November 21, 1997
edit: August, 1998

<center>

ACCOUNTING FOR HEALTH CARE:

INTEGRATING PRICE INDEX AND COST-EFFECTIVENESS RESEARCH

Jack E. Triplett
Brookings Institution

</center>

Personal health care *expenditures* in the National Health Accounts increased at an average annual rate of 10.2 percent per year between 1985 and 1992. Were health care *prices* also rising rapidly between 1985 and 1992? That is less clear.

Health expenditures are the product of price and quantity. Determining whether expenditure increases are caused by price increases requires a methodology for separating health care spending into price and quantity changes--into medical price inflation and increases in the output of health care. Economic statisticians lack an adequate methodology for this task.

For the same reason, it is also not clear why the U.S. commits 14 percent of Gross Domestic Product to health care, a far larger share than for any other major industrialized country. Does the U.S. health care sector have inefficiently high costs, as some seem to believe? Or in significant ways does the U.S. provide a higher level of health care than other countries with which it is compared? Better methodology for separating health care spending into price and quantity components would provide information to answer these questions.

Inadequate economic measurement can have serious consequences for health care policy analysis and for public debate. When economic statistics for the health care sector do not accurately separate price from quantity, health policy actions undertaken to suppress presumed medical price inflation will be inappropriate if in fact it is the quantities that are rising and the output of medical care is expanding. Separating price from quantity is not the end of the story: It has also been asserted that the U.S. employs too many complicated medical procedures (that is, the quantities are too many). But partitioning medical expenditure into price and output change is clearly necessary for subsequent analysis.

BACKGROUND

Accounting for price and quantity change in health care is an old problem that has been discussed repeatedly in the price index literature and in medical economics. Historically, the Consumer Price Index (CPI) medical care price index has been used to measure medical inflation. CPI medical care price index methodology has also determined the concepts for measuring medical care output, because output measures are typically produced through deflation--that is, by dividing the change in expenditures by a price index. In the U.S., medical care output measures appear in the national income and product accounts (NIPA), published by the Bureau of Economic Analysis, and the national health accounts (NHA), published by the Health Care Financing Administration.

The CPI medical care index was until recently constructed from a sample of medical care transactions: a hospital room rate, the price for administering a frequently-prescribed medicine, or the charge for a visit to a doctor's office. Transactions of this kind are sufficiently standardized that the same transaction can be observed repeatedly, which is required for a monthly price index.

It was widely noted, even 35 years ago, that CPI methodology did not adequately allow for improvements in medical care. For several reasons, an improvement in medical procedures that raised the cost of treatment but also improved efficacy frequently showed up as an increase in the CPI medical care component. When this CPI component was used as a deflator, the improved medical care procedure was thereby inappropriately removed from the medical output measure.

An early alternative to CPI methodology was the idea of pricing the "cost of a cure," estimating the cost of a medical procedure (the treatment of appendicitis, for example). George Stigler, in testimony on the "Stigler Committee Report" remarked (in U.S. Congress, Hearings (1961), page 533): "...we were impressed by some of the preliminary work that has been done...on problems such as the changing cost of the treatment of a specific medical ailment.... We think it would be possible...to take account of things such as the much more rapid recovery and the much shorter hospital stay...." This contrasted with the CPI's focus on hospital billing elements for a medical procedure, such as the hospital room rate and the administration of a pain medication.

Scitovsky (1967), using data collected from hospitals, estimated the cost of treating selected medical conditions, including appendicitis and otitis media, and compared the cost trends with the CPI. Her findings, though, did not confirm CPI upward bias, a result that most economists regarded as puzzling.[1] Rice and Horowitz (1967) reported on a special study conducted by the Bureau of Labor Statistics (BLS); again, the study indicated that prices for a selected group of surgical and hospital procedures[2] rose less than the CPI physicians' fee index, and less also than the CPI medical care services component. The study covered only 15 months, bridging the time that Medicare first became operational. Few similar cost of illness studies have been conducted until quite recently.

Reder (1969, page 98) proposed to by-pass the medical care pricing problem altogether by pricing medical insurance: "If medical care is that which can be purchased by means of medical care insurance, then its 'price' varies proportionately with the price of such insurance." Barzel (1969) estimated an insurance measure of medical price inflation, using Blue Cross-Blue Shield

---

[1] Scitovsky (1967) suggested that the CPI had understated the rate of medical inflation because actual average charges had advanced relative to the "customary" charges that presumably went into the CPI. In recent years, it has been asserted that the error from "list" prices goes the other way--it raises the CPI, relative to the true rate of medical inflation; see Newhouse (1989).

[2] The hospital procedures included the treatment of heart attacks, for which the "most frequent" hospital fee was $110, not including the cost of lab tests or medicines, for a hospital stay of 21 days. It is interesting that for heart attacks treated by medical management, which is most nearly comparable to the treatment that was in use in the 1960's, Cutler, McClellan, Newhouse and Remler (1996) report an average Medicare payment in 1991 of $10,783, including drugs and lab tests.

plans. Feldstein (1969, page 141) objected that the cost of insurance approach "is almost certain to be biased upward" because "average premiums will rise through time in reflection of the trend toward more comprehensive coverage" and because the insurance plans will purchase "more services or services of higher quality." Moreover, if an epidemic occurs that raises the cost of insurance (AIDS, for example), it will inappropriately show up as an increase in the price of medical care, and therefore not an increase in its quantity, unless the medical premium were calculated net of utilization rates. It has now generally been recognized that measuring the output of insurance is also hard conceptually (Griliches, 1992), so using insurance as a proxy is not an attractive solution to the problem of measuring health care. Indeed, HMO's and capitation plans pose more difficult measurement problems than traditional fee-for-service provision of health care.

As these references from the 1960's suggest, the major issues on health care output were joined years ago. Until recently, the debate on the measurement of prices and output in the medical sector largely repeated the arguments of 30 years ago. The empirical work and the data, too, had not advanced that much beyond the mid-1960's (Newhouse, 1989).

Several things have changed recently. First, the Bureau of Labor Statistics, in 1992 in the Producer Price Index and in 1997 in the CPI, has introduced new medical price indexes that are substantial improvements on what existed before (Catron and Murphy, 1996; Cardenas, 1997). However, the basic problem remains: If, for example, a better but more expensive treatment for heart attacks becomes available, the more expensive treatment will probably show up as medical price inflation unless the BLS can get hospitals to place a value on the improved treatment. Typically, hospitals cannot do this.

Second, a major new research initiative on health care price indexes, using new approaches and new sources of data, has been undertaken by a research group centered in the National Bureau of Economic Research. Some of these studies are presented at this conference.

Third, information on health care outcomes has been enhanced greatly by recent research on "cost-effectiveness analysis" within the medical establishment itself. Gold, Siegel, Russell, and Weinstein (1996) provide a comprehensive review of cost-effectiveness procedures.

The present paper is part of a larger project that is directed to the task of building these new price indexes and health outcome measures into a "health account" for the medical care sector. Present accounting systems for health care are organized around an *institutional* framework (hospitals, doctors' offices, and so forth). In Triplett (1998), I contend that improvements in measuring the price and output of medical care can best be made by shifting away from this institutional orientation toward cost-of-disease accounting, of the type pioneered in the U.S. by Rice (1966) and recently expanded and updated by Hodgson and Cohen and Cohen (1998), and by constructing cost-of-disease accounts in a time series context, as are other accounting systems (time series comparability has not been a priority for cost-of-disease accounts in the past). Cost-of-disease accounts can make use of economic information for specific

diseases, such as price indexes for, e.g., heart attacks (Cutler, McClellan, and Newhouse, this volume) and mental depression (Frank, Berndt, and Busch, this volume), and of information on cost effectiveness of alternative treatments for specific diseases. It is much harder to use this information in an accounting system for, say, expenditures in doctors' offices.

