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. This project was partly supported by The George Stigler Center at The University of Chicago (Philipson), 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). 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
JEL No. I0,I1,I11,I18

ABSTRACT

As health care costs continue to rise, governments and private payers are being forced to make difficult coverage decisions about new health care treatments. Cost-effectiveness (CE) analysis is the main method used to prioritize this spending. The self-evident efficiency rationale for CE is that resources should be spent where they have the highest health impact. This has led to perhaps the largest field in health economics which attempts to provide better estimates of value through CE analysis. However, the costs invariably used in CE analysis are prices set by producers rather than resources used to produce treatments. Therefore, observed CE levels are endogenous because the pricing of new technologies is chosen to maximize profits. This is important because optimal prices, and hence observed CE levels, are affected by demand factors such as patient/doctor demand and payer adoption policies. This implies that traditional measures of "costs" reflect these demand-determined mark-ups rather than resource costs and moreover, CE-based reimbursement policies affect the endogenous CE levels payers observe. Reimbursement based on endogenous CE may therefore bear little relationship with efficient use of scarce medical resources. Using data from technology appraisals by the National Institute for Health and Clinical Excellence (NICE), we test for conditions under which adoption based on standard CE analysis may lead to adoption of more inefficient technologies in terms of resource use.

Anupam Jena
Department of Medicine
Massachusetts General Hospital
55 Fruit Street
Boston, MA 02114
ajena@uchicago.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 relied on combined measures of the benefits and costs of new technologies. These measures include cost-effectiveness, cost-utility, or 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. Thus, CE analysis offers an important means to allocate scarce health care budgets, whether privately or publicly funded. CE thresholds, which dictate that a given technology will be reimbursed only if the incremental costs per quality-adjusted life year (QALY) they provide are below a given threshold, is one way in which CE-based adoption is implemented in practice. The most prominent examples are the UK's National Institute for Clinical Excellence (NICE) and Australia's Pharmaceutical Benefits Advisory Committee[3]. 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 static 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 spending.

[1] See e.g. Newhouse (1992).
[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 and Phelps (1997), Garber (2000), Cutler and McClellan (2001), and Cutler (2005).
[3] 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.

When CE analysis is used to guide reimbursement in practice, 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 costs. Therefore, prices – marked up over costs – determine the CE levels observed for new patent-protected innovations, not the production costs which ordinarily determine the efficient use of resources.

Because prices are chosen to maximize profits to producers, this implies that actual CE levels are *endogenous* and respond to a firm's incentive to mark up technologies above their production cost. This has the important implication that the *demand-side* factors that drive mark-ups also drive observed CE levels. In fact, because producers face two customers, payers adopting the technologies and patients/doctors using them, the price-sensitivity of both parties jointly determines the mark-up. In particular, the CE levels observed depend on how CE assessments are used in technology adoption. For example, if a payer only pays when technologies are cheaper than a fixed CE threshold, manufacturers may find it in their best interest to price up to that threshold regardless of production costs. Therefore, when demand-side factors affect optimal mark-ups in this way, using CE assessments to guide resource allocation will not necessarily have the intended results; the highest "bang-for-the-buck" rationale for CE fails because demand factors are included in the "buck". In general, the main argument of this paper is that the rationale for using CE assessments for health care adoption is weakened when those affected by such adoption behavior act in their own self-interest.

Section 2 of this paper begins by deriving a specific condition for when CE rankings 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 of the cheapest technologies with the largest health impacts. Reversals occur when mark-ups are negatively related with exogenous CE levels, so that the treatments with the lowest resource use are also those marked up the most.

Because the relationship between mark-ups and exogenous CE is what drives the possible discordance between exogenous and endogenous CE, we characterize what drives mark-ups in an environment of dual demand by patients/doctors and payers. In the simplest case when payers accept all treatments, mark-ups depend only on standard demand elasticities of patients and doctors. The focus of this paper, however, is on how adoption policies impact mark-ups and drive reversals in cost-effectiveness. We show how public CE-based reimbursement determines mark-ups in conjunction with patient demand, and hence determines the level of endogenous cost-effectiveness chosen by firms seeking reimbursement. We demonstrate that partial reversals may occur under common forms of technology adoption criteria, e.g. fixed CE thresholds partly utilized in the UK. Moreover, we show that when political or bureaucratic factors impact adoption behavior, so that differences in cost-effectiveness alone do not fully explain adoption decisions, full reversals may occur. This is because favorable adoption policies for certain disease classes drive up prices submitted by firms to payers, so much so that socially less cost-effective treatments in more favorably adopted classes may command higher prices than more cost-effective treatments in less favorably adopted classes.

