120)
Mental health	-0.310*
	(0.132)
Prevention	-0.008
	(0.171)
Constant	1.154
	(0.096)
R^2	0.38
F-test of equality of disease indicators	$p = 0.03$

Source: NICE published treatment guidances, 1999 – 2005. Table presents coefficients of a linear probability model of the impact of cost-effectiveness and disease class (excluded class: diabetes) on the probability of treatment adoption by NICE. Standard errors are in parentheses. * Significant at $p < 0.05$.

Table 3 demonstrates a statistically significant negative relationship between endogenous cost-effectiveness and the probability of treatment acceptance; the probability of acceptance declines by an estimated 0.009 (s.e. 0.002) for every 1,000 £/QALY increase in the submitted CE level. In addition, compared to the excluded class of diabetes, each of the diseases presented in Table 3 has a lower estimated probability of acceptance, with infectious disease and mental health being the only diseases with statistically significant effects (-0.322 (s.e. 0.120) and -0.310 (s.e. 0.132), respectively). This suggests the possibility of heterogeneity in treatment acceptance across disease classes, holding submitted CE constant.

The existence of the possibility of reversals therefore boils down to testing

whether there are class differences in adoption behavior, which can be succinctly summarized by an F-test of whether the coefficients on the disease-dummies are the same. An *F*-test of the equality of disease dummies, in fact, rejects the null hypothesis that adoption behavior is identical across disease classes ($p = 0.03$).

4. Conclusion

This paper examined the efficiency implications of CE-based technology adoption in the presence of optimal pricing by firms. Such pricing implies that observed cost-effectiveness levels are endogenous to the CE criteria used to guide payer adoption decisions. Our main finding was that endogenous cost-effectiveness may not be a good guide to resource allocation and may not relate in a systematic way to exogenous measures that reflect true resource costs. We showed that this may be important for threshold adoption policies commonly employed by NICE in the UK. This occurs because both patient/doctor demand factors and adoption policies determine prices but not resource costs. The end result is that the intended value of using CE analysis, to economize on resource costs used to deliver health care, may not be fulfilled.

Our analysis has several important limitations that future research may successfully address. First, one major identification issue facing any analyst is that actual production costs are unobservable to both econometricians and reimbursement authorities. Endogenous CE rankings are observable while exogenous CE rankings are not. The fact that markups are unobservable is, of course, well-known and long-recognized in empirical industrial organization. This issue led us to state our results as sufficient conditions for reversals across a distribution of treatments rather than as documented reversals in the sample at hand. More work is needed to derive results that apply to a given sample of observed quality-, price- and demand data, such as those reflected in the NICE data analyzed here.

Second, we did not consider the possibility of endogenous *effectiveness* or quality induced by technology adoption criteria. This would be important when pricing affects effectiveness through demand. For example, in the case of vaccines, lower prices lead to greater vaccination and socially beneficial "herd immunity", thereby raising effectiveness. Similar issues may arise for other links between demand and

effectiveness, for instance "learning by doing" in the adoption of new technologies (Chandra and Staiger, 2007). Reductions in price for a device used in surgery may lead to increased utilization, greater learning by doing, and ultimately increased effectiveness. The full endogeneity of both prices and effectiveness deserves further analysis in order to better understand the efficiency implications of cost-effectiveness based reimbursement. Adaptive cost-effectiveness adoption in which future prices are not restricted by initial launch prices may be an efficient method of dealing with both endogenous costs and effectiveness.

Third, our analysis did not consider the implications for comparative effectiveness of multiple competing treatments. Such an analysis would consider the duopoly and oligopoly pricing implications of making reimbursement decisions contingent on multiple industry prices and quality levels, as opposed to the single price and quality of a monopolist. When setting prices, producers presumably take into account how reimbursement authorities use CE levels of competing treatments, e.g. through branded or generic reference pricing, for similar conditions. The industrial organization of endogenous cost-effectiveness analysis, and its impact on health care spending, is an important area of future research.

Fourth, we did not analyze how transparency of public decision-making affects cost-effectiveness reversals. In our analysis of the NICE data, endogenous CE levels do not perfectly predict adoption decisions in the sense of goodness-of-fit. This suggests that other unspecified political considerations affect adoption. Making such criteria explicit would lead to increased efficiency if producers did not waste development and application costs on rejected treatments. This efficiency role of transparency needs to be better understood and can be assessed by the goodness-of-fit of the transparent criteria used in explaining payer adoption decisions.

