

NBER WORKING PAPER SERIES

MEDICAL CARE PRICE INDICES: PROBLEMS AND OPPORTUNITIES
THE CHUNG-HUA LECTURES

Joseph P. Newhouse

Working Paper 8168
<http://www.nber.org/papers/w8168>NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2001

Kennedy School of Government, Harvard University; Department of Health Care Policy, Harvard Medical School; Department of Health Policy and Management, Harvard School of Public Health; National Bureau of Economic Research. I thank Ernie Berndt, David Cutler, Dennis Fixler, Richard Frank, Clint McCully, Bonnie Murphy, and Jack Triplett for comments on a preliminary draft; they are not responsible for any remaining misadventures. I also thank Chapin White for research assistance, and the Alfred P. Sloan Foundation and the Hans Sigrist Foundation for support. The views expressed herein are those of the author and not necessarily those of the National Bureau of Economic Research.

© 2001 by Joseph P. Newhouse. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Medical Care Price Indices: Problems and Opportunities
The Chung-Hua Lectures
Joseph P. Newhouse
NBER Working Paper No. 8168
March 2001
JEL No. I1, O3

ABSTRACT

These Chung-Hua Lectures, given at the Academia Sinica in Taiwan in December 2000, summarize work that has been done by myself and others on biases in medical care price indices. I begin by reviewing various uses of price indices and therefore why biases in the overall indices - and changes in those biases - matter. I then describe briefly the assumptions and theory underlying the official price indices. I next turn to the problems of measuring medical prices, assuming the basic applicability of the theory upon which the official indices are based. Finally I take up the potential inapplicability of the assumptions made by that theory and the resulting issues for measuring medical price changes. I describe an alternative theory and its implications for the measurement of medical prices. I conclude that the biases in the official medical care index, while substantially reduced by recent improvements, likely remain substantial enough to affect the overall official indices in the United States, especially the GDP deflator, where the weight of medical care is around 13 percent.

Joseph P. Newhouse
Division of Health Policy Research and Education
Harvard University
180 Longwood Avenue
Boston, MA 02115
and NBER
newhouse@hcp.med.harvard.edu

I wish to devote these lectures to a topic that might strike many of you as somewhat arcane, namely medical care price indices. I shall try to convince you that there are important biases in the measurement of medical price changes and that these biases are large enough to have measurable consequences for the official overall price indices such as the GDP deflator and the all-items Consumer Price Index (CPI).¹ Because of the size of its medical care sector, the overall indices are probably most affected in the United States. I know the issues with the American indices much better than the issues elsewhere, so my examples will come from the American indices. Nonetheless, I suspect that many of the problems I discuss are found in every developed country.

Some of the problems with the medical care price indices are well understood conceptually, but persist because of difficult or costly issues in implementing the conceptually correct solution. Other problems arise because the theory and assumptions on which the official indices rely are violated in the medical care context. Although the violations of the assumptions are not limited to the medical care sector, that sector presents particularly good examples of them. I suggest an alternative theory that is relevant for the medical care sector and may be adaptable for other sectors as well.

From a more practical point of view, there are opportunities to improve the indices in important ways. Indeed, the agencies charged with responsibility for the American indices have in recent years substantially improved the accuracy of the medical care price indices and deserve commendation for doing so ((Cardenas, 1996), (Catron and Murphy, 1996)). Nonetheless, as I hope to show, important biases almost surely remain.

I begin by reviewing the importance of price indices in our everyday lives and therefore why biases in the overall indices – and changes in those biases – matter. I then describe briefly the assumptions and theory underlying the official price indices. I next turn to the problems of measuring medical prices, assuming the basic applicability of the theory upon which the official indices are based. Finally I turn to the potential inapplicability of the assumptions made by that theory and the resulting issues for measuring price changes. I describe an alternative theory and its implications for the measurement of medical prices.

The material presented here will generally apply to any country with insurance arrangements for medical care; in such cases, there will be a price the insurer, which may be the government, pays medical care providers, a price that in principle a statistical agency can observe. To some degree the material also applies to those countries with public production of medical care, such as the British National Health Service. For example, groups of physicians in the United Kingdom now receive monies from the state. From these funds they are to purchase hospital care, drugs, and other medical inputs in

¹ The chain-type price index for American domestic purchases is even more affected because of the greater weight of medical care among domestic purchases.

accordance with the medical needs of their patients, so that in principle prices can be observed in this case as well.

In public health care systems that operate on budgets, which are present even in the United States (e.g., the Veterans Administration and military hospitals), unit prices are not observed. Countries vary in how they treat such systems in constructing a GDP deflator, but the most common method is probably to apply an input price index (e.g., an index of wages) to nominal spending. Many of the problems discussed below apply to such a procedure. In particular, such a procedure assumes that productivity growth is zero; for example, it assumes the savings from a shift of a procedure to a less expensive setting outside the hospital is a fall in real output. This procedure also ignores the pervasive problem of quality change.² For further discussion see (Berndt et al., 2000b).

The material presented here grows out of work on medical price indices by a group of economists at the National Bureau of Economic Research over several years, a group in which I have had the good fortune to participate and to which I am clearly indebted. What follows builds on several papers from this group: (Berndt et al., 2000a), (Berndt, et al., 2000b), (Berndt et al., 2001a), (Berndt et al., 2001b), (Cutler et al., 1998), (Cutler et al., 1999), (Cutler et al., 2001), (Triplett, 1999), and (Triplett, 2001).³

The Uses and Importance of Economy-wide Price Indices

Price indices, of course, measure changes in nominal prices. The indices have many important uses that touch the lives of us all.

Monetary and fiscal policy. In modern times the measured rate of price change – and changes in the measured rate – have great importance for monetary and fiscal policy. The popular press is filled with articles that seek to divine how Alan Greenspan and other central bankers will react to changes in price indices. These articles, which used to be found on financial pages of newspapers, now are often found on page one of the main news section. The monthly announcements of price index changes by the US Bureau of Labor Statistics and the statistical agencies of other nations frequently move nominal interest rates, exchange rates, and equity prices.

For purposes of monetary and fiscal policy a constant bias in a price index, especially an additive bias, is not overly troublesome; it will presumably become incorporated into the expectation of a normal inflation rate, or a NAIRU, and those charged with implementing monetary and fiscal policy will generally react only to departures from that expectation.⁴ A change in the bias, however, would send a false signal of accelerating or decelerating inflation. And because the central bank is often

² As described below, existing price deflators also ignore measures of quality change.

³ The Bureau of Economic Analysis of the United States Department of Commerce, which is responsible for the GDP deflator, and the Bureau of Labor Statistics of the United States Department of Labor, which is responsible for the Consumer and Producer Price Indices, have helped support some of the work of this group.

⁴ NAIRU is the non-accelerating inflation rate of unemployment.

prepared to act on a small change in a true signal, changes in the bias may importantly change monetary and fiscal policy and hence affect the overall economy.

Indexing of taxes, transfers, bonds, and wages. Measured rates of price change are used to index taxes and transfers in certain public programs, and are also incorporated in private contracts. For example, in the United States and in several other countries changes in public pension benefits (Social Security in the United States) are tied to changes in the CPI, so that they will maintain their value in real terms. Overall, just under a third of total American federal outlays are indexed to the CPI (Boskin et al., 1996). American personal income tax brackets are also tied to the CPI on the premise that real income should be taxed, and that inflation should not change the progressivity of the personal income tax. The United Kingdom has for many years issued indexed Treasury bonds, and the United States has also recently started to issue bonds whose payments are indexed to the CPI so as to virtually eliminate the risk of inflation in holding bonds. Private multiyear wage contracts are also often indexed to the CPI. Finally, the Maastricht agreement on the European Monetary Union contained an inflation target for countries to join the Union.

Although a constant bias in the indices is not overly troublesome for purposes of monetary and fiscal policy, for the purpose of indexing pensions and tax brackets it can have serious consequences. A group of five distinguished American economists chaired by Michael Boskin examined the American CPI in 1996 at the request of the Finance Committee of the United States Senate (Boskin, et al., 1996), (Boskin et al., 1998).⁵ They concluded that a 1.1 percentage point bias in the American CPI, which they regarded as a conservative estimate of the bias, would lead to \$1 trillion ($\10^{12}) of additional national debt in a 12 year period, mainly because of additional pension benefits and reduced tax collections (Boskin and Jorgenson, 1997).⁶ To provide a sense of scale, \$1 trillion is around a fifth of the current total American national debt. Changes in the bias, at least if they are random, are also troublesome for purposes of indexing pensions and tax brackets, but the first order effects come from the overall bias. A constant upward bias could also lead to overly restrictive policies for countries desiring to join the European Monetary Union.

A constant bias, however, is not so serious for indexing bonds or wage contracts. Such a bias would presumably be factored in to the real interest rate that investors would be willing to pay for the bonds, although any changes in the bias would cause windfall losses or gains and make such bonds less appealing to risk-averse investors. Similar considerations would apply to wage contracts; a constant bias would presumably be accounted for in negotiation, but changes in the bias would cause windfall gains and losses.