Based on the analysis of the previous sections, section 3 proposes a test for whether a given adoption procedure leads to CE reversals. 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 mark-ups are unobservable, a standard and central empirical problem in industrial organization. Our test therefore concerns whether observed patterns of treatment adoption based on CE could lead to reversals. We find that reversals will always occur when political or bureaucratic factors lead to differential adoption behavior across *classes of treatments*. A simple test for the possibility of reversals can therefore be implemented by testing for the significance of class-dummies in a regression of adoption on observed endogenous cost-effectiveness levels. 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 such class specific adoption behavior.

In Section 4, we discuss some of the shortcomings of the paper and some of the many future research areas these suggest.

2. Endogenous cost-effectiveness

2.1 Basic framework

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. As a basic framework, 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:

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

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.

The *endogenous* cost-effectiveness ratio uses the price faced by public 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 mark-up above costs, it is defined as $p = m \cdot c(q)$. It follows immediately that the two forms of cost-effectiveness are related by:

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

This implies that resource allocation decisions based on endogenous cost-effectiveness may bear little relationship to the efficient choices that exogenous cost-effectiveness analysis would normally deliver. In particular, the difference between exogenous and endogenous CE is important because treatment adoption based on endogenous CE may

lead to the selection of less cost-effective treatments in terms of exogenous resource costs. To illustrate, consider two treatments, the first of which is more cost-effective in terms of exogenous societal resource use:

$$CE_X(1) \leq CE_X(2) \tag{4}$$

Endogenous cost-effectiveness will lead to a reversal in the most cost-effective treatment if:

$$CE_N(1) > CE_N(2) \quad \Leftrightarrow \frac{m_1}{m_2} > \frac{CE_X(2)}{CE_X(1)} \tag{5}$$

Such "CE-reversals" amount to changes in the ranking of CE levels from best to worst. These reversals occur when the offsetting mark-up differences are larger than the exogenous cost-effectiveness differences. This takes place when mark-ups are negatively related to production costs so that low-cost treatments are marked up relatively more. For example, compared to medical devices, small-molecule drugs may have smaller costs of production yet face larger mark-ups due to inelastic demand or lower competition.

2.2 Dual demand and mark-ups

As the mark-up of prices over costs is the key determinant of the concordance between endogenous and exogenous CE, it is important to understand what drives mark-ups. In privately or publicly insurance, mark-ups are non-standard as producers face two demand sides; the payer adopting the treatment and the patients or doctors using it. Therefore, the price-sensitivity of both sets of customers will affect optimal pricing.

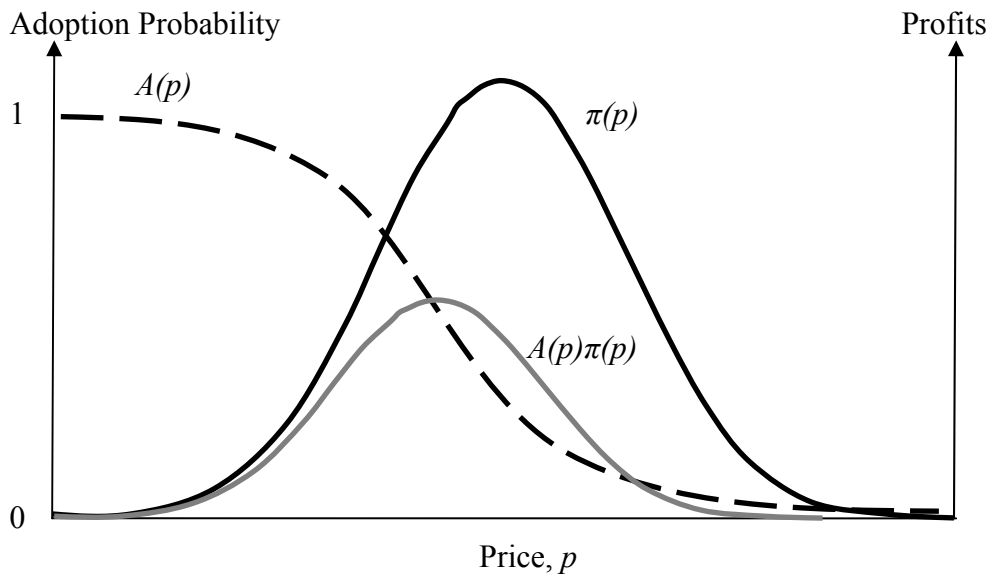
More precisely, consider a producer who chooses a price *p* for a given treatment sold to a payer. Assume that the patient cost-sharing is given by *s(p)* resulting in the quantity demanded *D(s(p),q)* given cost-sharing and quality. If the technology is adopted