Studies on price indexes for diseases and cost effectiveness studies on treatments for diseases have so far been conducted independently, with no interaction between them. Combining information from those two sources of information requires an analytic framework for how they fit together. Filling this gap is the subject of this paper.

COMPARING PRICE INDEX AND COST-EFFECTIVENESS METHODS

A matrix of hypothetical health care information is given in table A. I will use this information to demonstrate the relation between price indexes and cost-effectiveness research.

The subject of table A is a particular disease, a code or classification from the International Classification of Diseases, Ninth Revision, or ICD-9 (Puckett, 1997). An example of an ICD classification is ICD-9 code 410, acute myocardial infarction, or heart attack, which has been the subject of considerable research, encompassing both price index research (Cutler, McClellan, Newhouse and Remler, 1996) and cost-effectiveness studies. ICD-9 code 410 is located in ICD-9 "chapter 7," which is titled "Circulatory Diseases." Note that U.S. Diagnostic Related Groups, or DRG's, can be mapped into ICD-9 codes (Tambourine, Fish, and Ehrhardt, 1993).

The rows of table A designate four alternative medical interventions for the specified disease. Two of them (treatment A and treatment B) are available in period 1; a new intervention (treatment C) becomes available in period 2, and another new one (treatment D) in period 3, at which time the oldest and (by specification, for purposes of the example) least effective treatment (treatment A) is no longer employed.

In the cells of table A, four types of information are recorded. The first piece of information is the transaction price for each treatment in each period--p_{a1} is the transaction price of treatment A in period 1, p_{a2} the price for the same treatment in period 2, and so forth. By transaction price, I mean the charge made by a medical provider for administering the specified medical intervention. The medical price index literature is full of discussions about the difficulty of obtaining the transactions price (as opposed to the list or "chargemaster" price), and about whether "the" price makes any sense, given the tendency to price discriminate among patients according to income or ability to pay or class of payer. It is also the case that much variation exists among the medical conditions of patients who may fall into the same ICD-9, or DRG, code, and this variation affects the treatment, its cost, and therefore the amount charged. I abstract from these considerations, without implying in any way that I denigrate their importance and relevance, or the intractability of the problems they present. If necessary, one can think of the p-terms in the table as "average transactions prices."

The treatment methods are arrayed in table A from the least expensive to the most expensive, that is, $p_{at} < p_{bt} < p_{ct} < p_{dt}$, for each time period, t. This implies that newer methods are always more expensive than older ones,[3] and also implies that when treatment methods go out of use, they are the less expensive ones (compare the arrays for periods 2 and 3). It is not necessary to make any assumption about the movement over time of prices for a given treatment, but the exposition will be easier, and probably more realistic, if we assume that prices for each treatment, I, are increasing over time, that is $p_{i1} < p_{i2} < p_{i3}$.

The second piece of information in each cell of table A is the quantity of each medical procedure that is performed in each time period. Thus, q_{b2} is the number of treatment B procedures performed in period 2, and so forth. The pattern of table A suggests that newer treatments enter with small shares and expand and that older ones have declining shares and are eventually supplanted. This is consistent with the shares of treatment regimens shown in Cutler, McClellan, Newhouse and Remler (1996). However, the patterns of relative quantities and their changes are not crucial for present purposes.

The third and fourth pieces of information are a medical outcome measure (M_i), which is unique to a specific treatment but does not depend on time, and a valuation of the medical outcome (V_t) which does vary with time but, ideally, is not unique to the treatment (provided the M_i terms are measured correctly). These will be described and considered when I make use of them, at a later point.
.

Conventional Price Indexes and Output Measures

By "conventional price index," I mean price indexes of the type now produced by the Bureau of Labor Statistics, in which the pricing unit is, in effect, a specified treatment for a defined medical condition. I noted earlier that the CPI medical care price index was once (prior to January, 1997) formed from the price of a hospital room, and so forth, and that the BLS recently has moved to an improved approach. In the current CPI (and in the Producer Price Index, the PPI), the BLS first takes a sample of medical procedures (actually of DRG's). It then samples, from among all the cases in a hospital or physician's office records, an individual patient's bill (or several bills). The characteristics of the patient and of the treatment are specified in great detail. The agency then collects the charges in subsequent periods for the precisely-specified medical case that was chosen in the initial period.[4]

A statistical agency is not likely to produce a price index at so disaggregate a level as an ICD-9 three-digit code, so the following should not be taken as an explicit description of any

---

[3] Frank, Berndt, and Busch (1997) show that this is not always the case, for they present data on a new, and equally effective, treatment that is far cheaper than the old one.

[4] See Catron and Murphy (1996). This BLS procedure is not necessarily employed by the statistical offices of other countries. Very few countries, for example, do probability sampling of price quotations for price indexes.

country's methodology for health price indexes. The BLS, however, does produce PPI indexes at the level of an ICD-9 chapter (for example, "Circulatory diseases"). The principles can be illustrated efficiently by assuming a price index for a more narrowly specified category (ICD-9 code 410--heart attack).

Suppose, then, that the statistical agency has decided to construct a price index covering all the treatments in table A. Typically, the agency will choose a sample. For example, it might choose, either judgmentally or by use of probability sampling (the BLS method), treatment B from among the treatments available in Period 1, or if the sample were chosen in Period 2, perhaps again treatment B, or perhaps treatments B and C (or some other combination). As noted above, sampling in the BLS methodology extends across patient billings, so other characteristics are also involved. Although sampling has substantial importance in practice, sampling considerations are not particularly relevant to the present discussion, and so I will neglect the details of sampling in the following. The array of data in Table A will illustrate the basic problems of constructing price indexes, whether done with samples of treatments or with an exhaustive survey of them.

Statistical agencies typically use the Laspeyres formula to compute the price change for medical treatments. In words, a Laspeyres price index weights the change in each of the prices by the number employed of each treatment in period 1. Symbolically, the calculating formula for this price index (which I denote with a capital letter, P) is [5]:

EQUATION (1): $P_{12} = \Sigma \, p_{i2} \, q_{i1} \, / \, \Sigma \, p_{i1} \, q_{i1}$

Note that the price measure holds constant the treatment (and other aspects of the hospital or physician encounter), so the comparisons that are made for the price index are exclusively *across the rows* of table A.

In the Laspeyres calculation, only period one quantities, q_{i1}, are employed, no use is made of the period 2 quantities, q_{i2}. In period 1, two treatments, only, are available, A and B. The new treatment, C, was not available in period 1. Accordingly, treatment C can have no role in the price index calculation of equation (1)--it has no price index weight in period 1 because $q_{c1} = 0$. This is not an innocuous point, but rather a potentially serious source of error; I will return to it in a subsequent section of the paper.

The next step is constructing an output measure for medical treatments. It is generally true that output measures in economic statistics are calculated indirectly. Output, or quantity, measures are estimated through *deflation* by a price index. The principles are as follows.

---

[5] Actually, at this level the agency typically does not have expenditure weights, though it may have sampling weights, the use of which will give the same formula, as an expectation. It might also use alternative calculations, including a geometric mean, rather than the arithmetic mean implied by equation (1). The best method for calculating such "basic components" in price indexes is a topic that is currently the subject of substantial professional debate; the subject cannot be pursued here.