⁵ The four economists other than Boskin were Ellen Dulberger, Zvi Griliches, Robert J. Gordon, and Dale Jorgenson.

⁶ An upward bias in the CPI acts to reduce tax collections by increasing the income level at which higher marginal tax rates take effect.

Changes in productivity and real wages. Another important use for price indices is to measure changes in productivity, that is to measure the change in the quantity of real goods and services being produced in the entire economy, or in a sector of the economy.

Changes in multifactor productivity come from a measure of the change in real output(s) relative to a measure of the change in real inputs.⁷ The real quantity of output is usually measured by $\Delta(\sum p_i q_i)/\Delta P$, where the numerator is the change in nominal GDP or in the nominal monetary value of the output of a sector, and ΔP is the change in an output price index. A constant upward (downward) bias in the output price index obviously biases down (up) the measure of change in real output and similarly for changes in the bias. Measurement of the real change in inputs similarly requires an input price index, but any bias or changes in the bias of output price indices will generally differ from those of input price indices and hence will not cancel.⁸ If any upward bias in the output price index exceeds that in the input price index, the growth in productivity will be understated and conversely.

Although changes in measured productivity do not usually have the immediate policy implications that measured changes in inflation have for monetary policy, they can nonetheless have an important influence. For example, measured productivity in many developed economies slowed down in the early 1970s. Initially this slowdown was attributed to the shocks from the large, rapid increase in the price of oil. By the 1980s, however, this explanation of the slowdown seemed less plausible. Enough time had passed so that it seemed economies should have adjusted to the price shock, but measured productivity increases in the United States and several other countries remained low.

And the Japanese economy, which should have been as vulnerable to the oil price shocks as the American economy, was performing substantially better. As a result, at the time some American commentators touted the merits of an industrial policy, such as found in Japan (Thurow, 1984, 1987). An industrial policy would have substantially increased government involvement in the economy, since the government would have been trying to choose winning industries for investment, much as the Japanese Ministry of International Trade and Industry did. In the 1980s, however, the American political climate was not receptive to such proposals. This was probably fortunate, because in the 1990s the Japanese and American economies reversed their performances (Table 1). The American economy performed well, and productivity improved sharply. In fact, the 3.0 percent gain in American private business multifactor productivity in the four years from 1993 to 1997 exceeded the 2.7 percent gain over the entire 9 years from 1984 to 1983.⁹

⁷ A specific parameterization of the change in multifactor productivity m comes from the Solow residual and equals $y - \sum \beta_i \eta_i$, where y is the growth in real output, β_i is the real growth rate of input i , and η_i is the elasticity of output with respect to input i . This measure assumes a constant returns to scale production function.

⁸ Measurement of the productivity of single factors, for example, output per labor hour, may be in physical terms and not require input price indices. Similarly, output could be measured in physical terms, but this is unusual.

⁹ These percentages are cumulative not annual. See <http://146.142.4.24/cgi-bin/surveymost> for the raw data. I chose 1984 and 1993 as the beginning years for this measurement because the economy had substantially come out of the early 1980s and early 1990s recessions by that time. Had I chosen 1985 as a

Today one hears virtually no discussion in the United States of an industrial policy akin to the Japanese model.

Although undoubtedly much of the gain in American productivity in the 1993-1997 period was real, some is likely attributable to improvements in the accuracy of official price indices, which decreased the upward bias in those indices. We shall encounter some examples of those improvements in the medical care component below. But we shall also show that the underlying bias in the medical index may have increased in those years, so that the true productivity gain was even greater than measured.

Closely related to the issue of measuring the change in productivity is the measurement of real wages and income. Real American average hourly earnings fell 3.5 percent and real average weekly earnings fell 10 percent between 1970 and 1998.¹⁰ But any upward bias in the CPI will obviously reduce the measured change in the real wage. (Boskin and Jorgenson, 1997) estimate that average hourly earnings increased 13 percent instead of falling by 14 percent between 1973 and 1995 if there was a 1.1 percentage point upward bias in the CPI. Similarly, they estimate that real median household income in the United States increased 36 percent rather than 4 percent between 1973 and 1995 if the bias in the CPI was 1.1 percentage points annually.

(Nordhaus, 1998) estimated that a 1.1 percentage point upward bias in the price indices used to calculate productivity would have meant that American total factor productivity in fact rose 1.8 percent per year between 1973 and 1995, rather than the anemic 0.6 percent per year computed from official indices. He also estimated that real wages in the United States would have increased 61 percent between 1960 and 1995 rather than the 10 percent value one would calculate using the CPI.

Thus, three important reasons to be concerned about biases in the overall price indices are the possibility of errors in monetary and fiscal policy, improper indexing of taxes and transfers and government bonds and the subsequent effect on public outlays, as well as improper indexing of wages, and mismeasurement of productivity and real wages. Insofar as biases or changes in biases in the medical care price indices contribute to problems in overall price indices, there is clear reason for concern about the medical indices.

Medical Care Price Indices

Three other reasons for concern about the medical care price index are specific to the medical sector.

beginning year on the grounds that the economy was still not out of the recession in 1983 and measured cumulative productivity gains to 1993, the productivity gain over those eight years would have been a paltry 1.8 percent (0.2 percent annually). And the productivity gains continued through 1999; the change in multifactor productivity from 1996 to 1999 was 1.16 percent annually (3.5 percent cumulatively) (Oliner and Sichel, 2000)

¹⁰ This calculation, uses the CPI to deflate nominal wages, (U.S. Council of Economic Advisers, 1999), Table B-47.

Measurement of changes in real output and productivity in medical care. Many countries are now debating whether competitive arrangements in medical care improve productivity through the stimulus to more efficient production or whether they detract from it because of socially unproductive efforts at selection of good risks (Cutler and Zeckhauser, 2000), (Hurley, 2000), (Newhouse, 1996), (Rice, 1998). This debate takes place against a backdrop of a natural experiment in which several countries have, to varying degrees, attempted to increase the degree of competition in medical care delivery and financing (van de Ven and Ellis, 2000). Among those countries are Israel, the Netherlands, Switzerland, the United Kingdom, and the United States. The debate, however, continues in part because empirical content on what has happened to the real quantity of medical services as the degree of competition changed has been lacking.¹¹ A standard method for obtaining information on changes in the real quantity of services in any industry is to deflate changes in nominal spending by changes in an industry specific price index. Biases in the price index, of course, will proportionately bias measured changes in the real quantity of services in the opposite direction.

Measurement of medical prices is also important in other policy decisions toward the medical sector. I give two American examples. Between 1927, when measurement of the medical care component of the American CPI began, and 1999, the medical care component has increased at annual rate of 4.53 percent, whereas the overall index only increased at an annual rate of 3.19 percent. Compounded over the 72 years since 1927, the medical care index has risen by more than a factor of 24, while the overall index has risen by only a factor of 9.6.¹² The above average increase in the American medical care price index has at times led to a public perception of increasing rents to medical care providers.¹³ This perception may well have played a role in the 1993 Clinton Administration decision to press for expenditure controls in medical care, a feature that in turn may have played a part in the defeat of its health care proposals, its major domestic initiative upon taking office.¹⁴

A much more specific use for a measure of medical productivity is the administered price system in the American Medicare program, which covers almost all elderly Americans and certain other individuals. In absolute size the Medicare program is the largest health insurance program in the world, now amounting to over 2 percent of

¹¹ There is a further question about whether an increase in medical services is beneficial; I take up the possibility of services with a negative marginal product below. In addition to effects on efficiency, distributional concerns are part of the debate.

¹² In the case of the Personal Consumption Expenditure deflator and its medical care component, the discrepancy is less; between 1929, the first year it was measured, and 1999 it rose 3.10 percent per year (a factor of 8.5) versus 4.10 percent per year for its medical care component (a factor of 16.7).

¹³ Among the first to discuss rents in physician prices were two future Nobel laureates, Milton Friedman and Simon Kuznets. Their work was an outgrowth of Friedman's doctoral dissertation. See (Friedman and Kuznets, 1945).

¹⁴ The Clinton Administration's felt need to obtain a low "score" from the Congressional Budget Office in order to avoid proposing a tax increase to finance its proposal also played a large role in its decision to propose expenditure controls. Nonetheless, the public perception that medical care prices were increasing rapidly and were accounting for a considerable part of the medical care spending increase surely factored into the decision to propose controls as well. And biases in the pharmaceutical component of the index may have played a role in the Clinton Administration's specific proposals to control drug prices.

American GDP and projected to grow to 4.4 to 5.6 percent over the next three decades.¹⁵ Those managing this system are in principle to account for any change in productivity at hospitals when adjusting the prices paid hospitals from year to year. Specifically, the annual increase in prices is to be reduced by the amount of the change in productivity. Because a reliable deflator for hospital spending has been lacking, however, the adjustment for changes in productivity has in practice been a judgment (Medicare Payment Advisory Commission, 1999).

