This revision: Sept 15, 2009

Output and Productivity Measures for Medical Care and Education

Jack E. Triplett

Brookings Institution

Prepared for the University of Rome Workshop on
Medical Care and Education Productivity:
Rome, April 2009

Of services output generally, "…it is not exactly clear what is being transacted, what is the output and what services correspond to the payments made to their providers (Griliches, 1992). Medical care and education are services sectors where output measurement is ambiguous or difficult, both for conceptual and empirical reasons. Much recent progress has been made on methods for measuring medical care output. Far less progress is evident on education output, for reasons I will explain in section II of this paper.

In the U.S., concern for improving data on medical care output and productivity extends well beyond the national accountants. A National Academy of Sciences panel is currently reviewing measures of medical care output and accounts for medical care and for health.[1] In the U.K. as well, concern for these topics extended outside the rarified circle of national accountants: Is there any other country where health system productivity figured in Parliamentary debate?[2]

In the last section of this paper, I contend that measuring the output of medical care should be a concern, not just for the national accountants, but also for politicians and managers of health care systems: The *data needed* by national accountants to measure health care output appropriately are data that are also necessary for analysis and understanding of advancing health care costs and for improving health care delivery. The public policy analysis need is a more urgent one than improving the national accounts, but fortunately the two agendas coincide.

I. The Output of Medical Care

A dozen years ago, I wrote: "What's Different About Health? Human Repair and Car Repair in National Accounts" (published as Triplett, 2001). In that paper, I proposed that national accountants think about medical care output by analogy with

[1] The National Academy of Sciences was established by President Abraham Lincoln to provide scientific advise to the government. I am a member of the NAS panel, but this paper incorporates solely my own views and not necessarily those of any other panel member.

[2] The debate led ultimately to the Atkinson Report (Atkinson, et al. 2005).

the way they measure the output of car repair: One can measure a human repair as a completed episode of treatment for a disease.

To implement the human repair proposal, the first step involves grouping health care expenditures by a disease classification system, such as the World Health Organization's (WHO) International Classification of Diseases (ICD). The second step is to estimate quantity measures[3] of treatments by disease categories, e.g., the quantity of medical services for treating heart attacks or broken legs. As well, there are the associated measures of price or cost change for these medical services. As with other commodities in national accounts, the product of the change in medical care quantities and the change in prices equals the change in medical expenditures.[4]

The human repair proposal was quite controversial at the time, partly because many health care policy analysts viewed the "units" of medical care to be the hospital-doctor's office-drug company trichotomy traditional to the national health accounts. In the human repair view, the hospital is the industry and treatments for various diseases are the products that the industry produces, though in human repair several industries may combine to produce a human repair unit, which poses some problems. The human repair proposal had antecedents in economics going back at least to Anne Scitovsky's (1964, 1967) measures of the cost of treating diseases, and some contemporary price index research used an identical concept, such as Cutler, et al (2001). Scitovsky's work, however, had been viewed for some reason as relevant only to price indexes and in any case it had mostly and unjustly been ignored.

Now, a growing consensus has endorsed the human repair (cost of treating disease) view, not so much because of my paper but because it is a natural way for an economist to think about the subject. The consensus extends to statistical agencies in the U.S. and Europe, so the conceptual approach to medical care is mostly harmonized internationally, at least at a gross level. Differences in implementation reflect mainly differences in our medical care systems. In the U.S., medical care is (mostly) privately delivered, and prices for medical transactions are available and collectible. In consequence, in the U.S. national accounts the quantities of medical care services are estimated by deflation (dividing the change in expenditures by the price index). Implementing the human repair model in the U.S. thus requires medical care price indexes as deflators. In European countries, governments provide most of medical care and prices are not relevant for most of medical care delivery. Accordingly, in European national accounts, the quantity of medical care services is estimated by a direct quantity index—except for those

---

[3] In the System of National Accounts, these quantity measures are called "volume" measures, for reasons that make no sense in English.

[4] For this paper, I neglect the distinction between the prices of medical care services and their costs, because *for the issues I address* it does not matter, conceptually, whether a computation is based on prices or costs. However, for other questions it matters very much.

countries where output is still regrettably estimated by use of medical care inputs as proxies for a proper output measure.