National accountants, and particularly national health accountants, go to great lengths to collect, estimate and record total national expenditure on medical treatments (Lazenby, et al., 1992). In our example, this is simply the aggregate spending in period 1 on the *two* treatments that were available in period 1; similarly, total expenditure in period 2 is the aggregate spending on the *three* treatments that were available in period 2. Algebraically, these two expenditures amount to (using the information in table A):

EQUATION (2a): $Exp_1 = \Sigma \, p_{i1} \, q_{i1}$, I=a,b
EQUATION (2b): $Exp_2 = \Sigma \, p_{i2} \, q_{i2}$, I=a,b,c

The ratio of (2b) to (2a) gives the change in expenditure between the two periods, which in national accounts language (at least in the U.S.) is often called the change in "current dollar" expenditure. [6]

In national accounts, the change in current dollar expenditure is "deflated" by the appropriate price index to obtain the change in output, which is a quantity change measure. I will denote this quantity change measure with a capital letter, Q. Then, the change in output between Periods 1 and 2 is[7] :

EQUATION (3): $Q_{12} = (Exp_2 \, / \, Exp_1) \, / \, P_{12}$

An important property of this conventional deflation method is: The change in the output of medical procedures between periods 1 and 2 *includes* the output of treatment C in period 2, even though, as noted above, the price index that is used for deflation *does not include* treatment C. Including treatment C in the output measure makes sense because the quantity of treatment C went from zero in period 1 to some positive number of treatments, designated by q_{c2}, in period 2. However, the introduction of the "higher quality" treatment C is outside the CPI sample, it is not included in the CPI. Thus, treatment C's quantity increase in national accounts is implicitly valued by the prices for treatments A and B.

---

[6] In the usual presentation, the ratio is converted to a percentage increase or decrease by subtracting 1 from the ratio, or: $(Exp_2 \, / \, Exp_1) - 1$ = percent change in expenditures.

[7] Readers who are well versed in national accounts practices will note that the algebra in equation (3) results in an unfamiliar expression. Statistical agencies that produce price statistics will almost always use index weights from the earlier period, as shown in equation (1), and will also typically hold the weights fixed over a number of periods (often called a "base-period weighted" index). When the base-period weighted (or Laspeyres) price index is used as a deflator, as in this example, the result is a current-period weighted quantity index, usually called a "Paasche" index (this is apparent from substituting equations (1) and (2) into equation (3)). This differs from the traditional "upper level" aggregation structure of national accounts for most countries of the world, where the upper level corresponds to a Laspeyres quantity index and a Paasche implicit deflator. It also differs from the approach introduced by the Bureau of Economic Analysis in 1996, in which both the quantity measure and the price measure are described by a different index number formula, the Fisher Ideal index number (*Survey of Current Business*, January/February, 1996). But at the very lowest level of national accounts, the implicit weighting structure for output measures is determined by the formula for the deflating indexes, which is what is represented in equation (3).

It is very appropriate to ask whether the price movement for treatments A and B provides a valid basis for valuing the quantity increase in treatment C. The greater an improvement is treatment C over treatments A and B, the more dubious is the use of prices for treatments A and B to value the increase in the quantity of treatment C.

 In the price index literature, this problem is known, somewhat inappropriately,[8] as the "new good" problem. With conventional price index procedures and conventional national accounts procedures, there is real danger that the introduction of new goods will be under valued when the price movements of existing goods are used for valuing the new ones. A discussion of the economic measurement problems posed by new goods is contained in the "introduction" to Bresnahan and Gordon (1997).

In period 3, another price index problem arises, which I will describe briefly to avoid confusion by its omission. A new treatment (treatment D) becomes available in period 3, and an old treatment (A) is no longer used, so it disappears from the data that are collected for price indexes and national accounts.

The introduction of treatment D in period 3 is exactly parallel to the introduction of treatment C in period 2. But presuming that the price index sample contains treatment A, the disappearance of treatment A requires that the price compiling agency carry out an "item replacement" within the price index sample.[9] All commonly used methods for dealing with item replacements create potential errors in price indexes, some of which are discussed below.

In summary, the essence of the traditional price index method is to use only the prices of *matched* treatments--the method holds constant the characteristics of the medical intervention and the other terms of the transaction. Price comparisons are made only within rows of table A. Any change in medical costs that arises outside a particular row is ignored.

Cost-Effectiveness Measures

Cost-effectiveness studies and price index studies have very different purposes. Although cost-effectiveness studies may have uses for economic analysis, the cost-effectiveness technique is

[8] By "inappropriate" I do not mean that the subject here is a service, rather than a good. Rather, it is not clear that the four treatments in Table A should be thought of as "different" services; since they all address the same problem (the treatment of a specified disease) they might better best be thought of as different "quality" levels of the same service or commodity, so we should be using language such as "new variety of good" or perhaps "new quality level of service." The language gets cumbersome. I therefore use "new good" without necessarily implying that it is a commodity or service that has never been seen before, only that it has not been available in a previous time in the form in which it is made available today.

[9] Within methodological materials published by the U.S. Bureau of Labor Statistics, this is often referred to as an "item substitution." Methods for handling item replacements in price indexes are discussed in Triplett (1997), and in Moulton and Moses (1997).

mainly an analytic tool for decision-making (Gold, Siegel, Russell, and Weinstein, 1996).

Cost-effectiveness studies use information from table A, but it is somewhat different information from the information that goes into conventional price indexes and national accounts output measures. A cost-effectiveness ratio is computed from two kinds of information. The numerator contains the cost of a medical procedure. *If* the only cost of an illness, or of a treatment, were the charges made by physicians and hospitals, then the numerator of the cost-effectiveness ratio could be calculated from the p_{it} terms in Table A. I explore below cases where medical provider charges are not the full cost of an illness.

The second piece of information in cost-effectiveness studies is a measure of medical outcome, often also called a health outcome. The outcome of an operation for a life-threatening condition might be the survival probability, or the expected number of years of life prolonged by the operation. If an improved operating procedure is developed, its increment to effectiveness could be measured by increased survival probability, or the increase in expected number of life years that it generates. In table A, the medical outcome associated with each treatment--however measured, but measured consistently across treatments--is designated M_i. Because I have assumed, for expository convenience, that more expensive medical treatments have higher effectiveness (though not necessarily proportionately), $M_d > M_c > M_b > M_a$.

Suppose one desired to perform a cost-effectiveness analysis on treatments B and C in table A, and the study were carried out at period 2. Period 2 is, of course, the point at which treatment C was introduced. Then, assuming that medical provider charges are the only costs, the difference in cost for the new intervention (C), compared with the older one (B), is $p_{c2} - p_{b2}$. The increment to effectiveness (the change in medical outcome from use of treatment C, compared with the previously best treatment, B) is $M_c - M_b$. The cost-effectiveness ratio is then:

EQUATION (4): $(C/E)_{cb} = (p_{c2} - p_{b2}) / (M_c - M_b)$

Obviously, with this formulation the lower the cost-effectiveness ratio the more desirable the new treatment is relative to the older one. As Gold, Siegel, Russell and Weinstein (1996, page 27) remark, "interventions that have a relatively low c/e ratio are 'good buys' and would have high priority for resources."

The literature on cost-effectiveness contains a great deal of discussion about the costs that belong in the numerator of the cost-effectiveness ratio and about the way one measures medical outcomes in the denominator.

Costs. Generally, the costs included in a cost-effectiveness study are broader and more comprehensive than the p_{it} terms of table A, because the full costs of a medical procedure include more than the charges for treatment made by the hospital or physician. Recovery from a stroke may involve an extended stay in a nursing home, for example, and other costs, such as the value of care provided by friends and family members, are incurred by the patient. "A primary objective

of cost-effectiveness analysis is to incorporate a consideration of resource consumption into decisions about health care" (Gold, Siegel, Russell, and Weinstein, 1996, page 176). Thus, there are *direct* resource costs, the charges for which are made by medical care providers, including any public subsidies, and *indirect* costs[10], which Gold, Siegel, Russell and Weinstein (1996) categorize into use of non-health resources (they give as an example the use of transportation to obtain treatment), use of unpaid or informal care giver time, and use of the patient's own time in the course of treatment. An extended discussion of costs appears in Gold, Siegel, Russell, and Weinstein (1996), especially Chapter 6.