Updating payments in a defined contribution framework for health insurance. A related reason why one might care about medical care price indices has to do with updating payments or subsidies from government or employers, if the structure of financing uses a defined contribution approach with competing health plans. In such a case the government or the employer contributes a lump sum toward the cost of any health plan that an employee or beneficiary chooses, with the employee paying the marginal dollar.¹⁶ Such a method has been proposed as a reform of the American Medicare program (Aaron and Reischauer, 1995), (Cutler, 1995), (Wilensky and Newhouse, 1999).¹⁷ A major point of contention in the debate over the desirability of such a proposal is how the government contribution will grow over time to reflect changes in prices and spending, because this will critically affect the division of payment between taxpayers and beneficiaries of the Medicare program.¹⁸

It may appear at first blush that the appropriate update factor is a medical care price index. After all, the all-items CPI is the update factor in the public pension program (Social Security) through which beneficiaries receive annual cost-of-living increases. Elsewhere, however, I have suggested that the major explanation of the sustained spending increase on medical services is technological change in medicine, including the addition of many new product lines (Newhouse, 1992). If that is correct, using a medical care price index to update the lump sum would put the entire burden of paying for cost-increasing technological change on the current elderly.¹⁹ Because medical care spending

¹⁵ The Board of Trustees of the Federal Insurance Hospital Trust Fund, "2000 Annual Report of the Board of Trustees of the Federal Insurance Hospital Trust Fund," (corrected April 20, 2000), page 82, available on the web at <http://www.hcfa.gov/pubforms/tr/>; Congressional Budget Office, "The Long Term Budget Outlook," October 2000; Table 4, available on the web at <http://www.cbo.gov/showdoc.cfm?index=2517&sequence=0&from=1#c2>.

¹⁶ Although such an arrangement exacerbates problems of selection, it also can reduce health plan markups by increasing the price elasticity of demand that health plans face (Newhouse, 1996), (Cutler and Reber, 1998).

¹⁷ A defined contribution approach gained political prominence in the United States in 1997 as a result of the proposals of the Bipartisan Commission on the Future of Medicare, which can be found at <http://thomas.loc.gov/medicare/bbmtt31599.html>. The Clinton Administration in 1999 proposed an analogous approach, in which a Medicare beneficiary would receive 75 percent of the savings if he or she chose a less expensive health plan. President George W. Bush endorsed a defined contribution approach in the 2000 Presidential campaign.

¹⁸ If there is a voluntary element, as in the current program, the size of the subsidy may also importantly affect the degree of adverse selection.

¹⁹ In a lifetime context in which the economy is in steady state and in which individuals have the same life expectancy, this choice should not affect expected welfare; individuals would either pay less when young and more when old or vice-versa. In practice, of course, in addition to the issue of how to treat the current

increases have historically run well ahead of increases in the income of the elderly, having the current elderly pay for all the technological change is not a likely outcome of the political process.²⁰

An alternative proposal is to index the lump sum by a measure of changes in private health care spending or private health care premiums.²¹ Because current Medicare spending is principally financed from current taxes (i.e., it is a pay-as-you-go system), such an update method would place the burden of financing the future technological change on taxpayers, who are mainly non-elderly.²² In practice, the burden is likely to be shared between the elderly and non-elderly, meaning the outcome will probably fall between these two possibilities. An accurate medical care price index, however, may be considered a reasonable lower bound on an appropriate update factor.

Testing Baumol's cost-disease theory. A final reason for wanting a medical care price measure is to test an interpretation of medical care spending increases that comes from a more general and well-known theory of William Baumol's (Baumol, 1967), (Baumol, 1993), (Baumol, 1997). Baumol suggested that the relative price of labor-intensive goods will rise over time. This result follows from the plausible assumptions that productivity in the economy and therefore wages in all sectors will rise, but that productivity will not rise as rapidly in the production of labor-intensive goods. Baumol's original example was live chamber music, where he postulated that any productivity

elderly, population demographics are not in steady state (both because of fertility changes and immigration), and individuals do not have the same life expectancy.

²⁰ Implicitly the American Congress has historically rejected the option of placing the entire burden on the current elderly. The portion of the Medicare program that covers physician and certain other outpatient services (Part B) is voluntary and requires a premium payment by the beneficiary. When the program began in 1966, Part B was to be financed half from premium payments by the elderly and half from general revenues. Because the rate of increase in Part B spending exceeded the rate of increase in the income of the elderly, the Congress over time shielded the elderly from the increased premiums by reducing the share of the program financed through the elderly's premium from 50 to 25 percent. Because of the growth in medical spending and the growing share coming from general revenues, Part B grew to be the largest domestic program that the American government financed from general revenues. A major issue in the 1995 "shutdown" of the federal government was a political impasse over whether this share should be stabilized at 25 percent or should be further reduced. For now the share has been stabilized at 25 percent.

²¹ Such a proposal may not be feasible outside the American context because of the small and generally non-representative size of the private sector.

²² The portion of the Medicare program that covers hospital and other institutional services such as nursing homes (Part A) is almost entirely financed from payroll taxes and constitutes 60 percent of the program. Because payroll taxes are only levied on those who work, they fall upon almost entirely on the non-elderly. The remaining 40 percent of the program is subject to the 75-25 split between general revenues and premium payments noted in the previous footnote. General revenues are mainly derived from income taxes and probably fall disproportionately on the non-elderly, since the elderly have little wage and salary income and most Social Security payments as well as unrealized capital gains are exempt from tax. Overall, therefore, close to 90 percent of the program (60 percent plus 75 percent of the remaining 40 percent) is tax financed. In addition to the Part B premiums, however, the elderly also bear substantial costs in relation to their income for uncovered services, especially long-term care and prescription drugs, and for cost sharing amounts for covered services or for supplementary insurance.

increase will be minimal.²³ If demand for the labor-intensive good is price inelastic, the price increases will produce spending increases, which Baumol termed the “cost disease.”

Because demand for medical care is price inelastic (Newhouse et al., 1993), Baumol’s theory has been interpreted as suggesting that spending increases in medical care are inevitable.²⁴ Given the rising share of GDP devoted to medical care in every OECD country (Table 2), Baumol’s interpretation has a certain superficial plausibility. The implications for policy, however, are exactly the opposite of the call for price or spending controls that some have drawn from the interpretation of increasing rents described above. In Baumol’s theory the price and spending increases in medical care are the inevitable byproduct of progress in a market economy.

I shall show below that, when properly measured, relative prices for medical care may not have risen in real terms, in which case Baumol’s cost disease cannot explain the rise in medical spending. If instead the reason for the increase in spending is the increased capabilities of medicine, the relevant policy question becomes the value and cost of the real services that society is buying with the increased spending. Unlike most goods and services, one cannot make the usual assumption assume that the value received at the margin for a medical care service equals the opportunity cost of the resources used to produce it because of widespread health insurance, among other reasons.²⁵

The Theory and Assumptions Underlying the Official Price Indices

A cost-of-living index seeks to measure the change in income necessary to maintain an individual at a given level of utility. The American Bureau of Labor Statistics correctly notes that the CPI is only an approximation to a cost-of-living index because it measures the change in income needed to purchase an initial period market basket (i.e., it is a modified Laspeyres index). In fact, however, it is often described as or

²³ Baumol reasoned that it will presumably take four musicians approximately the same amount of time to give a live performance of a Beethoven string quartet whether it was played twenty years ago or twenty years hence. Jack Triplett, however, has commented to me that rehearsal time is an input and has in fact decreased greatly for any specific work.

²⁴ This notion has even filtered through to the United States Congress. See, for example, the statement of Senator Daniel Patrick Moynihan (D-NY), then the chairman of the Senate Finance Committee, in United States Senate, Committee on Finance, “Hearing on Underlying Causes of the Rise in Health Care Costs,” October 6, 1993, in which he refers approvingly to Baumol’s theory in the context of commenting on the futility of the expenditure controls proposed in the Clinton health care reform proposal. Mr. Moynihan, like President Clinton, is a member of the Democratic party, so his remarks were not a partisan attack on the President’s proposal.

²⁵ This is not, of course, a criticism of health insurance, which acts to shift resources to states of the world where they are more desired. How health insurance is structured, however, can have important consequences for the degree of moral hazard (Newhouse, et al., 1993), (Pauly, 1999). Moreover, in the case of a new technology the average rather than the marginal gain is the relevant measure of the welfare gain. My personal view is that the evidence of a low marginal payoff to additional medical spending at a point in time is substantial (e.g., (Chassin et al., 1987), (Chassin et al., 1998), (McClellan et al., 1994; McClellan and Newhouse, 1997), (Newhouse, 1992), (Newhouse, et al., 1993)), but that the payoff over time from technological change in medicine has been large. Such a conclusion is clearly consistent with the rising share of GDP that virtually every developed country, encompassing a wide range of financing institutions, has chosen to devote to medical care.

used as if it were a cost-of-living index, for example in indexing wages, pensions, and tax rates.²⁶ For ease of exposition I shall often refer to the CPI as a cost-of-living index, although at times I will take note of the difference between the actual index and a cost-of-living index.