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)* and assume that it is a differentiable and decreasing function of price. A special case would be when the chance of coverage is decreasing in the cost-effectiveness ratio of the technology, i.e. *A(p)*≡ *A(p/q)*. The monopolist's expected profits at the time of approval are his post-approval profits discounted by the probability of treatment adoption:

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

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 doctors using the product once adopted. The optimal price balances the profit impacts of the two demand sides and satisfies the necessary first-order condition:

$$A'\pi + A\pi' = 0 \tag{7}$$

The gain in profits conditional on adoption at a higher price must be balanced with the larger chance of not being adopted. Under the maintained assumptions, it follows that producers will not set price low enough to guarantee acceptance since the probability of rejection, *1-A*, is strictly positive at the optimal price. Producers take the risk of rejection in exchange for the larger profits obtained when the submitted treatment is adopted. However, since the probability of adoption falls with price, $A'<0$, the first-order condition directly implies that the price that maximizes expected profits ($A\pi$) is lower than the price that maximizes profits conditional on reimbursement (π). Producers do not maximize ex-post profits for fear of not getting the technology approved at such a high price.[4]

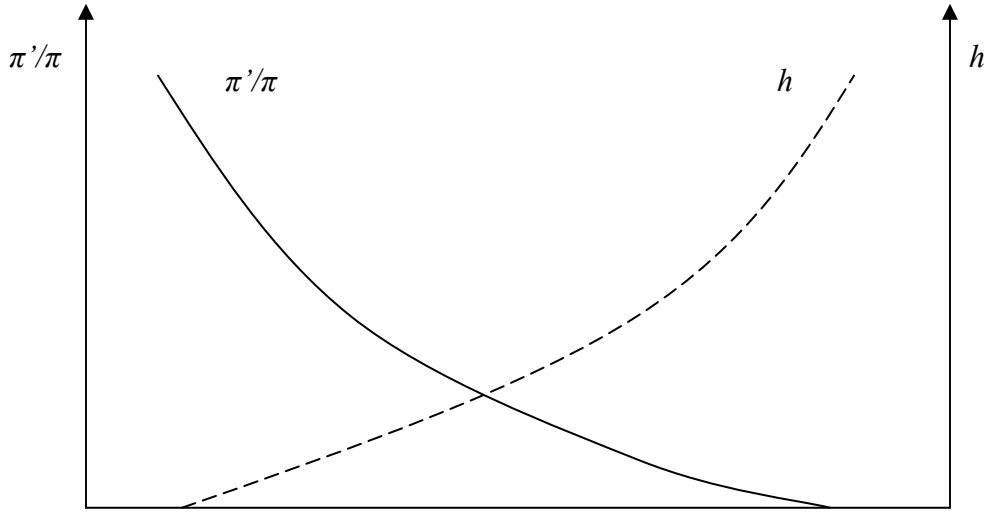
Equation (7) can be rearranged to further elucidate the conditions which determine the profit-maximizing price:

$$-A'/A = \pi'/\pi \tag{8}$$

Figure 2 illustrates that under standard conditions on profit functions (increasing and concave in output price), π'/π is downward sloping in price and there is a unique optimal price whenever $h = -A'/A$ (the hazard rate of technology rejection) is weakly increasing.