International statistical practice may converge. First, prices have decreasing relevance in the U.S., and it is not clear how long the U.S. delivery system will continue to differ from those of other industrialized countries. Second, private provision of medical care services is increasing in some countries where the bulk of medical care is governmentally provided. In any case, the major measurement issues are the same, whether the quantity measure is estimated through deflation or by a direct quantity index.

A. Measurement Issues. To estimate medical care output under the human repair model, the unit is an episode of treatment for a disease, a heart attack, for example. One begins by estimating medical care expenditures according to a disease classification. The ICD is not only readily available and in use for other purposes but conforms well to what the national accountants—and the national health accounts compilers—need. A closely related alternative is a Diagnostic Related Group (DRG) system. DRGs, which do differ internationally, are nonetheless usually consistent with the ICD at the chapter level, and with the present stage of development, the chapter level is the lowest practical disaggregation, for the most part. In some countries the first step is likely to proceed by disaggregating the aggregate medical care expenditures presently in the national accounts or in the country's national health accounts, though in other cases disease-classified medical care expenditures may be built up from the bottom.

There are to be sure difficulties in obtaining the data and carrying out the disaggregation, and alternative methods have been proposed for it. But I will by-pass the details, on the grounds that cost-of-illness accounts have been done in many countries (though seldom constructed as time series), so much experience exists on the first stage of implementing the human repair model. I focus instead on a few major methodological issues that have arisen in recent discussions.

(1) ***Grouping provider data into episodes***. If collection is from medical care providers, which is usually the case for the data used in the production or industry accounts portions of national accounts, one must link expenditures on, e.g., hospitals, doctors' offices, clinics, laboratories and so forth, into expenditures on a disease. A heart attack involves treatments from various institutional units. Thus, the national aggregates for institutional expenditures (hospitals and so forth) must be disaggregated by disease and the data must then be re-aggregated into expenditures on heart attacks (or more likely in the immediate future, into expenditures on circulatory diseases, the relevant chapter of the ICD, though more detail is of course always preferable to less). I do not see that this poses any major issues, though there has been a bit of confusion about it.

Sometimes one hears the contention that data on an episode of disease can only be collected from the patient and not from providers. At the micro level, this contention has some merit, and if the objective is to produce microdata for the analysis of patients, then obviously data must be collected from the patient, or from claims data or utilization data, where it exists. But to produce medical care output measures in national accounts, expenditures data must be aggregated. Whether one starts with patient data or provider data merely implies a different order of summation: With patient data, expenditures have already been summed over providers to get total expenditures for the patient's episode; then, summation over patients yields aggregate expenditures on heart attacks. Provider data will already be summed over patients by their very nature; the aggregate expenditure on heart attacks is found by summing over providers in the second step (rather than the first, as with patient data). Provider data do pose some issues of their own (see below), but it is not necessary to start with patient data to estimate the costs of episodes of disease.

(2) **Co-morbidity.** Illnesses frequently occur together. Can the cost of an episode of illness be separated into the costs associated with separate illnesses that may each belong in different ICD chapters? [5]

There is no question that co-morbidity presents problems for estimating expenditures by disease. Regrettably, this undeniable difficulty has been elevated in some discussions (or by some objectors) into some sort of empirical impossibility theorem.

Methods for dealing with co-morbidities can be found. Indeed, the difficulty has been confronted in past estimates of cost-of-illness accounts—see, for example, Hodgson and Cohen (1999), who allocated 87 percent of U.S medical care expenditures into ICD chapters (most of the unallocated amounts were in drugs—where does aspirin go?—and in non-structured miscellaneous categories such as "other professional services" and "other personal health care," where the allocation difficulty has nothing to do with co-morbidities). The issue is not whether it can be done. The issue is finding ways to improve what has been done in the past.

Cutler and Rosen (2007) have been experimenting, using patient data, with a regression method that allocates a patient's total annual medical expenditures among all the diagnoses (whether co-morbid or not) that are reported for each patient. The regression coefficients allocate the co-morbid costs. Another method is to first attribute to each of the co-morbid diseases the cost of treating that disease in non co-morbid cases; the total of these allocations will either exceed the actual cost of treatment (implying economies of scope in medical care) or fall short. The allocated total is then compressed or elevated to match the actual data. No doubt other methods will be devised as research proceeds.