Lastly, the impact of endogenous innovation incentives needs to be better understood. The work of Lakdawalla and Sood (2006, 2008) analyze the implications for innovation of the two-part pricing induced by insurance reimbursement with co-pays. The implications of payer reimbursement policies, such as those based on CEA, on appropriation by innovators is an important area of future research.

Despite these possible further implications of our analysis, however, we believe the overall concern that we raise deserves serious consideration in evaluating the gains from using CE analysis for efficient resource allocation. The overall point, that adoption policies may not have their intended efficiency effects when those affected by them act in their own interest, should feature more prominently into the design of future reimbursement systems and enable more appropriate growth in health care spending.

References

Buxton, M. Implications of the Appraisal Function of the National Institute for Clinical Excellence (NICE). *Value in Health* 4 (2001): 212-16.

Chandra, A. and Staiger, D. "Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks," *Journal of Political Economy* 115 (2007): 103-140.

Chandra, A. and Skinner, J. "Technology Growth and Expenditure Growth in Health Care," Working paper (2010).

Chandra, A., Fisher, E., and Skinner, J. "Outcomes Assessment and Health Care Reform," Working paper (2010).

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

Garber, Alan M., "Advances in Cost-Effectiveness Analysis of Health Interventions," in Handbook of Health Economics. Edited by J. Newhouse and A. Culyer, 2000, Elsevier B.V., North Holland.

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

Jena, A.B. and Philipson, T.J. "Cost-Effectiveness as a Price Control," *Health Affairs* 26 (2007): 696 – 703.

Jena, A.B. and Philipson, T.J. "Cost-Effectiveness Analysis and Innovation," *Journal of Health Economics* 27 (2008): 1224-1236.

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

Lakdawalla, D., and Sood, Neeraj, "Health Insurance as a Two-part Pricing Contract," *NBER Working Paper 12681* (2006).

Lakdawalla, D., and Sood, Neeraj, "Innovation and the Welfare Effects of Public Drug Insurance," *Journal of Public Economics 93* (2009): 541-548.

Meltzer, D., "Accounting for Future Costs in Medical Cost-Effectiveness Analysis," *Journal of Health Economics* 16 (1997): 33-64.

Newhouse, J. "Medical care costs: how much welfare loss?," *Journal of Economic Perspectives* 6 (1997): 3–21.

National Institute for Clinical Excellence. *NICE's Guide to the Technology Appraisal*

Process, London: NICE, April 2004.

Raftery, James. NICE: Faster Access to Modern Treatments? Analysis of Guidance on New Health Technologies. *British Medical Journal* 323 (2001): 1300-03.

Raferty, James. "Review of NICE's recommendations, 1999-2005," *BMJ* 332 (2006): 1266-1268.

Schulper, M., M. Drummond, and O'Brien, B.. Effectiveness, Efficiency, and NICE. *British Medical Journal* 322 (2001): 943.

Weinstein, M.C., and Stason, W.B., "Foundations of Cost-Effectiveness Analysis for Health and Medical Practices," *New England Journal of Medicine* 296 (1977): 716-721.

Appendix

Relationship between social surplus and cost-effectiveness criteria

Given the inverse demand function p(y,q) corresponding to the demand function d(p,q), the social surplus from an invention is given by:

$$\int [p(s,q)-c]ds = g - yc$$

Here, the gross consumer surplus is denoted by $g = \int p(s,q)ds$ and the second term, yc, is the total cost of production. Therefore, the social surplus is larger for one product over another if:

$$g - yc > g' - y'c'$$

At equal quantities, $y = y'$, this is equivalent to:

$$(g/y - g'/y')/(c-c') > 1$$

where the numerator is the incremental gross surplus for the average individual. Treatment of quality q is incrementally more cost-effective than treatment of quality q', i.e. $(q-q')/(c-c') > 1$,, whenever the average gross surplus equals the quality level:

$$g/y = q$$

In other words, when quality levels represent average willingness to pay, the two criteria coincide. This may be interpreted as a cost-utility- or a willingness-to-pay criterion and previous analysis has analyzed the differences between these criteria at length. For example, standard justifications using $100,000 per quality adjusted life year as a threshold for payer adoption of technologies is based on the value of a healthy life-year of $100,000, $g/y = \$100,000$, derived from empirical studies of labor markets or other tradeoffs between income and mortality.