There is still debate on cost-effectiveness methodology for some aspects of cost, especially certain future costs--see, for example, the positions taken by Garber and Phelps (1997) and Meltzer (1997) and the review of this debate by Weinstein and Manning (1997). For present purposes, particular interest attaches to the methodology applied to the cost of the patients' own time in a cost-effectiveness study. If a new medical treatment economizes on patient time, some contributors to cost-effectiveness studies have argued that the saving should be treated as a reduction in the cost of care, others that it should be treated as an increment to the effectiveness of medical treatment (see below), and still others (Garber and Phelps, 1997) that it makes no difference how time savings are handled in the study so long as the method is applied consistently across studies.

But whatever the existing debate about the elements of cost that belong in a cost-effectiveness ratio, the cost concept that is used in cost-effectiveness studies is broader than the cost or price concept that is measured in traditional price indexes and national accounts. Cost-effectiveness studies, at least in principle, address the full costs of illness; price indexes typically address only the fraction of the full cost that is identified with charges made by medical providers.

Effectiveness measures. Turning now to medical outcomes, many measures of effectiveness have been employed. Probability of survival or expected lifetime is one straightforward quantitative measure that, if not a comprehensive measure of medical outcome, is not inappropriate for some diseases and treatments. For example, the probability of survival from a heart attack is a very relevant measure of medical outcomes for comparisons among alternative treatments for heart attacks, and it remains relevant even if lifestyle restrictions and ability to carry out normal functions ought also to be included in the medical outcome measure.

---

[10] Gold, Siegel, Russell and Weinstein (1996) argue against the use of the term "indirect" costs on the grounds that it has different interpretations in different social science literatures. Although their point is well-taken, it is nevertheless standard in accounting for health costs to partition costs into those that are measured or estimated from health care provider data or from health care expenditure data (direct costs of illness) and those that must be estimated from other kinds of data, such as earnings foregone (indirect costs). Usually, the latter estimates are also less direct in terms of estimating methods than are the former ones, because direct costs are "what are" estimates (the amount paid to a hospital, for example), where indirect costs are "what would have been" estimates (the earnings that would have been received, had the patient not become ill).

For other medical interventions, survival probabilities may be irrelevant (orthopedic treatment of an ankle injury, for example). Indeed, as the cost-effectiveness panel (Gold, Siegel, Russell, and Weinstein, 1996, page 84) remarked:

> In industrialized nations, where length of life has shown steady increases over the past century, it is the improvement in quality of life produced by health care inputs that is often the truer gauge of how well the health care system is performing. For example, in evaluating the effectiveness of cholesterol screening, mortality from health disease is certainly an important outcome. But simply counting deaths, or even life years gained, may leave out other important health outcomes, such as the morbidity repercussions of angina and heart attacks, as well as the psychological concerns that may accompany a diagnosis of hypercholesterolemia. All of these outcomes may be highly relevant in assessing the value of an intervention.

As the passage quoted suggests, a medical outcome measure should incorporate both morbidity and mortality effects into a single measure. For this purpose, the panel recommended a measure known as QALY, the "quality adjusted life year"(Gold, Siegel, Russell and Weinstein, 1996, page 122).

For present purposes, two characteristics of QALY can be noted. First, future years of life gained are discounted in QALY. Second, QALY incorporates community preferences across various health states, in order to weight different changes that are brought about by medical interventions. For more information on the QALY measure, its limitations and research needed to improve it, see Gold, Siegel, Russell and Weinstein (1996), chapter 4, and the range of authors cited at page 93.[11]

It has been clear for many years that measures of health outcomes are a vital step in the ultimate solution to the health care measurement problem (see, for example, Reder, 1969). The development of the QALY and other related measures provides a solution, in principle, to some of the problems with the cost of illness approach that emerged in some of the early research. For example, Scitovsky (1964, 1967) discussed a new treatment for acute appendicitis that, because of potential adverse side effects, was better in some respects (or for some care recipients), but worse in others (or for other recipients). Though it was not recognized at the time, the Scitovsky study showed that the outcome of a medical procedure is generally multi-dimensional, one cannot just consider a "cure" or the principle or primary outcome. Put another way, looking only at the cost of a unidimensional "cure" (appendicitis treatment), without considering the multidimensional attributes or characteristics of a medical procedure, could produce its own bias. Though this problem was intractable with the analytic tools that were available in the 1960's, it has been addressed in the medical outcome research of the past 10-15 years, because both increased benefits of a new procedure, and side effects, if any, are valued in QALY.

In practice, however, much work remains. Bailar has long cautioned that uncareful

---

[11] See also the DALY (Disability Adjusted Life Year), which is approximately 1-QALY, in Murray and Lopez (1996). The latter reference also contains an extended discussion of the measurement problems.

application of outcome measures may distort the apparent effects of a medical treatment (Bailar and Gornick, 1997). Their example was early detection of cancer, with survival rates after treatment as an outcome measure. Some cancers do not kill the patient; if these cancers are detected early and are treated, the treatment may appear effective in terms of patient survival, even though the untreated patient might also have had a high survival probability. This problem exists for all medical outcome measures.

Some additional considerations. The numerator of the cost effectiveness ratio is time-dependent. If the costs diverge for the two treatments being compared, the cost difference in the numerator of the cost-effectiveness ratio will also change. Numerators may vary across countries as well, if the structures of medical costs differ across countries. Intertemporal changes or international differences in costs are more probable the more expansive is the definition of cost. For example, the cost of family care givers is not constant over time, certainly differs across countries, and probably does not have a constant ratio to other included costs.

For these reasons, a cost-effectiveness ratio is a calculation that is unique to one particular time and country. This causes problems because one would typically like to apply cost-effectiveness studies to at least a medium-term period, and would like to apply research carried out in one country to decisions in another.

I have assumed that the denominator--the effectiveness part of a cost-effectiveness ratio-- is not time-dependent. That is, once a researcher has determined the difference in effectiveness between two treatments, there is no reason to assume that the difference in effectiveness will change in some future period. Likewise, if a medical procedure makes a net contribution to effectiveness in one country, one presumes that this research finding will carry over into data for another.

However, medical outcome measures do depend on medical research. New research sometimes shows that a treatment is more effective than originally estimated, and sometimes that it is less effective. This means that, even though effectiveness measures are not themselves time-dependent, they are subject to revision, sometimes large ones. Moreover, outcomes may change because a given procedure may change in its implementation or become more effective with increasing use (evidence suggests learning by doing in operations improves the quality of outcomes). If so, a revision to the outcome measure may also be called for. Revisions are well known in national accounts, but not well liked.

Garber and Phelps (1997) argue that the patient's own time may go into either the numerator or the denominator of a cost-effectiveness ratio. Because the value of the patient's own time is time dependent, and because it will not necessarily move with other changes in costs that are also time dependent, I believe it would be better to put all costs that are time dependent in the same part of the cost-effectiveness ratio. This reason complements other reasons that have been advanced for putting time costs in the numerator of the ratio (see Meltzer, 1997).

I do not want to give the impression that medical outcome measures are fully developed, nor minimize the controversy that still exists. Many problems limit the comprehensiveness and accuracy of medical outcome measures. They are discussed in Gold, Siegel, Russell and Weinstein (1996). In part because of the problems, outcome measures have not found ready acceptance among statistical agencies who compile price indexes and construct national accounts. But there is an additional reason. There has been great reluctance among statistical agency administrators to place an economic value on life years or on QALY's, which seems to be required.[12] I consider this matter in a following section.

A Comparison: Price Indexes and Cost-Effectiveness Studies

As indicated above, price indexes and cost-effectiveness studies differ in a number of ways, largely because their respective purposes differ.

A price index study is designed to produce information about larger or smaller *groups* of medical procedures, taken together. A price index, and its related national accounts output measure, provides information about the aggregate price and quantity movement of some set of medical treatments, at the lowest level of aggregation, probably a level such as an ICD-9 chapter or subchapter, although more detailed indexes are feasible. Of necessity, price index measurements abstract from the detailed movements of prices (and quantities) for individual treatments. Sometimes one also wants to measure international price and output differences prevailing at the same time; the index number procedures are similar.