Briefly put, the theory that underpins a cost-of-living index is that of a competitive economy (Fisher and Shell, 1972). The theory has operational difficulty with new goods, which are common in the medical sector. It also assumes that changes in levels of utility occur from changes in the consumption of marketed goods and that the production of goods and services is always on a production-possibility frontier. Both these assumptions are problematic in the context of medical care. Before coming to the larger issues of the applicability of the theory, however, I take up a series of more practical problems that have plagued, and in some cases continue to plague, the official indices. These problems, although sometimes mundane, have important consequences for the accuracy of the indices.

Medical Care Price Indices: Problems within the Context of the Standard Theory

As Table 2 shows, the United States devotes a larger share of its GDP to medical care than any other developed economy. Although the figure naturally fluctuates somewhat from year to year, in recent years the share has exceeded 13 percent (Levit et al., 2000) and that is approximately the weight of medical care in the American GDP deflator as well.²⁷ Most other developed economies now have shares that are between 7 and 11 percent. These figures, however, include public health spending, research, and construction. For the purpose of deflating medical care spending the subset of spending for personal health care would be more appropriate, and it is between one and two percentage points less. I have kept the more inclusive figure, however, to maintain comparability with data from other countries.

At this size biases in the medical care price index can create non-trivial biases in the overall American GDP deflator. For example, if there is a constant 3 percentage point annual upward bias in the medical care index used in the GDP deflator, the figure the Boskin Commission estimated for the medical component of the CPI, there was potentially about a constant 0.4 percentage point annual upward bias in the overall

²⁶ See www.bls.gov for the Bureau's view. As described below, the modification to a pure Laspeyres index is to use a geometric mean for elementary aggregates, but not at higher levels. There is a further issue concerning the CPI of whether different market baskets are appropriate for different groups of individuals, and, if so, whether there are important distortions from using the same market basket for all persons. I ignore this issue as well.

²⁷ The figures for GDP share given in Table 2 and the text are based on various countries' national health accounts, whereas the weight for the GDP deflator is based on spending in the National Income and Product Accounts. In the United States there are minor definitional differences between these values that I shall ignore. In 1996, for example, total medical spending as measured in the National Income and Product Accounts and in the National Health Accounts differed by only 3 percent (Sensenig and Wilcox, 2001), and the correlation in annual spending between the two series exceeds 0.99 (Huskamp and Newhouse, 1994). I have not seen data for other countries, but in principle the discrepancy should also be small in those countries. Netting out intermediate goods, the share of health care in the American GDP in 1996 was 5.9 percent (see (Lum, 1997), Table 7, page 28, line 69).

American GDP deflator in the 1990s just from the medical component.²⁸ The weight of medical care in the American CPI, about 5 percent, is substantially less than in the GDP deflator for reasons described momentarily; hence; there was potentially about a 0.15 percentage point annual bias in the overall CPI. Although there is a great deal of uncertainty about the amount of the bias, I attempt to show below that a 3 percentage point value is very likely too low, so that the bias in the overall indices was likely greater.

Moreover, I also try to show that between 1993 and 1997 the annual bias in the CPI may have increased by an additional 2 to 3 percentage points, to around 5 to 6 percentage points. Because it is constructed differently, the change in the bias in the GDP deflator is less, though how much less is hard to know. This is clearly a non-trivial change in the bias in the medical care component of the CPI, which would have raised the bias in the all-items CPI from 0.15 to 0.25-0.3 percentage points. Nonetheless, because the average annual change in the CPI in the 1993-1997 period was only 2.7 percent (and in the GDP deflator only 2.0 percent), both the size of the bias and the size of the change in the bias are potentially consequential.

Finally, in addition to the potential change in the bias in the mid 1990s, the size and sources of the specific biases described below make it plausible that there are other non-trivial annual changes in the bias, although neither I nor anyone else has attempted to quantify them. If, for example, the bias in the medical care component changed annually up or down by a random one percentage point, there would be somewhat more than a 0.1 percentage point change in the bias in the overall GDP deflator.

An analogous problem arises if the overall index is chain-weighted (i.e., if the shares of various goods are reweighted annually), as the American GDP deflator now is, and there are errors in the measure of the share of the GDP in medical care. Figure 1 shows the annual change in the share of GDP devoted to health care. The coefficient of variation around the mean change of 2.56 percent is nearly 1. It seems unlikely that the true shares vary this much; in other words, a non-trivial portion of this variation could represent measurement error in the share, which would result in error in the overall index.²⁹ I now describe several specific problems with the official indices that appear to cause substantial biases.

Omission of the government sector and employer-paid premiums. Most of the biases I discuss apply to both the GDP deflator and the CPI, but two important differences between them concern the weight of medical care in the overall indices and the medical care market basket that each index uses. The market basket for the GDP deflator includes the entire economy, whereas the market basket for the all-items CPI uses only those goods and services that consumers purchase with their after-tax incomes. This definition reflects the history of the CPI, an important initial purpose of which was

²⁸ The bias in the overall index is $\sum_i w_i b_i$, where w_i is the share of component i in the overall index and b_i is the bias in component i .

²⁹ Consistent with this interpretation, the Office of the Actuary of the Health Care Financing Administration makes annual revisions in the national health accounts. In recent years these have often been substantial.

to measure changes in purchasing power in order to keep wages constant in real terms between rounds of collective bargaining.

In the United States 37 percent of all medical care was purchased by various levels of government in 1994 and was thus excluded from the CPI market basket, although not from the GDP deflator. In most other developed countries this fraction is substantially higher, and it has been rising in the United States. Another 26 percent of American medical care was nominally paid for by employers as a fringe benefit of employment (Cowan et al., 1996). Neither government nor employer-financed care is considered in the market basket for the CPI, because neither is viewed by the Bureau of Labor Statistics, the agency responsible for the American CPI, as part of take-home pay. As a result, almost two-thirds of medical spending is outside the scope of the American CPI, and the weight of medical care in the CPI is therefore much lower than in the GDP deflator, just a little over 5 percent versus the 13 percent in the GDP deflator, as already noted.

Despite the important role of government in financing medical care throughout the world, however, some medical care is purchased from disposable, or take-home, income in all countries. This reflects both spending on health care services that are not covered by any public insurance plan, as well as cost sharing arrangements (e.g., coinsurance) in a public plan. Only these amounts in principle enter a CPI market basket. The exclusion of government- and employer-financed services from the CPI, while arguably correct from the point of view of measuring the cost of a market basket that can be bought from take-home pay, can seriously distort a medical care sector price index to the degree that various services are covered differentially by insurance.³⁰ In the United States, for example, hospital services have the most complete insurance coverage. They therefore receive a disproportionately low weight in the American medical care CPI precisely because of the coverage. Hence, the medical care CPI cannot be used to measure the productivity of the medical sector.

The distortion in the medical care CPI as a measure of the cost-of-living that is caused by the exclusion of government payments can be seen in the negative correlation between the prices paid hospitals by the American Medicare program and those paid by the privately insured. On average, Medicare accounts for about a third of hospital revenues. Since 1984 the Medicare program has paid hospitals an administered price, which is a lump sum per admission that is contingent upon the patient's diagnosis and whether certain medical procedures are carried out.³¹ Vis-à-vis Medicare, the hospital is a price-taker. In game-theoretic terms, Medicare moves first and sets a price; hospitals then set their prices to private insurers.

³⁰ Even on the logic of the CPI, the exclusion of employer-paid health insurance premiums from the CPI is questionable, since standard economic theory suggests they are mostly shifted to wages and hence are part of total compensation, the real value of which is clearly affected by changes in medical prices. See (Gruber, 2000), (Pauly, 1997), (Summers, 1989).

³¹ The administered price system is called the Prospective Payment System and is based on Diagnosis Related Groups (DRGs).

Suppose Medicare reduces the rates it pays. If there are not rents in hospital prices, hospitals must raise private prices to recover their joint (across-payers) costs. And although hospitals probably have some market power, especially in rural areas, there is empirically a substantial negative correlation between Medicare and private prices, as shown in Figure 2. The “prices” shown in Figure 2 are actually margins, that is, revenues per admission divided by the accounting cost of those admissions.³² Since 1984, when the Medicare program began using an administered price per admission, rather than pay a proportion of the hospital’s total cost, the negative correlation (r) between the two margins is -0.84 ($R^2=0.70$). In other words, if Medicare pays less, hospitals have tended to charge private payers more and conversely.³³

As a result, when Medicare has reduced the prices it pays hospitals, the CPI has increased and conversely. Consumers as a group, however, may be no worse off, because the payroll taxes they pay to finance the Medicare program are less than they otherwise would have been.³⁴ In short, if Medicare pays less, consumers’ standard of living may be unchanged, even though the increase in private prices means the CPI would have registered a decrease in real wages. Over the entire period from 1984 to 1998 Medicare margins have somewhat increased and private margins have fallen. This suggests a small downward bias in the CPI over the entire 14 year period, although, as Figure 2 shows, for much of the period there was an artificial increase in the CPI from this cost-shifting phenomenon followed by an artificial decrease.