---

[5] Co-morbidities that belong to the same ICD chapter (an aneurysm and an irregular heart rhythm) do not pose problems, so long as aggregation is at the chapter level. This itself reduces the incidence of the co-morbidity problem, though obviously more detail is usually desirable. Hence, there will be an inevitable conflict between the demand for detail in the data and the ability to produce accurate and meaningful detail.

The existence and prevalence of co-morbidities is not a reason for refusing to estimate the costs of illnesses.

(3) The prototypical cataract surgery problem. In a widely cited paper, Shapiro, Shapiro and Wilcox (2001) studied the evolution of cataract surgery, which was once an expensive in-hospital procedure involving many days of recovery. Now, it is normally an outpatient procedure, performed in a clinic or a doctor's office. Not only is the cost now much lower, but from the patient's point of view the modern operation produces much less disutility during recovery than was formerly the case, so the outcome of the cheaper procedure is at least as good as the older, more expensive one.

If data collection is from the provider, the cost saving from the medical innovation will normally be missed, because a statistical agency does not usually compare costs of operations across providers. The agency's hospital collection will gather data on the in-hospital operation until it disappears; collection from the clinic or doctor's office will gather data on the new operation after the point when it is first observed. Each collection will track the change in cost *within* each institutional unit, but will miss the change in cost that occurs when the procedure shifts *across* medical institutions.[6]

Alternatively, suppose the output measure is estimated through a quantity index, aggregated by price or cost weights (usually the case). Then, the shift from an expensive treatment (high weight) to a less expensive treatment (lower weight) understates the growth in output after the shift, because the two procedures have equal outcomes.

It has sometimes been thought that collecting data from patients would obviate the cataract surgery problem. If a cataract surgery is a homogeneous operation (that is, if patients present the same or nearly the same diagnoses and severities), then one could just compare an operation received by the patient who patronized the hospital with an operation received by a (different) patient in the out-patient clinic, provided of course we also knew that the medical outcomes of the in-hospital and outpatient procedures were the same. Merely to state those conditions indicates the potential hazards of direct comparisons of what may be dissimilar patient profiles and medical outcomes, and the likelihood of error.

Note, additionally, that if one were willing to accept the restrictive conditions of the preceding paragraph when employing patient data, or if one knew they were valid empirical approximations, one could also justify comparing across providers on the same assumptions. Thus, there is no real advantage *in this case* to moving

[6] A technical matter not pursued here: The shift of treatments between providers is not the standard "substitution bias" that is familiar from the price index literature, and which is usually resolved by use of an appropriate index number formula. It is more nearly analogous to the problem arising in the CPI from the spread of discount stores, where the same product can be bought for less than in conventional retail outlets.

from the traditional provider collection favored by statistical agencies to the patient or claims data that have typically been employed in research projects.

The cataract surgery example is compelling. It is less clear how important it is, quantitatively. To my knowledge, only one relevant study exists that attempts such a review across the whole of medical care: Aizcorbe and Nestoriak (2008). They present data indicating that the cataract surgery type of problem is not so serious, now, in the ICD chapter that contains cataract surgery (the shift in those treatments is history), but this general type of problem is more severe in some other ICD chapters.[7] Research on this general problem is in its infancy.

(4) The Quality Change Problem. Medical care output measures are applications of the general problem of estimating price and quantity indexes. It is commonplace that changes in the qualities of goods and services pose the most serious difficulties in index number measurements.

Index numbers are based on samples of goods (or services) that are matched in two periods—they present the change in the price (or quantity) of the same group of goods or services. If the match is not exact, if a product or service included in the sample changes, then part of the price change (in the price index case) will reflect price variation that is attributable to change in quality. An equivalent statement applies to direct quantity indexes. In medical care, the quality change problem often takes the form of changes or improvements in treatments, though changes in, e.g., hospital amenities, waiting times, and so forth can also occur and bias the index number in the same way.