[4] Prices failing to satisfy the classic Lerner condition of optimal pricing may therefore still be consistent with profit-maximization.

FIGURE 2—Impact of adoption on optimal price



We will maintain this sufficient condition on the adoption rule for a unique price throughout. This figure therefore shows how both demand sides, patients and payers, affect pricing and mark-ups. The figure has the direct implication that the larger the chance of technology rejection, the higher the *h* curve is, and the lower is the optimal price. For example, the case of a constant rejection hazard, $A(p) = e^{-rp}$, would correspond to a horizontal h function at the level r, implying that the optimal price would be decreasing in r. In the special instance in which the hazard r is equal to zero, so that all technologies are adopted regardless of the price, π'/π would intersect h at the x-axis and the optimal price would be that which maximizes ex-post profits.

2.3 Reversals in cost-effectiveness

Given that both demand sides affect mark-ups according to a function denoted by *m(A,D)*, differences in patient demand and competition, as well as differences in adoption

rules, may lead to differences between observed endogenous cost-effectiveness rankings and unobserved exogenous rankings.

For example, first, consider the extreme case when all treatments are adopted, $A=1$, regardless of their cost-effectiveness. This is arguably the case in the US where FDA approval of a technology is sufficient for Medicare reimbursement. 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 mark-up determination, the elasticity of patient or doctor demand ε yields a mark-up of:

$$m = 1/(1+1/\varepsilon) \tag{9}$$

Therefore, reversals may occur when the most cost-effective treatments are also the most inelastically demanded by patients or doctors. In particular, a negative relationship between mark-ups and costs of production may occur when low cost treatments are produced in less competitive markets.

Next consider when higher prices lower the chance of coverage, $A' < 0$, as would be the case when cost-effectiveness drives adoption. An adoption rule that leads to a negative relationship between the mark-up it induces and exogenous cost-effectiveness could possibly result in ranking reversals in cost-effectiveness. To illustrate, consider when public payers follow a *reservation price policy*, adopting only treatments priced below the reservation price. In the UK, this may be displayed by a strict "CE-threshold" policy in which technologies whose CE levels exceed a given threshold T are not adopted while those whose CE levels fall below are; $A(p/q) = 0$ 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, the

proportionate effect on profits is $\pi'/\pi = 1/(p-c)$ and the hazard function is zero below the threshold and infinite above it. Optimal pricing would lead to the endogenous CE set to the threshold, which would in turn induce mark-ups that are inversely related to the exogenous CE levels:

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

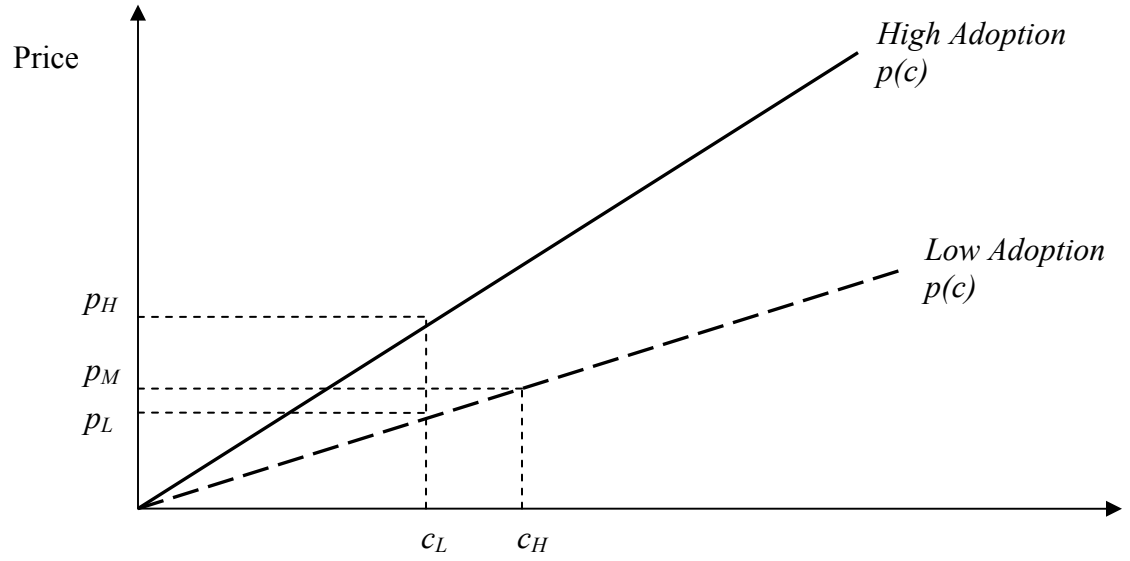
The adoption rule A (represented by T) may therefore induce mark-ups that are negatively related to resource costs. Because of this negative relationship, changes in real resource use – as reflected by exogenous CE levels – would have no impact on endogenous levels of CE used for payment purposes.[5]