Because Figure 2 shows margins and not prices, I cannot map the relationship shown in Figure 2 into the amount of the bias in the CPI. Nonetheless, the change of around 20 percentage points in the private payer margin, both up and down in a 14-year period, suggests the bias may have been both substantial and variable. This bias, of course, does not apply to the GDP deflator since it includes the Medicare payments.

The Nature of the Market Basket. A second problem with the official indices is that they often price a market basket of specific services, such as the cost of a hospital service or the cost of a physician visit or the cost of a drug. I proceed from the premise that what the consumer wants from the medical care system is the diagnosis and treatment of whatever ails him or her. When the method of treating the ailment changes, as it often does because of new knowledge, the market basket used in the price index does not change and so the index fails to register the consequences for the cost of treatment. For the moment, assume that there are no consequences for health of the change in the method of treatment. I will come below to improvements in quality.

³² Although allocating joint costs to particular patients (payers) is arbitrary, the procedures for allocating costs to specific payers have remained reasonably stable over time and cost increases have probably been reasonably similar across payers. As a result, changes in margins are likely to be a good measure of how prices have changed.

³³ Consistent with this interpretation, in the few years before 1984 for which there are data and when Medicare did not set a price but rather reimbursed its share of cost, there is almost no correlation between the two series; the overall correlation if one begins in 1980 falls from -0.84 to -0.68 (the R^2 falls to 0.47).

³⁴ Medicare precludes hospitals from billing its beneficiaries directly.

In earlier years the failure to price the treatment of an episode of illness may have caused the change in the CPI to be understated (Scitovsky, 1967, 1985). But in recent years it is likely that the bias has been in the other direction. For example, much surgery, as well as the treatment of some diseases, has dramatically shifted out of the hospital to outpatient settings, with, as best is known, no or minimal consequences for the patient's ultimate health (Table 3). As a result, procedures that in past years required a several day inpatient stay now require no inpatient stay. (Shapiro and Wilcox, 1996) estimate that between 1969 and 1993 the price of cataract surgery was overstated 4.6 percent per year because of the assumption that it was performed as an inpatient procedure, whereas during those years it shifted to become for the most part an outpatient procedure. The 4.6 percent estimate does not even consider the gains from reduced recovery time, as well as a lower likelihood of medical error and nosocomial (hospital-induced) infection. An analogous phenomenon occurs when drug treatment replaces surgical treatment of a problem, as was the case for the treatment of stomach ulcers in the past decade.

Starting in 1997 the CPI put inpatient and outpatient procedures in the same category of hospital services, thus capturing the saving from the shift of site. The Index continues to not account for the substitution of drug therapy for inpatient or outpatient procedures, however. Thus, before 1997 when a new market basket was defined that included less use of inpatient services, it was simply linked in with the old market basket in a way that left the average unchanged, so the Index never captured the savings, and the problem remains for substitution of drug therapy.

A similar problem occurs from the fall in hospital length of stay over the last two decades, shown in Table 3.³⁵ Until 1997 the American CPI priced a hospital day. As stays fell, of course, the cost of an episode fell.³⁶ To the degree the shorter stays caused worse outcomes, there should be an adjustment to the fall in the price of an episode, but the most definitive study of the sharp decline in length of stay among the elderly when the Medicare program introduced prospective payment of hospitals on the basis of stays (i.e., no marginal revenue from another day) found little evidence of worse outcomes (Kahn et al., 1990).³⁷

Although the fall in the length of stay presumptively reduced the price of a hospital episode, the CPI's pricing a hospital day rather than a hospital stay or an entire episode of illness sent the opposite signal. Patients are sicker and require more nursing time in the first days of their stay, so a fall in the length of stay *increases* the average cost of a day in the hospital. As a result, not only did the official indices fail to measure the fall in the cost of a stay from fewer days, they actually signaled an increase in hospital

³⁵ Before 1980, however, length of stay was rising.

³⁶ In the 1990s there is substantial evidence of unbundling by hospitals in the form of substitution of post-acute services for the marginal day(s) of the hospital stay, which, if properly accounted for, would reduce but not eliminate the amount of the bias from measuring the cost of a day rather than the cost of an episode of illness.

³⁷ Recent, and as yet unpublished, work by Yu-Chu Shen suggests that there may have been increased 30 day mortality among patients who had a heart attack although one year mortality did not increase; at conventional values of life, Shen finds that the savings in treatment cost much more than outweighed the value of the increase in mortality.

prices from an increased price per day.³⁸ In other words, when a true price index should have fallen from the decrease in the stay, the official price indices instead tended to rise. In a study of heart attack care that I will describe in more detail below, we found that the bias from not using the stay as a unit of service was on the order of 2 percentage points per year (Cutler, et al., 1998).

When the Producer Price Index (PPI) for hospital services began in December 1991, it priced a hospital stay rather than follow the lead of the CPI and pricing a day.³⁹ In 1997 the CPI changed to pricing a stay,⁴⁰ but for the period of time in which the PPI priced the hospital stay and the CPI priced the hospital day (December 1991 through 1996), the CPI for Hospital and Related Services increased 2.2 percentage points more per year (77 percent) than the PPI for General Medical and Surgical Hospitals. Although consistent with the notion that using price per day results in a greater rate of increase than price per stay when length of stay is falling, as it was in this period (Table 3), and that the substitution of outpatient surgery lowers prices, this test is not conclusive because of other differences between the two indices; in particular, the PPI includes prices of all payers, whereas the CPI does not.⁴¹

The change in the CPI to pricing a hospital stay in 1997 improved the accuracy of the Index by addressing the bias from falling length of stay, but the bias from ignoring the substitution of cheaper, alternative methods of treatment, such as the move to perform many procedures out of the hospital and the use of drug therapy instead of surgical intervention, remains. Other than the value for cataract surgery I cited, I have not found estimates of the amount of bias from this cause, but the magnitude of the shift to outpatient surgery shown in Table 3 suggests that the bias is non-trivial.⁴² This bias also remains in the PPI, although starting in January, 2001, the PPI is eliminating the distinction between inpatient and outpatient treatment, which will resolve the problem of bias from inpatient procedures migrating out of the hospital for that index.⁴³

³⁸ As I come to below, the CPI mainly measures list prices, so this statement needs to be qualified to the degree that list prices for a hospital day may not necessarily reflect cost per day.

³⁹ The American GDP deflator accounts for hospital prices using the PPI rather than the CPI.

⁴⁰ If the hospital bills the payer on a per diem basis, the price quote used in the Index is the reimbursement the hospital receives for the stay. Before 1997 there was a limited amount of aggregation in the definition of the service that the CPI priced; anesthesia services, for example, bundled drugs and supplies with nurses and technologists (Cardenas, 1996). Nonetheless, although I have not attempted to quantify it, the change in 1997 to pricing the stay clearly introduced substantially more aggregation into the Index.

⁴¹ Indeed, the difference between the two indices widens between 1997 and October 2000, consistent with the importance of other factors that affect the indices. In particular, the Balanced Budget Act of 1997 substantially reduced the prices Medicare otherwise would have paid, which entered the calculation of the PPI but not the CPI. Between 1993 and 1996, the CPI hospital index increased 2.91 percent per year and the PPI 5.14 percent. Between 1997 and October 2000 the two indices increased 1.54 percent per year and 4.00 percent respectively.

⁴² Some of the gains from managed behavioral health care, described below under the rubric of movements toward the production possibility frontier, also involve the substitution of cheaper outpatient services for inpatient care.

⁴³ The PPI will also implement a new item substitution policy, in which it will attempt to determine on a real time basis whether a new treatment has begun to replace an old treatment for a set of seven Diagnosis Related Groups. If it has, the Index will replace the old treatment with the newer treatment method.

Several recent changes in the price indices have been incorporated into revisions in the American National Income and Product Accounts (NIPA) back to 1978. The NIPAs change in 1993-94 to use the PPI for medical care rather than the CPI, however, has not yet been incorporated in these historical revisions. (Gordon, 2000), citing Oliner and Sichel, estimates that this creates an understatement of 0.09 percent annually in the growth of real output from 1973 on.⁴⁴ Although this figure demonstrates that the treatment of medical care can have non-trivial consequences for the overall indices, there are in fact many other causes of potential overstatement in the medical care price indices that this figure does not consider.