A cost-effectiveness study provides information for making comparisons among treatments, not for extracting information that is common to all of them. How much better is aspirin, or angioplasty, than alternatives for treating heart attacks? Cost-effectiveness studies are not aggregative, they are designed to produce information at a detailed level--comparisons between or among alternative treatments for a disease diagnosis. And for the most part, the comparisons they make are at the same time period, they are not normally designed to provide information about time trends or about international differences.

Another way to compare price indexes and cost effectiveness studies is to contrast the way they use information from table A. Both types of studies make comparisons among the elements of table A, but they make different comparisons.

The price index compiler always make comparisons *within the rows* of table A, never down the columns. A medical care price index measures changes in the prices of medical treatments over time, holding the treatment constant. This is exactly the information within the rows. Any medical cost changes that occur because new treatments replace old ones (changes that are not within the rows of table A), any price differences that arise from differences in

---

[12] For example, Cutler, McClellan and Newhouse, this volume, place a value on increased life expectancy in order to compute a cost-of-living index for heart attack treatments.

treatments--comparisons that go down the columns of table A--are excluded from the price index. They are regarded as quality change, not price change. As noted already, the within-row information that is used is exclusively the p_{it} and q_{it} data. No use is made of the M_i information in price indexes--but of course the M_i's do not vary across rows, so there is nothing useful in the across-the-row M_i.

The cost effectiveness study, on the other hand, is intended to make comparisons *down the columns* of table A. A cost effectiveness study that compares treatments C and B, say, compares price changes and medical outcome changes between treatments, at a fixed point in time. This is exactly information one gets from going down the columns of table A. The cost-effectiveness ratio uses exclusively the p_i and M_i information, it normally makes no use of the q_i terms. As explained in the following section, the cost effectiveness ratio answers a question that traditional price index methodology cannot answer, and needs to answer: How much better, in relation to its increased cost, is treatment C relative to treatment B?

THE "QUALITY PROBLEM" IN MEDICAL PRICE MEASURES (AND HOW MEDICAL OUTCOME MEASURES CAN HELP TO SOLVE IT)

The biggest problem in economic accounting for medical care is: How can we be sure that an improvement in medical care *is* counted as an increase in medical output and is *not* counted as medical inflation? What methods can be devised to allow for improvements in medical care in price indexes and output measures? This is the medical care version of the price index "quality problem" that has long been discussed in the context of other goods and services (see, for example, Griliches, 1971).

Four separate categories of price index "quality problem" can be distinguished. Sometimes clear distinctions have not been drawn among them in the medical economics literature.

(1) Changes in treatment methods that are not recorded by the statistical agency. For example, a heart by-pass may be employed in modern medical care of heart disease instead of medical management, which might have been the only possibility, or the predominant method, in the past; if the statistical agency fails to address the change in treatment, an error is introduced into the price index and into the deflated output measure for medical care. Unnoticed change in treatment method was certainly a problem with the old CPI procedure, in which the daily charge for a hospital room was collected, because in the old CPI procedure no consideration was given to the treatment for which the patient occupied the room. This case amounts to price comparisons that shift among the rows of table A, rather than price comparisons that are constructed from matched treatments (price comparisons within the rows). The new BLS procedures that track the cost of a specified medical case have undoubtedly greatly reduced or eliminated this problem.

(2) Un-noticed (by the statistical agency) improvement (or deterioration) of care, *within*

the rows of table A. One may control for the treatment, thus avoiding the gross
noncomparabilities inherent in case (1), but is a 1982 heart by-pass exactly the same as a 1997 by-
pass? The operation has become less invasive, learning by doing has perfected the techniques
(studies show that heart operations performed at hospitals, and done by surgeons, that do the
operation more frequently have more satisfactory outcomes than otherwise), and recovery times
have diminished. How does one handle drift in the effectiveness of what may be described, in
ordinary and in medical language, as the "same" treatment? Notice that this is not just a price
index problem: It arises in cost-effectiveness studies as well. If procedures are modified in ways
that were not incorporated into the original study, the effectiveness measure will be out of date,
and must be re-done.

Cases (1) and (2), above, are the traditional ones that have been the subjects of most past
discussion in the medical economics literature. Standard price index methodology restricts price
comparisons to *matched* medical treatments (comparisons within the rows of table A). Cases (1)
and (2) both amount to errors in holding the match constant.

It has been recognized more recently in the price index literature that the restriction to
within-row comparisons, though useful to prevent errors from gross noncomparabilities, may also
contribute its own problems. When new treatments are introduced and old ones disappear, those
entries and exits may imply changes in value--measured either in price or in quantity--that are not
equal to what one gets from "matched" comparisons within the rows. These are cases (3) and (4),
discussed next.

(3) New treatments are developed and introduced. The matched treatment method in
traditional price indexes ignores treatment C when it first is introduced in period 2 (and treatment
D, in period 3). This corresponds to what might be called "quality change outside the price index
sample." The matching methodology focuses attention on the match and getting the match right,
and not on changes that occur outside the set of matched price quotes. To put it another way, the
methodology maximizes the likelihood of collecting prices that are exact "constant quality"
matches, and minimizes encounters with cases of "quality change." However, ignoring C creates
potential bias in the price and output measures, even though the bias may be small if (as often
asserted) the quantity of the "new good" (q_{c2}) is inconsequential in the periods just after it is
introduced.

(4) The exit of old treatments, and their replacement in the CPI sample. In table A,
treatment A exits from the price index sample in period 4. Its disappearance *forces* a comparison
of some kind among the rows, a comparison between treatment A and some other treatment (C,
say) that will replace it in the sample. This might be called "quality change inside the price index
sample." A measure of relative effectiveness, or of willingness to pay--or a "quality adjustment,"
in usual price index terminology--is required to make this comparison.

Note that when treatment A disappears (period 3) it is not necessarily replaced in the
sample by the treatment that was newly introduced in period 3 (which is treatment D); the quality

problem in case (4) is logically separate from the one in case (3), because it is not necessarily a "new good" problem. But in any case, shifting the sample from A to C may be late, in the sense that the really significant change in treatments, and the really significant change in costs, might have been the introduction of treatment C some time earlier (case 3), well before the disappearance of treatment A forced a "quality adjustment" in the price index (case 4).

Both case (3) and case (4) require a quality adjustment. The next section discusses the use of medical outcome information to carry out adjustments in medical price indexes for improvements in medical treatments.

Using Medical Outcome Measures in Medical Care Price Indexes

When treatment A disappears and is replaced in the price index sample by treatment C (case 4), how much of its increased cost should be counted as a quality change? The entire amount (that is, $p_{c2} - p_{a2}$)? Some price index methods imply that answer, approximately, but the correct answer depends on the extent that treatment C is better than treatment A, not on the difference in their prices.

Or, considering case (3), what are the implications of the introduction of treatment C for the cost of medical care? Should the index be unaffected when treatment C is introduced? The answer depends on whether the higher cost of treatment C is or is not exactly equal to the value of the quality difference between C and the previously available treatments.

Recall from above that the cost-effectiveness ratio compares the cost difference and the effectiveness difference between two treatments (equation 4). We express the cost differences by the p_{it} terms of table A. The effectiveness differences are measured as QALY's.

Suppose that the cost-effectiveness ratio for treatment C compared with treatment A (or treatment B) equals unity. That means that treatment C's cost increase is just offset by the improvement in its effectiveness. No true price change takes place when C replaces A (case 4), or when C is added to existing treatments (case 3). The price index that measures the cost of medical care should not change in either case.

Suppose, on the other hand, that the cost-effectiveness ratio is less than unity--the cost increase when C is introduced is less than the improvement in effectiveness it brings about. This is the case that ought to prevail when a new, more effective treatment is actually adopted and comes into general use. The true cost of medical care has gone down in this case, and the price index should fall.