Fixed CE thresholds highlight an instance in which treatments of differing exogenous CE may have identical endogenous CE levels. Strict differences in exogenous CE would therefore be ranked the same way. Adoption policies can more generally cause "full" reversals in cost-effectiveness to occur, a situation in which strictly *more* cost-effective treatments in an exogenous sense are strictly *less* cost-effective endogenously. For example, suppose that payers employ differential adoption policies towards treatments based on disease. For example, certain diseases may be deemed more politically important (e.g. HIV/AIDS or breast cancer in the US) and therefore face easier approval. In this situation, factors other than a treatment's submitted cost-effectiveness may play an important role in whether the treatment is ultimately adopted. Holding quality constant, this would be reflected by shifts in the acceptance (A) and hazard (h) functions that lead to different optimal prices being charged for treatments of the same exogenous cost-effectiveness. The effect of this heterogeneity in adoption is illustrated in

[5] 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.

Figure 3. Among treatments of the same quality, this figure shows two curves that map out the optimal price of a treatment as a function of its exogenous CE level. These curves therefore represent the intersection of the proportionate profit curve π'/π and the hazard curve h as the proportionate profit curve shifts outward due to increasing costs c for a given quality level.[6]

FIGURE 3—Reversals in cost-effectiveness under differential adoption behavior



Holding quality constant, the x-axis corresponds to both 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, all else equal, are less favorably

[6] Price is an increasing function of c as the level of profits is decreasing in costs and the marginal profits π' is increasing in costs; $d(\pi')/dc=-D'>0$.

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.

Furthermore, 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.

More generally, this type of reversal will always occur under a compact support of costs of treatment, as long as there is heterogeneity in adoption behavior across classes. In that case, two costs that reverse rankings would always exist. This result suggests that as coverage or adoption decisions grow to be increasingly politically motivated – being determined by forces other than and not implicitly or explicitly referencing cost-effectiveness – the possibility of reversals in cost-effectiveness will rise. While precisely identifying the presence of CE reversals is impossible since mark-ups are unobservable, our result implies that a sufficient generic condition for reversals to occur is that adoption behavior vary across classes of diseases. In the section that follows, we test for this empirically using data on adoption behavior by the National Institute for Health and Clinical Excellence (NICE) in the UK.

3. Adoption behavior and reversals: An illustrative analysis of NICE

Using data on treatment adoption decisions by the National Institute for Health and Clinical Excellence (NICE) in the United Kingdom, this section illustrates how one may test for CE reversals through heterogeneity in adoption behavior.

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 manufacturers, independent academics, and professional and patients' groups. 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 appraisal by NICE

Since its inception in 1999, NICE has published 141 guidances. Our data analyzes the 86 guidances submitted to NICE between 1999 and 2005—the dates of guidance publication range from 2001 to 2007.[7] 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

[7] We are thankful to James Raftery for providing us with his detailed collection of these guidances.

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		
Category of disease addressed by treatment	Percent		
Arthritis	6		
Cancer	18		
Diabetes	6		
Heart	10		
Influenza	6		
Leukemia	3		
Mental health	11		
Surgery	3		
Therapy, not mental Health	11		
Other	26		
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		
Between £30,000 and £40,000	16		
More than £40,000	18		

Source: NICE published treatment guidances, 1999 – 2005.

The 145 NICE guidances present in our data span a relatively large group of diseases and categories of treatment. The largest share of guidances dealt with treatments for cancer (18%), heart disease (10%), and mental health (11%). Of the treatments included 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 or without restrictions.