Use of list prices and the spread of managed care. Hospitals and physicians have a list price or customary charge, but few American payers pay on this basis. This is especially true since the advent of managed care in the United States, because managed care plans virtually never pay list price. Between 1987 and 1995 the proportion of the American privately insured population enrolled in managed care plans rose on the order of 50 percentage points, from roughly one-quarter to three-quarters of the population (Gabel et al., 1989; Jensen et al., 1997).⁴⁵ In addition, much of the Medicaid insured population was shifted to managed care plans, as was about ten percentage points of the Medicare population.

Those enrolled in managed care plans pay less out-of-pocket if they use health care providers that are in the plan's network of providers. The plan can threaten to exclude individual providers from its network unless it obtains a favorable price, thereby increasing the elasticity of demand facing any provider and enabling the plan to reduce rents in provider prices. Indeed, one type of managed care, Preferred Provider Organizations, the most rapidly growing form of managed care, limits itself almost entirely to obtaining discounts from physicians and other providers (McDonnell and Fronstin, 1999).⁴⁶

An accurate price index, of course, should use the prices actually paid, or transaction prices, rather than list prices. One study of hospitals in California showed that the use of list rather than transaction prices caused a 40 percent overstatement in the hospital price component of the CPI in the 1983 to 1988 period, or seven percentage points per year (Dranove et al., 1991). This was a time when managed care and thus discounting was spreading in California, so it was particularly important to use transaction prices. Hence, the seven percentage point annual bias is more in the nature of a one-time bias; nonetheless, it is clearly a substantial bias for those years.

Another study showing the importance of measuring transaction prices is prices paid Massachusetts hospitals for the care of heart attack patients. Prices for patients

⁴⁴ See Gordon's note 8 on page 55.

⁴⁵ Consistent with these values, unpublished data from the Bureau of Economic Analysis suggest that between 1992 and 1997 there was a 38 percentage point increase in insured private workers in managed care plans.

⁴⁶ I estimate that less than 20 percent of the population has insurance that pays something approximating list price. This estimate is based on the 21 percent of employment-based insurance that is indemnity insurance plus individual insurance; see (McDonnell and Fronstin, 1999).

enrolled in Health Maintenance Organizations were on average 42 percent less for the treatment of heart attacks than for those with indemnity insurance, whose insurers paid charges or list price (Cutler et al., 2000b). And the intensity of treatment was essentially the same, so the savings in total cost was also around 42 percent.⁴⁷

This magnitude of difference between list price and transaction price, if it is typical for all diagnoses in the nation, is clearly important. And a rough calculation shows that it may be typical because the 42 percent figure for Massachusetts heart attacks approximates the following calculation of the effect of managed care on total American medical costs. From the 1940s on, annual real per capita increases in American medical care spending averaged around 4 to 5 percent in each decade (Figure 3).⁴⁸ During the 1993-1997 period, however, the rate of growth fell to around 2 percent per year (Figure 4). The low rate of growth during this five-year period is unprecedented in the past half century. Although one hesitates to attribute the change entirely to the spread of managed care, it is hard to identify any other factor that could be responsible.

More specifically, the annual rate of real spending increase from 1940 to 1998 was 4.42 percent, whereas in the 1993-1997 period it was 2.06 percent. This difference of 2.4 percentage points, compounded over five years, amounts to 12.4 percent. One can compare this with the Massachusetts number by making an assumption about the percentage point increase in the managed care population and the percentage of spending to which it would apply. For the percentage point increase in managed care, I have used the 38 and 50 percentage point figures cited above.⁴⁹ Although they are not from comparable time periods, it is unclear what the appropriate lag is, and for my purposes a very approximate figure will suffice. For the spending percentage, I have used the percentage of spending accounted for by private payers in 1995 (56 percent) plus 9 percent of Medicare spending (representing the proportion of the Medicare population in managed care in 1995 (Physician Payment Review Commission, 1997)) and the portion of Medicaid spending that is not chronic long term care (75 percent); this yields a figure of 69 percent (Cowan et al., 1999). Using these figures, a 42 percent savings from managed care, the Massachusetts heart attack figure, would translate to an 11-14 percent reduction in overall spending ($11-14 = 0.38 \text{ or } 0.5 \times (.69) \times 42$). Of course, the 12.4 percent reduction in spending nationally could be attributable to reductions in quantities as well as price. And it is certainly coincidental that these two independently determined values, one a highly inexact calculation, turned out to be so close to each other. Nonetheless, their similarity suggests that managed care may have had an important effect on transactions prices.

The Bureau of Labor Statistics attempts to use transaction prices in constructing the American CPI. The problem lies in implementation. The Bureau has a difficult time

⁴⁷ Not surprisingly, since treatment was essentially the same, outcomes were essentially the same.

⁴⁸ Data from 1960 to 1997 for the G-7 countries show similar rates of change among the countries if one deflates data on nominal spending in dollars shown in (Anderson and Poullier, 1999) by the US GDP deflator.

⁴⁹ The relevant percentage point figure for the Medicare population is much less, but the figure for the non-elderly Medicaid population is probably greater.

obtaining transaction prices from hospitals and physicians. It relies on voluntary reporting, and hospitals and physicians fear that the prices they have negotiated with managed care plans may become known to their competitors if they report them to the Bureau of Labor Statistics. As of 1996, only 15 percent of the hospital quotations in the American CPI were based on transaction prices (U.S. Government Accounting Office, 1996).⁵⁰ There are also difficulties obtaining transaction prices for physician services and drugs.

In light of the potential fall in transaction prices in the 1993-1997 period from the spread of managed care, it appears as if the bias in the medical component of the CPI may have increased by up to two percentage points or more during that period, an important change in the bias.⁵¹ Increasing the proportion of transaction prices in the indices seems like a high priority.

Substitution bias, the frequency of reweighting, and the treatment of new goods. As noted above, the official indices are Laspeyres indices; that is, they price the market basket that consumers bought in the initial period. It is well known that this induces an upward bias in the index, because consumers can generally achieve a higher level of utility by substituting against goods whose relative prices have risen, but Laspeyres indices do not measure this type of substitution (Boskin, et al., 1996), (Boskin, et al., 1998). (Boskin, et al., 1996) estimated that this substitution bias increased the overall American CPI by 0.4 percentage points per year.⁵² This estimate reflected improvements made in 1995 and 1996 that BLS and the Commission estimated reduced substitution bias by about 0.25 percentage points per year.

In an attempt to reduce substitution bias further, starting in 1999 the Bureau of Labor Statistics began using the geometric rather than the arithmetic mean for several elementary aggregates in the CPI, but not in the professional services and hospital portion of the Index.⁵³ Moreover, for aggregation above the elementary aggregates the assumption of fixed weights and any associated substitution bias was maintained.⁵⁴

⁵⁰ According to Dennis Fixler of the Bureau of Labor Statistics, however, by 1997 transaction prices comprise a majority of the hospital prices in the Index.

⁵¹ The bias in the GDP deflator was probably less, though how much less is hard to know. The GDP deflator for hospitals used the wage index for the 68 percent of hospital spending in not-for-profit hospitals. Although a wage index is problematic as a price index (see above), it is not subject to the omission of the potential fall in markups attributable to the spread of managed care. For the 32 percent of spending in for-profit and public hospitals, the GDP deflator used the Producer Price Index. Although the PPI also intends to measure transaction prices, it is unclear how successful it is in doing so. The GDP deflator also used the PPI for services of physicians, which was also intended to measure transaction prices. It is not clear that the PPI should be any more successful than the CPI in measuring transaction prices, but for whatever reason, the PPI for physician services only increased 7.2 percent from 1994 to 1997, whereas the CPI increased 11.6 percent.

⁵² The Commission estimated that the upper-level substitution bias was 0.15 percent. This interpretation is non-controversial, but the remainder, which the Commission attributed to lower-level substitution bias, could include other elements.

⁵³ The geometric mean was used within the category of drugs and medical supplies.

⁵⁴ For example, the bias from the substitution of Delicious for Granny Smith applies if the relative price of Delicious apples fell was in principle eliminated by using the geometric mean within the elementary

(Busch et al., 2000) estimate the consequences for a price index for the treatment of depression from using arithmetic rather than geometric weights. Their estimate was a bias of 0.7 percent per year, so the substitution bias appears to be present in the medical indices as well.⁵⁵

To the degree that the change in relative prices increases as the market basket ages, the frequency of reweighting is also a critical issue. Historically, both the American PPI and CPI were reweighted every ten years, but both have now moved to more frequent reweighting, every four and five years respectively, another recent improvement in the indices. The American GDP deflator is now chain-weighted, meaning the market basket changes every year.