Now suppose the third possibility, that the cost-effectiveness ratio is greater than unity. Treatment C causes cost to rise more than effectiveness. Implementing treatment C in this circumstance is not a wise decision, but it could happen (and the frequent talk about technology-driven medical price inflation suggests a belief that it does happen). Because the cost of medical

care rises when C is introduced, the price index should also rise.

A medical outcome measure, such as QALY, provides the adjustment to apply to price indexes when a new treatment is introduced or when it is encountered in compiling the price index sample.[13] One's general presumption--and good social policy--is that new treatments will be introduced when they have low cost-effectiveness ratios, when the increment they make to the cost of care is lower than the increment they make to medical effectiveness. When that favorable condition prevails, *the traditional "matched treatment" price index methodology will miss part of the medical cost reduction associated with the introduction of a new medical procedure*; use of QALY as an adjustment will permit the price index to record the true cost decline. Conversely, if new medical techniques worsen the cost-effectiveness ratio, then matched treatment price index methodology will miss part of the increased cost of medical care; use of QALY brings this price increase into the price index.[14]

For those who are familiar with the usual price index literature on quality change, it will appear surprising that the "value" of quality change (the value of medical outcomes) did not come up. In the method for employing medical outcome information that is outlined in this section, I make no actual use of the V_t terms of table A, I do not need to put a price on the quality-adjusted life years.. This is not parallel with what is done for non-medical goods and services, and requires further discussion, which is the topic of the next section.

VALUE OF LIFE IN PRICE AND OUTPUT MEASURES

Cutler, McClellan, Newhouse and Remler (1996) calculated the number of life years extended for heart attack victims (their medical outcome measure), and then multiplied by a value of life estimate in computing their price index for heart attacks.

There is great reluctance to put a value on lives saved or extended. Statistical agency administrators have objected to this aspect of the Cutler, McClellan, Newhouse and Remler (1996) price index research. Gold, Siegel, Russell and Weinstein (1996, page 28) note that many users of cost-effectiveness studies are "uncomfortable about attaching dollar valuations to health outcomes such as life expectancy." They also note that because cost-effectiveness analysis "offers much of the same information" as studies that do value health outcomes, "the distinction may be

---

[13] A more complete discussion would cover several points: What would a consumer be willing to pay for a QALY, and the fact that it depends on the income distribution? Social valuation may differ from individual valuations: the individual wants "the best" but society may not have the resources to pay. The various concerns about willingness to pay are not unique to QALY, they apply to any quality adjustment for medical price indexes.

[14] For simplicity, I have ignored a number of complications that arise in case (3). Price index biases caused by item replacements can go either way (even when quality is improving), because the bias depends on the nature of the method used for item replacement and the paths of price change and quality improvement. See the extended discussion in Triplett (1997).

more important for the sake of appearance than for its practical consequences." This is essentially the point made by Phelps and Mushlin (1991): Cost-benefit analysis, which puts an explicit monetary valuation on medical outcomes, is nearly equivalent to cost-effectiveness analysis, which does not.

Responding to such concerns by saying that economists routinely value human life for other purposes (regulation analysis, suits for wrongful deaths and so forth) is no real answer. The ethical concerns are real, and even if we have to value lives for some purposes, the necessity for doing so does not make the choice a more comfortable one (Schelling, 1968).

Accordingly, there is great value to a methodology that does not require valuing human life, and the methodology of the preceding section does not. Explaining why it does not, and why "quality adjustments" for medical care are not parallel to those for non-medical goods and services, is the topic of this section.

Suppose one were told that a new car was improved by adding an air conditioning system, including it in the price when it was not before; one would not know how to adjust the automobile price index for the quality improvement unless one also knew how to put a value on the air conditioning system. Why does a comparable valuation requirement not hold for medical care?

One part of the answer lies in the metric for recent measures of medical outcomes, particularly QALY: Outcomes always have the same units. If the medical effectiveness measure is the expected number of years of life extended, for example, the metric is always years of life. If the outcome measure is QALY, then the outcome is always measured in a QALY unit. If one improvement in the procedures for an operation increases expected years of survival and another reduces the amount of pain during recovery, both gains in medical outcome would be valued on the common QALY scale.

On the car, the quality improvement might be ten horsepower one year, an air conditioning system in another, and extra cup holders in the third. We have to put a value on quality changes in cars partly because air conditioners and cup holders are "apples and oranges" that can only be combined with a monetary value. There is no natural scale to determine how much "more car per car" each of these improvements implies, so economic measurement use a "money metric" for combining dissimilar quality improvements to the car. Similarly, a computer is viewed in economic measurement as a bundle of speed and memory (Triplett, 1987); partly because computer speed and memory are measured in different units, changes in computer quality must be valued in money terms.

In contrast to the cases of cars and computers, the medical community has put a lot of effort into quantifying medical effectiveness, into devising a common scale that applies to different medical interventions. Years of life extended and pain relief have been put on a common scale in the QALY. Because of that common scale, economists do not have to value changes in medical effectiveness in the same way--or for the same reason--that they must put a value on changes in

automobile or computer quality. One can take the percentage improvement in QALY as a direct adjustment for the percentage change in cost that is associated with the change in medical technology.[15]

Another way to make the same point is to consider the V_t terms from table A. I specified earlier that V does not depend on the treatment, *I*. Indeed, when M_i is measured on the same scale across treatments, V cannot differ across treatments.   If the outcome measures M_i and M_j are both valued by V_t (multiplying both by V_t) to form a quality adjustment, M_iV_t/M_jV_t, then V_t falls out of the ratio. It cannot affect the quality adjustment for the medical price index.

A second part of the answer lies in the nature of that increment to effectiveness that is measured by QALY.   Suppose the increments to car quality were also always in the same units (one cup holder the first year, a second the next time, then a third). No economist, I suspect, would want to value the *nth* cup holder the same as the first, because of the familiar idea of diminishing marginal utility to increments of consumption. For this reason, in the car case one again needs to value the cup holders, indeed to find out the value of the *nth* holder.

The same proposition, however, does not hold for QALY. Here the increment is always in years. Incremental years of life are attained in the future, and future incremental years are valued lower because future years need to be discounted. In measuring QALY, future years are already discounted (Gold, Siegel, Russell and Weinstein, 1996). One cannot consume additional QALY's in timeless units, as one can the services of incremental cup-holders.

 Many economists would argue that an extra year of life is subject to diminishing marginal utility just like everything else, that the value of the 10th year of life extended is worth less to a 90-year old than to a 30-year old person. On the same line of reasoning, living 90 percent of one's days pain free, compared to 80 percent, might be worth less than an equivalent percentage improvement for someone who is in greater discomfort. If so, one might need to value changes in QALY. Those are important qualifications. Their empirical importance needs quantification.

A third reason adds to the two presented above: The QALY scale is constructed from information about preferences over different health states. It is noted in the medical outcomes literature (Gold, Siegel, Russell and Weinstein, 1996) that the QALY is a measure of preferences, that it contains rather strong assumptions that yield something close to a measure of the economist's idea of cardinal utility.[16] Having a measure of preferences across health states

---

[15] One minor reservation arises because price index quality adjustment would normally be done in ratio form, rather than in ratios of differences, as cost-effectiveness studies are usually calculated. That is, one would like the ratios p_{c2}/p_{b2} and M_c/M_b, and not the expressions of equation 4. I presume that when a cost effectiveness study is published, the basic information in it could be used to reconstruct the ratios, or that, alternatively, these calculating differences would make only a small effect on the actual calculation of the price index.

[16] It should be acknowledged, however, that this cardinal property of QALY has itself been criticized.

permits economists to do something in the case of health care accounting that cannot be done for automobile or computer accounting. For cars and computers, one does not have information on preferences for improvements in quality and therefore economists must find market values for them.

In summary, there are three reasons why QALY adjustments can be used directly in medical care price indexes, without the necessity for placing monetary values on them, as is necessary for quality adjustments in other goods and services. First, medical improvements are always measured in the same units. Second, incremental units are mostly attained in the future, which is already discounted in the measure. Third, the QALY measure itself has community preferences for states of health built into it. None of those conditions applies to price index quality adjustments that are available for most other goods and services. I conclude that access to QALY means that economists can incorporate improvements in medical care into price and output measures directly, without needing to put a value on life.