In the U.S. Producer Price Index (PPI), the Bureau of Labor Statistics (BLS) uses the change in cost as a quality adjustment when changed treatments are encountered in PPI medical care components. For example, suppose that in the initial period a hospital provided a figure of $3000 as the cost of the existing treatment for a diagnosis. In the subsequent period, an improved treatment is used for this diagnosis, and the hospital reports $3600 for it. The hospital, however, also reports that the cost difference between the two treatments (considered at the same time, that is, in a period when both were in use) was $500. In this example, the PPI would record $100—not the full $600—as the price increase for treating the diagnosis.

This BLS quality adjustment method corresponds to the dictates of the theory of the output price index (Fisher and Shell, 1972, and Triplett, 1983). Directly estimated output indexes would use the same method.

[7] They reach their result by comparing two calculations, both of which are unit value indexes. Alternative interpretations of their comparisons are possible, so this may qualify the statement in the text. The matter is too complex to be explored here. As noted, it is the only study and it is an interesting one, so it deserves attention.

However, the assumptions of the theory are very restrictive because it is a theory of a constant input, *fixed technology* price index.[8] Some medical care improvements can be so described. For example, greater hospital resources put into cleaning and sanitation in order to reduce in-hospital infections fits the fixed technology model, as probably do efforts to reduce waiting times.

In medical care, however, many treatment changes involve new technologies. The cataract surgery that moved to a sutureless procedure (Shapiro, Shapiro and Wilcox, 2001) was not a constant technology innovation—it not only improved the treatment but also reduced its cost. The costs of the old and new operations correspond to points on cost functions from two different technologies; the theory does not say that ratios of these costs provide appropriate quality adjustments.

Thus, there are no alternatives to medical outcome measures, even though outcomes do not, strictly speaking, concord with the theory that underlies the measurement of output. Letting η_{21} be the quality-adjusted price measure, p_2 and p_1 the prices in periods 2 and 1, respectively, and m_a and m_b the medical outcome measures corresponding to treatments a (the old one, used in period 1) and b, the outcomes-adjusted price measure is:

$$(1)\ \eta_{21} = p_2 / (p_1 (m_b / m_a))$$

Expenditures on this treatment are $p_1 q_1$ and $p_2 q_2$ in the two periods, so that the deflated quantity change in treatments (λ_{21}) is:

$$(2)\ \lambda_{21} = (q_2 (m_b / m_a)) / q_1$$

Where the direct quantity index method is used, the same result would be obtained through an adjustment to the quantities, equivalent to equation (2). Note that the adjustment makes the improvement in outcomes part of the quantity change in medical care.

Few outcome measures for treatments of disease are available, so this is not an easy solution to implement. The outcome measure should specify the consequences of a change in treatment, not a treatment change combined with something else. Heidenreich and McClellan (2001) link outcome measures to clinical trials of treatment changes for heart attacks. A clinical trial does link securely the outcome and the treatment. Some estimates of medical outcomes may not.

B. Treatments or Outcomes? The human repair model implies that the output of medical care consists of treatments. Improvements in treatments are allowed for by adjustment by medical outcome measures, so we can think of medical care output as the quantity of treatments augmented by medical outcomes, as in equation (2), above.

---

[8] See Fisher and Shell (1972) and also Archibald (1977).

Why not by-pass the treatments and measure output by medical outcomes? This has been proposed by, among others, Dawson, et al. (2005). Under this proposal, the output of heart attacks would be, e.g., a weighted index of the Quality Adjusted Life Years (QALY) gained from heart attack treatments. Especially when interventions are not making positive incremental contribution to health (misprescriptions and misdiagnoses, for example), bypassing the treatment measure approximates more nearly, the proponents contend, the medical care system's contribution to health.

The distinction between output and welfare figures prominently in the national accounts literature. Output measures and welfare measures are both useful, but they are different measures. In the present case, the welfare measure is health. Medical care is the output of the medical care industries; it is an input to the production of health (and it is well known that medical care is only one of the inputs).

Consider the intriguing question: How much does medical care contribute to improving health? To answer this question, it will be necessary to model as well the contributions from other inputs that contribute to health, but leave this aside for the present, and assume that we can estimate the incremental contribution of the medical care input to the production of health.