As shown in Table 1, NICE does not always explicitly calculate or report cost-effectiveness for each treatment so estimates only exist for roughly half (76/145) of the observations in our data. For those guidances for which cost-effectiveness estimates do exist, NICE measures CE 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. 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.

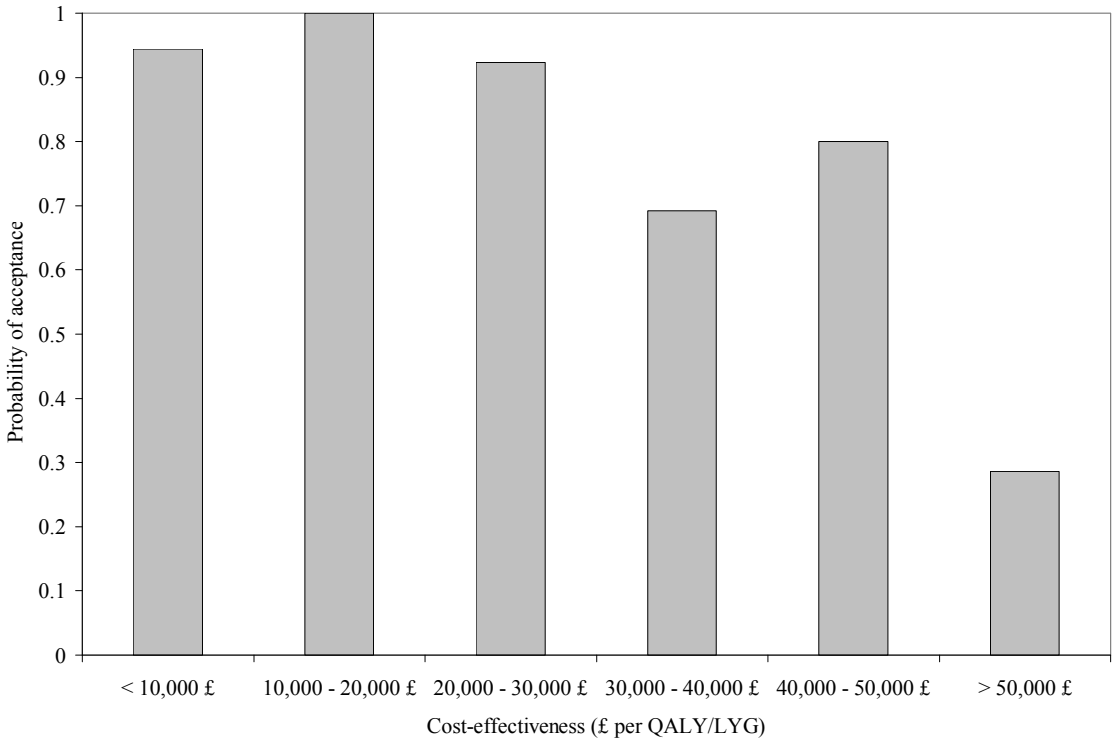
Because measuring effectiveness precisely can be difficult, NICE guidances often report high, mean, and low estimates of cost per quality-adjusted life year or standard life year gained for each treatment. For those treatments for which high and low estimates exist, Table 1 presents the average cost per QALY or LY gained within each range. The within-treatment uncertainty in these estimates is clear—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. There is substantial variation across treatments as well. Among those guidances reporting only costs per QALY gained, the mean cost per QALY was just over £28,000, with a standard deviation of nearly £19,000. At the same, among those guidances reporting only life years gained, the mean cost per LY was roughly £27,000 with a standard deviation of approximately £11,000. 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.

Table 1 also provides information about the distribution of estimated mean cost-effectiveness levels present 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 negative 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. In particular, it plots the proportion of treatments that are recommended by NICE within six groups of cost-effectiveness: less than £10,000, £10,000 - £20,000, £20,000 - £30,000, £30,000 - £40,000, £40,000 - £50,000, and above £50,000.

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



Source: NICE published treatment guidances, 1999 – 2005.

Among the most cost-effective treatments, the probability of a positive NICE recommendation is nearly unanimous and generally declines with the level of the cost-effectiveness ratio. Figure 4 is consistent with the discussed prediction that under probabilistic reimbursement, 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.