Some may contend, and have, that the QALY does not do what it purports to do, and these contentions are to be taken seriously. This is not the place to consider them. Rather, the present discussion concerns only a single problem: If the QALY does what it purports to do, what are the implications for using it in economic measurement?

Additionally, Mark Pauly points out, in his paper for this conference, that even if QALY measurements are extended and perfected, questions of accuracy remain. For medical decision making, it is only necessary to determine if the value of QALY (properly, the cost-effectiveness ratio in which QALY is the denominator) exceeds some bound. Price indexes require point estimates, which implies more accuracy. Pauly's point is correct; but the accuracy of quality adjustments for non-medical goods and services is not always high, and the choice is between using a number that is relevant (QALY), but imprecise, and using no number at all.

DO STATISTICAL AGENCIES ALREADY VALUE LIFE, IMPLICITLY?

As already noted, there is great reluctance to place a value on life, and statistical agency administrators have objected to price and output research in medical care that does so. The previous section showed that it may not be necessary to value life to use medical outcome measures. The present one responds to statistical agency concerns by explaining that price indexes (and therefore national accounts that use these price indexes as deflators) already value life *implicitly*.

Consider again the rows of Table A. Why do price index compilers only carry out comparisons *within* the rows? A simple answer is that moving down the columns introduces non-comparability--the content of the treatment in each row is not strictly comparable to that in another row. That answer is acceptable so far as it goes, but there is more to it.

For nonmedical goods and services, the price index literature often presents an explicit

justification for restricting price comparisons to those that move across the rows of table A. Any difference in price that is observed for different treatments at the same time (that is, price comparisons down the columns at the same time period or, e.g., p_{b2}/p_{a2}), equals the value of the quality difference--it is a measure of *willingness to pay* for the more expensive treatment by those who benefit from the more expensive treatment. This presumption arises in many forms, both explicit and implicit, which need not be explored here. It justifies, for example, the common practice of "linking" across rows (bringing one treatment into the price index as a substitute for another in a way that does not change the price index, which is standard price index "matched treatment" methodology for the "new" PPI and CPI medical care indexes).

What is the medical benefit? In the case of treatments for a heart attack, it seems wholly unreasonable to contend that the benefit is not increased survival probabilities or increased life expectancy. If it is, this implies that the price collecting agencies are *implicitly* assuming that price differences among medical treatments measure the value of improved life expectancy. It is true that they are seldom explicit about it. But price index procedures are implicitly valuing life whenever the differences in treatment that are controlled for in the price index (differences among the rows of table A) have survival difference implications.

An alternative or analogy may be helpful. Consider an improvement in an automobile that is included in the CPI price index for automobiles. Suppose the manufacturer includes on a new model car an air conditioning system. There is no natural way to answer the question: How much more car is this? One asks instead: What is the consumer willing to pay for this improvement? What charge is made for it, what does it sell for? What one incorporates into consumer price indexes, in general, is not a quantitative measure of the quality improvement itself, but rather a quantitative measure of the value of the improvement to the user.

In medical care price indexes, the situation is a bit different. Everyone seems uncomfortable with the question: What would the consumer be willing to pay for an additional year of life? So a price index procedure--or a cost-effectiveness procedure--that *seems* to require *explicitly* putting a monetary value on life extensions becomes controversial, essentially for ethical reasons. Yet, if another procedure values life implicitly, the same ethical reservations come into the picture. There is no reason for choosing the methodology where the ethical issue is less transparent over the other one where it is more transparent.

It is true that one might engender a lively discussion about what is being bought when the more expensive treatment is chosen, and also over who is doing the choosing. But that debate applies fully to the price indexes that are currently produced. It is not unique at all to the proposals to use QALY. If someone were to contend that consumers who purchase more expensive treatments do not do so to gain medical efficacy, then the same contention destroys the entire price index logic for refusing to judge treatments A and B (or C or D) as comparable for price index purposes and for refusing to make direct price comparisons between treatments A and B (or C or D). Traditional price index methodology already requires and makes a medical efficacy assumption--the assumption that the more expensive treatment provides some medical

benefit that is equal in value to the additional price charged for it.

CONCLUDING REMARKS

This paper has developed the relation between two bodies of research--price indexes and cost-effectiveness studies--that are important for accounting for health care output and prices. Something needs to be said about the context in which that relation is of interest.

One might use effectiveness measures as quality adjustments within price indexes, national accounts, and national health accounts, as they now exist. They could be applied, that is, to the four categories of price index "quality problem" discussed above. More interesting, however, are new possibilities for health accounting that are opened up by research on medical effectiveness.

The Production Boundary in an Accounting for Health: What Expenditures Are Deflated?

As noted above, traditional national accounts (and price indexes) place a restrictive boundary on the costs that are included in economic accounting for medical care. Conventional price indexes are defined only on the charges made by medical care providers, the market transactions designated by the p_{it} and q_{it} information in table A. The logic behind this restriction is apparent from equation (3): If a measure of output is to be formed by deflation, then the concept of "price" in the price index must match the measure of expenditure that is to be deflated. The expenditures that matter for conventional national accounts are monetary expenditures paid to medical care providers, not some wider concept of the full cost of illness or of treatment. Price indexes have been defined accordingly.

However, neither monetary expenditure nor the transaction price charged by the provider is a very good measure of the cost of illness or of the treatment of illness. Do we want to know the deflated value of expenditures on hospitals? Or do we want to compile an accounting for the total cost of illness? Of course, we want both. However, the usefulness of traditional national accounts for understanding the use of resources in health care, and for measuring health care output and productivity, is greatly limited by their restriction to market transactions .

Fortunately, broader accounts for medical costs already exist. These are sometimes called "burden of disease" studies. In place of the restricted concept of cost or expenditure in national accounts, the total cost of illness--direct and indirect--is estimated. This cost concept is closer to the numerator of cost-effectiveness studies than to the transactions price concept that is used in traditional national accounts.

The first government burden of disease study seems to be Rice (1966). Subsequently, these accounts have been improved and updated several times in the U.S. (an update to 1995 is Hodgson and Cohen). They have also now been constructed for a number of other countries, so there are not only limited time series for some countries, but also the beginning of a basis for

making international comparisons of costs.

So far, burden of disease studies have been conducted in constant dollars. Neither the total burden estimates, nor the various components that are also published, have been deflated by price indexes to obtain estimates in real terms. That is the next step in improved accounting for health care (see also Triplett, 1998).

The Classifications: What Prices Do We Want to Measure?

In national accounts, and national health accounts, expenditures are classified by provider of funds and by recipient of funds. The classification tells who provides the funds (e.g., Federal government) and who gets them (e.g., hospitals) , but it tells relatively little about what the funds are spent for.

Burden of disease studies classify expenditures into disease categories, not funding categories. Generally, they disaggregate to the level of ICD-9 chapters, or by disease categories within ICD-9 chapters. A burden of disease study tells how much is spent on, say, circulatory disease or on heart disease (or how much is spent with hospitals on heart disease). The burden of disease classification yields information that is a bit like the "product side" of conventional national accounts, which tell how much is spent on, say, consumption of restaurant meals.

The burden of disease classification system naturally suggests the question: What has happened to the cost of treating, e.g., heart attacks or strokes? The usual national accounting classification asks instead: What has happened to the cost of hospital services? The burden of disease classification is a much more natural one for integrating medical outcome measures into economic accounting for health care, because outcome measures are constructed for alternative treatments for a specified disease.