For this purpose, the data must be constructed so that it is possible in principle for the change in medical care to differ from the change in health. If the medical care sector were measured by health outcomes, there can be no difference, by construction of the data. Thus, even if we want medical care data for the purpose of estimating the sector's contribution to health, we need treatments data, not medical outcomes. [9]

Additionally, output data are useful in their own right. By an output standard, treatments, whether appropriate or not, are still produced in the medical care sector; treatments, not health, absorb resources in the sector. By this (conventional) way of looking at output, they are outputs of the medical care sector. We do not subtract botched and inappropriate car repairs from GDP, even though they hardly contribute to consumers' welfare.

I understand the old adage that a measure that is "approximately right" is preferable to one that is "precisely wrong." But trying to put elements of the welfare measure into the output measure produces a muddled measure, not one that is approximately right. The history of economic measurement demonstrates that muddled concepts produce much confusion, not more useful measures.

---

[9] To avoid confusion, it is not inconsistent to make the quality adjustment for a changed treatment depend on the *ratio* of medical outcomes for the new and old treatments (though it is to an extent conceptually somewhat awkward, as noted above).

Another reason is compelling: An output measure that is based on disease treatments (even with medical outcome measures as quality adjustments) is grounded on a more precise statistic than an output measure that is based entirely on medical outcome measures. Information on expenditure by disease, on numbers of treatments by disease, and even on health care prices by disease is further developed than are medical outcome measures. Treatment information is inherently more concrete and therefore more precisely measured information. For a medical care output measure to attain public confidence (recall those debates in Parliament) it needs to be seen as transparent, at least in relation to other comparable economic measurements. Measuring medical care output by treatments is not that different from the way we measure car repair in national accounts (Triplett, 2001), and can readily be understood within the usual framework of economic statistics. In contrast, even health economics professionals raise difficulties, both conceptual and practical, with existing medical outcomes measures (Meltzer, 2001). A sound measurement principle is to minimize the use of undeveloped and potentially controversial measures, using them only where they are necessary and not where more straightforward alternatives exist.

II. The Output of Education

Measuring medical care output is difficult. Measuring the output of education is *really* hard.

The fundamental difficulty in education has little to do with test scores, class sizes and similar attributes that have figured so intensively in the discussion so far, though those measurement problems deserve the attention they are getting. More crucially, the output of educational establishments is difficult to measure *because they are multi-product firms*. They do not produce *only* education, they produce other things as well.

The New York Times (21 April, 2009, page A12) reported that between 1987 and 2007 support staffs in U.S. higher education establishments doubled, when teaching staffs grew only about 50 percent. Though the *Times* predictably referred to “administrative bloat,” a more cogent appraisal came from labor economist Robert Bennett, who was quoted as follows:

> “Universities and colleges are catering more to students, trying to make college a lifestyle, not just people getting an education. There’s more social programs, more athletics, more trainers, more sustainable environmental programs.”

Growth in non-teaching employment is usually viewed by researchers (who, after all, are often teaching staff) as evidence either that administrators do not know what they are doing, or of self-aggrandizing behavior by non-academics. I suggest, instead, that the administrators likely do know what they are doing (though they may articulate it imperfectly for our purposes): They are attending to the non-educational outputs of their establishments.

Moreover, the evidence cited in *The New York Times* indicates that, rather than being a minor part of the activities of the educational firm, these non-education activities are large and growing at substantially higher rates than the traditional education function. Non-educational output growth in U.S. higher education establishments has generated *more than half of employment growth* in these establishments over the past twenty years. We cannot expect to measure adequately the productivity of educational establishments by taking, as the numerator, some measure of students (adjusted by standardized scores and so forth) over a denominator that consists of all university employment. That clearly biases the productivity measure downward, owing to omitting the output being produced by the employees whose employments have the greatest growth rates. No wonder, then, that estimated productivity growth rates for education are negative.[10]

"Purging" the employment data of non-educational employment is no solution. Ehrenberg (200x), reporting on his own experience as an educational administrator, cites many instances where educational decisions ran afoul of non-educational objectives. From a decision-making point of view, if not from technical properties of the production function, these are joint outputs, just as demonstrably as the classic economic example of beef and hides. The non-educational outputs may be "secondary products," in the language of industry classifications, but the resources that produce them are inherently co-mingled with the resources that produce education outputs.