But there is an additional delay in reweighting both the CPI and the PPI, because it takes time to analyze and implement the survey of spending patterns. In practice, a new market basket has been introduced approximately four years after the date of the spending survey. Thus, the actual age of the market basket in the American CPI and PPI has historically varied between four and fourteen years and has averaged nine years. With a survey now being done every four or five rather than every ten years, the age of the market basket has fallen to between four and nine years, assuming the four-year implementation lag remains unchanged.⁵⁶

Infrequent reweighting also exacerbates the problem of introducing into the index the new goods that in reality appear almost daily. Consider a new good that appears after the initial period in which the weights are established. For the purposes of the price index, one would like to value the new good at the price consumers would have paid for it, had it been available in the initial period and, in a Laspeyres index, using the hypothetical budget share weight from the initial period (Fisher and Shell, 1972). Such a price and share, of course, are not observed. In practice, therefore, the official indices commonly “link” the new good in at the average of the index, meaning that in the period in which the new good enters the index, the index is unchanged. This approach essentially ignores the new good until it officially enters the market basket and then accounts for price changes in the good from that point in time forward. I take up other methods for valuing new goods below but here focus only on the lags in accounting for them in the indices.

Unfortunately there is an inevitable delay in recognizing that a new good has in fact entered the market. Prices of new goods often fall dramatically after their initial introduction, as production runs increase and economies of scale are realized. As a result, the lags in incorporating new goods into the indices can cause serious upward bias

aggregate of apples, but the bias from the substitution of food for transportation if the relative price of transportation rose was not.

⁵⁵ The 0.7 percent per year figure ignores the use of the geometric mean within the drug and supplies category, but this would probably little affect the estimate because the main substitution relevant to the Busch study is between anti-depressant medication and psychotherapy.

⁵⁶ The Bureau of Labor Statistics is attempting to reduce this lag.

by failing to measure these price declines. Many new goods did not appear in the American CPI for years after their appearance in the marketplace. For example, personal computers, microwave ovens, and VCRs all appeared in the American CPI more than a decade after they became available for purchase. By the time they were introduced into the Index, all of their prices had fallen by more than 80 percent since their initial appearance in the marketplace (Boskin, et al., 1998). Cellular telephones were not introduced into the American CPI until 1998, and their price fell by more than 90 percent between 1989 and 1996 (Hausman, 1999).

The medical indices are not immune from these issues. In our study of heart attack treatment, we showed that the frequency of reweighting matters a great deal (Cutler, et al., 1998). Using a 1983 market basket and keeping that market basket unchanged over the entire 1983-1994 period, we found an annual rate of price increase for heart attack treatment of 2.8 percentage points above the GDP deflator. If we reweighted the market basket every five years instead of continuing to use the 1983 market basket, the measured rate of price increase fell to 2.1 percent. And if we reweighted annually, the measured rate of increase fell to 0.7 percent. In other words, reweighting annually rather than maintaining fixed weights for an eleven year period decreased the real inflation rate in heart attack treatment by a factor of four.

The importance of new goods and the potential lags in introducing them into the indices was also shown in the heart attack study. Even allowing for the equivalence of minor changes in inputs, we could only match 78 percent of the (dollar-weighted) hospital inputs used in 1994 with those used in 1983.⁵⁷ Over a five-year interval we could match 85 percent of the inputs. If we linked the indices annually, however, we could match 98 percent of the inputs.

Some evidence on the bias from deficiencies in measuring new goods comes from the drug price index of the American PPI. (Berndt et al., 1993) and (Berndt and Greenberg, 1995) show that the Index prior to 1995 oversampled older drugs, perhaps because of a failure to recognize that a new drug had entered the market. This caused an upward bias in the Index, because prices of older drugs tended to increase more rapidly than those of newer drugs. (Berndt, et al., 1993) and (Berndt, et al., 2000b) estimated that the oversampling of older drugs caused as much as a 3 percentage point per year bias in the measure of prescription drug prices in the 1986-1991 period. Later studies by the Bureau of Labor Statistics presented corroborating evidence; see (Kelly, 1997) and the discussion in (Berndt, et al., 2000b). As a result, the Bureau has now supplemented its sample of drugs that are in the market basket in order to address this problem, another recent improvement in the Index.

Generic drugs provide an example of bias from the method for introducing new goods into the CPI. Generic drugs come on the market after the patent protection for a

⁵⁷ That is, we could account for only 78 percent of the spending on inputs if we restricted 1994 inputs to those available in 1983. If we did not allow for minor changes and required strict equivalence, we could only match 27 percent of the inputs over an 11 year period and 42 percent over five years. We could still match 98 percent of the inputs annually, however.

brand name drug expires. They are the same chemical entity as the branded drug, although the branded drug retains its copyright on the brand name.⁵⁸ Before 1996 the American CPI considered a new generic drug as a new good and linked it in at the average of the Index. Subsequent to 1996 it has treated the generic as equivalent to the branded drug, relying on the Food and Drug Administration's certification of therapeutic equivalence. This change represents another recent improvement in the Index.⁵⁹ Because generic drugs typically enter the American marketplace selling at a one-third discount, a discount that typically rises substantially as more generic producers enter the market, the previous practice of simply linking in the generic drug as a new good missed the initial price fall to consumers who switched to the generic version of the drug.

(Griliches and Cockburn, 1994) assumed that consumer tastes for generic and branded drugs were uniformly distributed. Using this assumption they calculated that the change in a true price index for two antibiotics over a period in the late 1980s was very different than the change in an index that approximated the methods of the official PPI. For cephalexin (Keflex), Griliches and Cockburn's preferred index after three years was 46 percent lower than one that mimicked the BLS procedure of simply linking in the generic drug, an 18.4 percent per year difference. For cephadrine (Velosec) the analogous figure was 21 percent (7.6 percent per year). (Berndt et al., 1996) extended this work and showed that for anti-depressant drugs, one of the largest classes of drugs, the procedure of simply linking in the generic drug led to a 3.5 percentage point overstatement of the annual rate of increase in price.

An indication of something amiss. As noted above, beginning in the 1980s the managed care revolution changed the financing of medical care in the United States. Much (many would say most) of the effort of managed care was directed at obtaining lower prices from hospitals, physicians, and other medical providers. Arguably these efforts reduced rents that these providers had been receiving. One would therefore have expected that, relative to earlier decades, a medical care output price index would have fallen relative to an overall price index.⁶⁰ As Figure 5 shows, however, the opposite was in fact observed. The difference between the medical care index and the overall index actually rose in the 1980s and 1990s, especially relative to the 1970s. The data in Figure 5 strongly suggest something is, or at least was, amiss with the American CPI for medical care. A similar conclusion applies to the medical care component of the GDP deflator.⁶¹

The Medical Care Sector and Price Indices: Problems with Applying the Standard Theory and Its Assumptions to Medical Care

⁵⁸ For example, the generic drug for the brand drug Motrin® is ibuprofen.

⁵⁹ The PPI has revised its series back to 1978 to incorporate this change.

⁶⁰ This assumes medical care specific changes in unit costs were reasonably similar across decades.

⁶¹ The values for the annual excess of the medical care component over the Personal Consumption Expenditure component by decade, starting in the 1950s and ending with the 1990-98 period are respectively (in percentage points): 0.8; 1.4; 0.8; 2.5; 1.6. Although the excess is smaller than is the excess of the CPI, there is still a substantial contrast between the periods before and after 1980. I think Clint McCully for these figures.

Medical advances and technological change. The problem of accounting for changes in the quality of products is a well-known issue with price indices. Let an improved product be considered a new good. The official indices take five approaches in incorporating a new good into the index, once it is recognized in the market basket. As described above, a common approach is to link in the new good at the average value of the index and then track the change in price in subsequent periods. Even if there were no lag in introducing the good into the index, this approach ignores the gain to consumers from having a new or better quality good available in the first place. This can lead to dramatic overstatements of the change in the cost of living, defined as the income necessary to remain on the same indifference curve. (Nordhaus, 1998), for example, has calculated that the price of a constant amount of lighting over the past two centuries, relative to a procedure that mimics that of the official price indices, is overstated by a factor of 1000 (!) because of linking in new lighting technology and the resulting failure to incorporate technological advances in the amount of illumination. This is an annual bias of 3.6 percentage points. An approach with the same effect is the so-called addition-and-deletion approach, under which the old good is dropped from the index and the new good added in a fashion that leaves the overall index unchanged.

A third approach, which is usually used for “minor” improvements in an existing good, is simply to ignore the change in the characteristic and price the new good as if it were the old good. Any price increase that is attributable to better quality is thus ignored and treated as a pure price increase.⁶² To the degree products are improving, this approach also leads to an overstatement in the change in the cost of living. Because minor improvements in goods are being made all the time, the cumulative overstatement may be substantial over a period of years.