3.3 Class heterogeneity and reversals: NICE adoption behavior by disease class

Our analysis implied that differences in treatment adoption behavior across disease classes will induce reversals in CE rankings. To test for heterogeneity in adoption behavior across classes, Table 2 further characterizes NICE's adoption process by specifying how the probability of acceptance of treatments by NICE has varied by disease class and endogenous cost-effectiveness.

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	< 10	10 - 20	20 - 30	30 - 40	40 - 50	> 50
Arthritis	0/0	5/5	0/0	2/2	0/0	0/1
Cancer	6/6	8/8	3/4	5/5	2/3	0/0
Heart	6/6	1/1	4/4	0/0	0/0	0/0
Infectious	2/2	0/0	2/2	0/3	1/1	¼
Mental	0/1	4/4	0/0	1/2	0/0	0/1
Prevention	1/1	1/1	2/2	0/0	0/0	0/0
Other	2/2	1/1	1/1	1/1	1/1	1/1

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.

For a given range of cost-effectiveness, 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 6 treatments for cancer with submitted CE levels below 10,000 £/QALY, 6 were accepted by NICE with minor, major, or no restrictions. For cancer treatments with submitted CE levels in the range of 20,000 – 30,000 £/QALY, 3 out of 4 treatments were adopted by NICE, while in the range of 30,000 – 40,000 £/QALY, all 5 submitted treatments were accepted. Importantly, however, in the same range of 30,000 – 40,000 £/QALY, 0 out of 3 treatments for infectious disease were accepted as well as 1 out of 2 submitted treatments for mental health. 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 for whether the probability of treatment acceptance depends not only on submitted cost-effectiveness, but on the disease being treated 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 diseases shift the relationship between submitted CE levels and the probability of adoption.

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 $p < 0.05$.

Because individual guidances issued by NICE often include a range of CE estimates, we estimate the impact of CE on adoption by using the mean endogenous CE level reported by NICE. The excluded disease class was the smallest class, diabetes. Consistent with the descriptive data reported above, Table 3 demonstrates a statistically significant negative relationship between submitted cost-effectiveness and the probability of treatment acceptance; the probability of acceptance declines by an estimated 0.009 (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 (0.120) and -0.310 (0.132), respectively). Because several of the estimated disease effects are significantly different from zero, this suggests the possibility of heterogeneity in treatment acceptance across disease classes, holding submitted CE constant. In fact, a simple *F*-test rejects the null hypothesis that adoption behavior is identical across disease classes ($p = 0.03$).

4. Conclusion

This paper examines CE-based technology adoption in the presence of optimal pricing by firms. Such pricing implies that that observed cost-effectiveness levels are *endogenous* to the criteria used to guide treatment adoption decisions. Our main finding is that endogenous cost-effectiveness may not relate in any systematic way to exogenous measures that reflect true resource costs. This occurs because both demand factors and adoption policies determine prices; prices, in turn, affect endogenous CE rankings but not exogenous CE rankings. This implies that the intended value of using cost-effectiveness, to economize on resource costs used to deliver health care, may not be present.

Our analysis has several important limitations that future research may successfully attempt to deal with. 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 mark-ups are unobservable is, of course, well-known and long-recognized in empirical industrial organization. This issue lead us to state our results as a failure of a given reimbursement procedure, rather than a failure 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 do 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. For example, reductions in price for a device used in surgery would 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. In fact, 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 does not consider the comparative effectiveness of multiple competing treatments. Such an analysis would consider the duopoly and oligopoly pricing implications of making reimbursement decisions contingent on the vector of industry prices and quality levels, as opposed to the single price of a monopolist. When setting prices, producers presumably take into account how reimbursement authorities use CE levels of competing treatments, whether branded or generic, for similar conditions. The industrial organization of endogenous cost-effectiveness analysis, and its

impact on the growth of public health care spending, is an important area of future research.

Lastly, we do 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. 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 stated criteria in explaining public adoption decisions.

Despite the shortcomings of our analysis, however, we believe the overall concern that we raise deserves serious consideration in interpreting the impact of CE-based adoption policies on public health care spending. These adoption policies may not have their intended goals when those affected by them act in their own interest.

References

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

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

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

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.

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

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.

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.