The only way to get an adequate measure of the productivity of educational establishments is to measure their non-education outputs and combine them with the education they produce. That is a huge job for two reasons:

- Unfortunately for education establishments, the product lines into which they have diversified (entertainment, hotels and eating places, personal services) all have very low or declining industry productivity growth rates (Triplett and Bosworth, 2004, pages 18-19).
- These other outputs are also hard to measure: How does one compare last year's hit movie with the current one?

One might contend (hope?) that in other countries, and in primary and secondary education in the U.S., a smaller proportion of educational establishments' growth occurs in non-educational output. After all, diversification into minor league

---

[10] Triplett and Bosworth (1994) reported negative multifactor productivity growth in the U.S. educational services industry (which does not include public education) over the period 1987-2001, and negative labor productivity over part of the period (Tables 2-4A and 2-4B, pages 18-19; additionally, they reported negative labor productivity in primary and secondary and in higher education (both of which, using a different data source, covered public and private education) over 1980-2000 (Tables 9-5 and 9-6, pages 270-711). In all of these estimates, the output measure was based on students, and did not include the non-educational outputs. Significantly, the industry's measured productivity performance worsened over time in Triplett and Bosworth's estimates, which is consistent with the trends reported in the *New York Times* article.

professional sports seems unique to U.S. higher education. Yet, anecdotes suggest similar joint-output measurement problems exist, differing perhaps in degree (or lagging behind U.S. higher education). For example, resources to meet socialization and social goals and provision of advisory functions are a growing in secondary education, particularly.

We are still debating how to measure effectively the educational output of education, and little consensus has emerged. Focus on the other outputs may seem to increase the difficulty of the task, and it probably does. But by ignoring the other outputs, we are imposing too stringent a standard for the education output measure, because output measurement improvements tend to be judged by whether they reduce those improbable negative productivity growth rates. The negative growth rates in education industries are, at least in part, signals that we have omitted major parts of the output of educational establishments—those parts that are not education.

III. Conclusion

Better measures of medical care and education outputs would be of great value to economists. Likely, educational administrators and national education departments would find improvements in education output measures interesting, but not vital to administration and policy concerns. They are probably right.

Many health department officials seem to have the same view of efforts to improve medical output measures. They are wrong.

Nearly all industrialized countries are concerned with advancing medical care expenditures. Yet, they have very little information on what their medical expenditures buy. The National Health Accounts produced in many countries have become nearly irrelevant to current policy debates on medical care costs because they show only who provides the money for health care (consumers, governments, insurance companies) and who gets it (hospitals, doctors, pharmaceutical companies). They do not show what is bought for health care expenditures—treatments for disease.

It is naïve to think that expenditures across all types of disease are growing at the same rate: Are expenditures growing more rapidly for treatments of circulatory diseases, or for mental health care, or for cancer? How many countries have this information? How can one make sensible policy decisions on medical expenditures without knowing where the increases originate?

And of course, just knowing which disease categories have advancing expenditures is not enough: We then need to know whether expenditures in the more rapidly increasing sectors are growing because of rapidly increasing treatment costs or rapidly increasing numbers of treatments. Again, it is naïve to think that treatment costs are growing at the same rate in all disease categories. In the U.S. Producer Price Indexes for hospitals (which are published by ICD categories), the inflation rate for blood and blood forming organ diseases has been *two and a half times* the inflation rate for treating infectious diseases. These are hospital costs, unfortunately, not total costs of treating the disease, but they indicate strongly that the costs of treating different diseases are not advancing at the same rate. How can anyone design intelligent policy on advancing medical care expenditures without knowing what is going on with costs? And why? Once again, information on price or cost growth by disease is sadly not available in most countries—including the United States—that have had debates about medical care expenditure containment.

If expenditure growth is contained, is the effect to reduce medical care inflation? Or is it, instead, to reduce the growth of medical care services? We don't really know, because we do not have information about the quantity trends in medical services, by disease. Only after determining this can we move to the next stage and ask whether the growth in the number of medical care services is worth it, a question that is too often "answered" with a minimal amount of relevant data.

The data that national accountants are seeking to improve medical care output measures is exactly the policy-relevant data that are needed for the analysis of national health care policy questions. Filling in the database for the analysis of medical care productivity deserves high priority. It deserves high priority not so much because productivity analysis is necessarily that important, but because vital questions of health care policy demand exactly the same data.