It is true that hospital services are made up of expenditures on diseases, so the two groupings of information are not antithetical to one another. One could, for example, combine Cutler, McClellan, Newhouse and Remler (1998) estimates with the BLS hospital PPI for circulatory disease. But "hospital services" is such a broad category; it focusses attention away from what is being treated and toward institutional groupings (type of hospital) that may be problematic for accounting for health care because treatment of disease too often crosses the boundaries between institutional providers. Additionally, physicians' services price indexes are not, for the most part, organized by disease, so the PPI accounting for health costs is not consistently by disease categories. Classification by disease focusses attention on the questions: What has happened to the price of treating heart attacks, on the quantity of the treatments and of their outcome? That is a more useful focus for improving our information on the real economic performance of the health care sector than is the institutional focus of the present system, with its emphasis on institutional providers of care, and on recipients of funds and the funding agencies. If adopted for economic accounts for health, the cost of disease focus would also have the advantage of setting up the accounts in a form that takes maximum advantage of the burgeoning

research on medical outcomes and cost-effectiveness.

Structuring accounts around the cost of disease has additional advantages in constructing accounts for health for countries that have public health systems. In those countries, prices charged by medical providers do not exist, at least for the bulk of medical care. A price index for the hospital industry makes no sense, nor does constructing hospital output by deflation methods. But data exist on the cost of disease and the cost of alternative treatments, on the numbers of treatments (the q_{it} terms of table A), and there is growing information on medical outcomes in all countries. There is also great interest in accounting for the output and productivity of medical care in several of those countries. One can conceive of doing this by constructing direct quantity measures for health care, using medical outcome measures to "adjust" treatment quantities, instead of treatment prices. The application of that is another paper.

Table A:

Treatments	Period 1	Period 2	Period 3
Treatment A	$p_{a1}q_{a1}M_aV_1$	$p_{a2}q_{a2}M_aV_2$	
Treatment B	$p_{b1}q_{b1}M_bV_1$	$p_{b2}q_{b2}M_bV_2$	$p_{b3}q_{b3}M_bV_3$
Treatment C		$p_{c2}q_{c2}M_cV_2$	$p_{c3}q_{c3}M_cV_3$
Treatment D			$p_{d3}q_{d3}M_dV_3$

References:

Bailar, John C., III, and Heather L. Gornik. 1997. Cancer undefeated. *New England Journal of Medicine* 336(22): 1569-1574.

Barzel, Yoram. 1969. Productivity and the price of medical services. *Journal of Political Economy* 77: 1014-1027.

Bresnahan, Timothy F., and Robert J. Gordon, eds. 1997. *The economics of new goods*. Studies in Income and Wealth, vol. 58. Chicago: University of Chicago Press.

Cardenas, Elaine. 1996. The CPI for hospital services: Concepts and procedures. *Monthly Labor Review* 120(7): 32-42.

Catron, Brian, and Bonnie Murphy. 1996. Hospital price inflation: What does the new PPI tell us? *Monthly Labor Review* 120(7): 24-31.

Cutler, David M., Mark McClellan, Joseph P. Newhouse, and Dahlia Remler. 1998. Are medical prices declining? *Quarterly Journal of Economics*, November.

Ehrhardt, Melvin E., Raymond M. Fish, and Linda M. Tambourine. 1993. *DRG 1993: Diagnosis related groupings with ICD-9-CM cross reference*. Los Angeles: Practice Management Information Corporation.

Feldstein, Martin S. 1969. Discussion. In Fuchs (1969).

Frank, Richard G., Ernst R. Berndt, and Susan Busch. Price indexes for the treatment of depression. Unpublished paper.

Fuchs, Victor R., ed. 1969. *Production and productivity in the service industries*. Studies in Income and Wealth, vol. 34. New York: Columbia University Press.

Garber, Alan M., and Charles E. Phelps. 1997. Economic foundations of cost-effectiveness analysis. *Journal of Health Economics* 16: 1-31.

Gold, Marthe R., Joanna E. Siegel, Louise B. Russell, and Milton C. Weinstein, eds. 1996. *Cost-effectiveness in health and medicine*. New York: Oxford University Press.

Griliches, Zvi, ed. 1971. *Price indexes and quality change: Studies in new methods of measurement*. Cambridge, Massachusetts: Harvard University Press.

------, ed. 1992. *Output measurement in the service sectors*. Studies in Income and Wealth, vol. 56. Chicago: University of Chicago Press.

Hodgson, Thomas A. and Alan J. Cohen. 1998. Medical Care Expenditures for Major Diseases, 1995. Unpublished paper, National Center for Health Statistics, Centers for Disease Control and Prevention.

Lazenby, Helen C., Katherine R. Levit, Daniel R. Waldo, Gerald S. Adler, Suzanne W. Letsch, and Cathy A. Cowan. 1992. National health accounts: Lessons from the U.S. experience. *Health Care Financing Review* 13(4): 89-103.

Meltzer, David. 1997. Accounting for future costs in medical cost-effectiveness analysis. *Journal of Health Economics* 16:33-64.

Moulton, Brent R. and Karin E. Moses. 1997. Addressing the Quality Change Issue in the Consumer Price Index. *Brookings Papers on Economic Activity* (No.1): 305-49.

Murray, Christopher J. L., and Alan D. Lopez, eds. 1996. *Global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. 2 vols. Cambridge, Massachusetts: Harvard University Press

Newhouse, Joseph P. 1989. Measuring Medical Prices and Understanding Their Effects--The Baxter Prize Address. *Journal of Health Administration Education* 7(1): 19-26.

Phelps, Charles E., and Alvin I. Mushlin. 1991. On the (near) equivalence of cost-effectiveness and cost-benefit analyses. *International Journal of Technology Assessment in Health Care* 7(1): 12-21.

Phelps and Weinstein, 1991

Puckett, Craig D., ed. 1997. *1997 annual hospital version, the educational annotation of ICD-9-CM*. 5th edition. 3 vols. Reno, Nevada: Channel Publishing, Ltd.

Reder, M. W. 1969. Some problems in the measurement of productivity in the medical care industry. In Fuchs (1969).

Rice, Dorothy P. 1966. *Estimating the cost of illness*. Health Economics Series, no. 6. Washington: U.S. Government Printing Office.

Rice, Dorothy P., and Loucele A. Horowitz. 1967. Trends in medical care prices. *Social Security Bulletin* 30(7): 13-28.

Scitovsky, Anne A. 1964. An index of the cost of medical care--A proposed new approach. In *The economics of health and medical care*. Conference on the Economics of Health and Medical Care, 1st, University of Michigan, 1962. Ann Arbor: University of Michigan.

------. 1967. Changes in the Costs of Treatment of Selected Illnesses, 1951-65. *American Economic Review* 57(5): 1182-1195.

Schelling, T. C. 1968. The life you save may be your own. *Problems in public expenditure analysis*. Studies of Government Finance. Washington: Brookings Institution.

Triplett, Jack E. 1989. Price and technological change in a capital good: A survey of research on computers. In *Technology and capital formation*, edited by Dale W. Jorgenson and Ralph Landau. Cambridge, Massachusetts: MIT Press.

------. 1997. Measuring consumption: The post-1973 slowdown and the research issues. *Review--Federal Reserve Bank of St. Louis* 79(3): 9-42.

------. 1998. What's Different About Health? Human Repair and Car Repair in National Accounts. Paper presented at the NBER-Conference on Research in Income and Wealth conference on Medical Care, Bethesda, Md. June 12-13, 1998.

U.S. Congress. 1961. Joint Economic Committee, Subcommittee on Economic Statistics. *Government Price Statistics: Hearings before the Subcommittee on Economic Statistics of the Joint Economic Committee*, 87th Cong., 1st sess., 1,2,3,4, and 5 May, part 2.

U.S. Department of Commerce. Economics and Statistics Administration and Bureau of Economic Analysis. 1996. *Survey of Current Business* 76 (1-2).

Weinstein, Milton C., and Willard G. Manning, Jr. 1997. Theoretical issues in cost-effectiveness analysis. *Journal of Health Economics* 16:121-128.