References:

Aizcorbe, Ana and Nicole Nestoriak. 2008. The Importance of Pricing the Bundle of Treatments. Unpublished paper. Washington, D.C., U.S. Bureau of Economic Analysis. July.

Archibald, Robert B. 1977. "On the Theory of Industrial Price Measurement: Output Price Indexes." *Annals of Economic and Social Measurement*, 6(1) (Winter), pp. 57-72.

Atkinson, Tony. 2005. *Atkinson Review: Final Report: Measurement of Government Output and Productivity for the National Accounts*. Palgrave McMillan.

Cutler, David M. and Ernst R. Berndt, eds. 2001. *Medical Care Output and Productivity*. Studies in Income and Wealth, Vol. 62. The University of Chicago Press for the National Bureau of Economic Research.

Cutler, David M., Mark McClellan, Joseph P. Newhouse, and Dahlia Remler. 2001. Pricing Heart Attack Treatments. In Cutler and Berndt, eds., 2001.

Cutler, David M. and Allison B. Rosen. 2007. Measuring Medical Care Productivity: A Proposal for U.S. National Health Accounts. *Survey of Current Business*, June, pp. 54-58.

Dawson, Diane, Hugh Gravelle, Mary O'Mahoney, Andrew Street, Martin Weale, Adriana Castelli, Rowena Jacobs, Paul Kind, Pete Loveridge, Stephen Martin, Philip Stevens, and Lucy Stokes. 2005. *Developing New Approaches to Measuring NHS Outputs and Productivity: Final Report.* York and London: Centre for Health Economics, University of York and the National Institute for Social and Economic Research.

Ehrenberg, Ronald G. 1999. Adam Smith Goes to College: An Economist Becomes an Academic Administrator. Journal of Economic Perspectives, Winter, Vol 13, pp. 99-116

Fisher, Franklin and Karl Shell. 1972. *The Economic Theory of Price Indices: Two Essays on the Effects of Taste, Quality, and Technological Change*. New York: Academic Press.

Griliches, Zvi, ed. 1992. *Output Measurement in the Service Sectors.* National Bureau of Economic Research. Studies in Income and Wealth. Volume 56. Chicago and London: University of Chicago Press.

Heidenreich, Paul, and Mark McClellan. 2001. Trends in Heart Attack Treatments and Outcomes, 1975-1998: Literature Review and Synthesis. In Cutler and Berndt, eds., 2001.

Hodgson, Thomas A. and Alan J. Cohen. 1999. Medical Care Expenditures for Major Diseases, 1995. Health Care Financing Review, Winter,

Meltzer, David. 2001. Theoretical Foundations of Medical Cost Effectiveness Analysis: Implications for the Measurement of Benefits and Costs of Medical Interventions. In David M. Cutler and Ernst R. Berndt, eds., 2001.

Scitovsky, Anne A. 1964. "An Index of the Cost of Medical Care--A Proposed New Approach." In Solomon J. Axelrod, ed., *The Economics of Health and Medical Care*. Ann Arbor: Bureau of Public Health Economics, University of Michigan.

Scitovsky, Anne A. 1967. Changes in the costs of treatment of selected illnesses, 1951-65. *American Economic Review* 57(5): 1182-1195.

Shapiro, Irving, Matthew D. Shapiro and David W. Wilcox. 2001. Measuring the Value of Cataract Surgery. In Cutler and Berndt, eds., 2001.

Triplett, Jack E. 1983. Concepts of Quality in Input and Output Price Measures: A Resolution of the User Value-Resource Cost Debate. In Murray F. Foss, ed. *The U.S. National Income and Product Accounts: Selected Topics*. Conference on Research in Income and Wealth: Studies in Income and Wealth, Volume 47. University of Chicago Press for the National Bureau of Economic Research, 1983: 296-311.

Triplett, Jack E. 2001. “What’s Different About Health? Human Repair and Car Repair in National Accounts and in National Health Accounts.” In David Cutler and Ernst R. Berndt, eds., 2001.

Triplett, Jack E. and Barry P. Bosworth. 2004. *Productivity in the U.S. Services Sector: New Sources of Economic Growth*. Brookings Institution Press.

World Health Organization.