Two other approaches are less common. A hedonic approach estimates market prices for key attributes of the product that has been improved and then uses these estimates to adjust observed prices for quality change ((Griliches, 1988), chs. 7,8; (Berndt, 1991), ch. 4). In the case of computers, for example, one can estimate a market price for attributes such as clock speed of the chip, RAM, and hard drive capacity by observing the price of computers with different configurations.⁶³ One can then adjust the price of a newer model computer for improvements in these dimensions; that is, one can decompose a price change into the portion of the change attributable to these improvements and a pure price change.⁶⁴

Where feasible, hedonic adjustment is clearly desirable since it follows directly from the theory underlying the indices. Unfortunately, the spread of health insurance has led to the widespread use of administered prices in medical care, so that hedonic pricing has limited applicability to medical care.

⁶² Similarly, any price falls are understated.

⁶³ In the computer example, the market price for each of these dimensions has fallen dramatically over time and consumers have bought more powerful machines, leading to a standard type of substitution bias from using initial period weights.

⁶⁴ See (Oliner and Sichel, 2000) and (Gordon, 2000) for a further discussion of price indices for computers and their role in overall measures of productivity.

Finally, the American Bureau of Labor Statistics sometimes directly values quality improvements and adjusts the observed price accordingly, but this is rare. In my view the pace of technological change in medicine is so rapid and the advances so important to the health of consumers that the usual approaches of either linking in the new good or simply ignoring the product changes lead to substantial overstatements of the rate of increase in the cost of living. If hedonic adjustment is not feasible, that leaves direct adjustment, which, to preview my conclusion, is what I believe should be done.

As an example of the rapid advances in medicine, several studies have quantified the role of medical treatment in recent substantial reductions in mortality from heart attacks (acute myocardial infarctions), which have occurred in many countries. Between 1975 and 1995, for example, the American 30 day mortality rate from heart attacks declined from 22 to 12 percent, or nearly a factor of two. (Cutler, et al., 1999) combined data from clinical trials that established the effect of various therapies on heart attack mortality, together with data on changes in the use of those therapies between 1975 and 1995, to determine that about half of the decrease in the American mortality rate was attributable to new knowledge or new procedures. The largest single factor was the greater use of aspirin following a heart attack, which accounted for about a quarter of the total change in mortality.

(Hunink et al., 1997) addressed the same question with a different method. They used a model of factors influencing coronary heart disease mortality, of which heart attacks are the major component, to explain the change in mortality between 1980 and 1990 in the United States. Coronary heart disease accounts for about a third of all deaths in the United States. They found that mortality from coronary heart disease in 1990 was 34 percent below what would have been predicted if risk factors such as smoking and blood pressure, event rates (i.e., whether a coronary event such as a heart attack occurs conditional on risk factors), and case fatality rates (i.e., whether death occurs conditional upon an event) had all remained constant at 1980 levels. Nearly half the decline was attributable to the third factor, lower case fatality rates for heart attacks. Most of this decline was probably attributable to better medical and surgical management. A quarter of the decline was attributed to better primary prevention (i.e., better management of risk factors in those who had never had any prior history of disease). Even this quarter, however, was partly attributable to medical interventions such as cholesterol-lowering and anti-hypertensive drugs. The remainder was attributable to better secondary prevention (better management of risk factors in those who had already had a coronary event in 1980); a portion of that could similarly be attributed to medical care. Overall then, over half of the decline in coronary heart disease mortality between 1980 and 1990 as assessed in this study was probably attributable to technological change in medicine, a figure consistent with the (Cutler, et al., 1999) work just cited, as well as with some of our other work discussed below. An earlier study (Goldman and Cook, 1984) found that a smaller fraction, about a third, of the improvement in coronary heart disease mortality in an earlier period (1968-1976) was attributable to better medical and surgical management of coronary heart disease. In large part because of the reduced mortality from coronary heart disease, there has been a 6 year gain in life expectancy in the past 30

years in the United States (8 percent) and a 2.6 year gain (a 17 percent gain) in life expectancy at age 65 (U.S. Department of Health and Human Services, 2000).

But the gains from improved medical care affect not only life expectancy but also the quality of life.⁶⁵ Improvements in artificial hips and knee joints, intra-ocular lenses, and anti-depressant and anti-inflammatory medications are but a few examples of widely used advances whose primary effect is on the quality of life rather than on mortality. Reflecting such changes, (Cutler, 1998; Cutler and Richardson, 1997) have shown that individuals with several different diseases or conditions rate their overall health substantially better in 1990 than they did in 1970 (Figure 6). Moreover, the pattern of improvement across diseases is plausible; cardiovascular treatment, for example, has changed substantially more than cancer treatment.

Another, perhaps less obvious gain in the quality of life stems from the improvement in surgical techniques from learning-by-doing and especially the development of minimally invasive techniques that have permitted the shift to outpatient surgery described above. These improvements mean that many procedures are now carried out on patients where the risk of operative mortality would have previously precluded the operation. Figure 7 shows a striking difference in the rate of increase in several common medical procedures among those over 85 years of age and those between 65 and 69 between 1987 and 1995. This difference presumptively stems from physicians' increased confidence in their abilities to succeed in older, more clinically risky patients. In a price index context, this type of advance appears as a change in the quantity of services – indeed such changes partially account for increased medical spending – and thus is not incorporated in the official price indices at all. In many cases, however, the changed technology can lead to a dramatic change in the quality of life. For example, an angioplasty in an elderly patient with chronic angina pectoris may mean the difference between being bedridden and being ambulatory. And, as already noted, the faster recuperation time from less invasive techniques is also a gain.

The theory of the official indices and gains in health. The standard theory of price indices is based on a utility function that is defined on goods and services that are traded in the marketplace. Health, of course, is not directly traded in the marketplace, but there can be little doubt that consumers place a high value on their health. In many languages the customary toast refers to wishing for good health. Hence, I will modify the usual utility function by including health as an argument, in addition to including goods and

⁶⁵ (Gordon, 2000), argues that the Second Industrial Revolution, especially the development of electricity and the internal combustion engine in the 1860-1900 period, overshadows the Third Industrial Revolution, especially the development of information technology in the latter half of the 20th century. In so doing, he downgrades the importance of the medical advances of this period, noting that the rate of increase in life expectancy at birth was three times as large from 1900 to 1950 as from 1950 to 1995. Although his reference to life expectancy is more in the nature of an aside, it misses the point that many medical advances have had important effects on the quality of life but not life expectancy. Similarly, life expectancy from 1950 to 1970 was not much changed, perhaps for reasons having little to do with medical care, such as the great upsurge in smoking in the first half of the century, with a lagged effect on mortality, as well as potential less healthy diets. Some of the increase in life expectancy since 1970, of course, can also be attributed to lifestyle changes, such as the fall in smoking rates.

services that are traded in the marketplace. Doing so has important implications for the measurement of changes in the cost of living.

One can ask whether including a non-marketed good in the utility function is really necessary; for example, consumers primarily value automobiles for transportation and food for (among other things) nutrition, but economists do not normally include transportation or nutrition as arguments of the utility function, at least for the purposes of constructing price indices, a treatment with which I concur. Why, then, should one base price indices on a utility function with health as an explicit argument? There are several reasons for doing so.

Most importantly, appropriately valuing changes in medical technology in a cost-of-living index requires valuing their effect on health. In the case of medical care this is perhaps most obvious if the consumer purchases an insurance policy that covers the cost of all medical services. The only observed price of medical care in the consumer's market basket is thus the premium.⁶⁶ Indeed, some have proposed that the medical care price index should be based on premiums (Reder, 1969), (Pauly, 1999).⁶⁷ Suppose premiums increase because of a costly medical advance. To ignore the health benefits of the change is to assume that the increase in cost is pure inflation and, in the standard index, welfare decreasing. But if the change is welfare decreasing, one is hard put to explain why the advance was adopted in the first place. Certainly no one is seriously proposing that we would be better off returning to 1950s medicine even if we could pay 1950s prices. In other words, in this context a measure of the value of any change in quality requires one to value the change in health.

Second, the standard theory assumes that prices reflect the value of the resources elsewhere in the economy, but the presence of health insurance means that an informed consumer will value the marginal unit of medical at less than its opportunity cost elsewhere in the economy. In short, the standard theory of price indices when applied to medical care assumes no moral hazard, but there is compelling evidence that more extensive insurance coverage increases the demand for services (Newhouse, et al., 1993), (Zweifel and Manning, 2000). Any gains in health are therefore not properly valued using observed prices that the consumer pays at the time of service.

Third, a principal good the consumer buys from the medical care sector, and especially from physicians, is information (Arrow, 1963). A diagnosis is usually the physician's first order of business when seeing a sick patient, and in many cases there may be multiple treatment options. As a result, the incentives of the physician to provide accurate information are relevant, which is not the case in the standard theory underlying the price indices, where the consumer is assumed to have full information. In short, in

⁶⁶ One could in principle observe the prices that the insurer pays hospitals, physicians, and other medical providers, which is in effect what the Consumer Price Index does now, treating any change in retention by the insurer as a change in price. In a world, however, in which the insurer can deny services as not medically necessary and in which there is no ready way for the consumer to express willingness to pay at the point of service (i.e., when seeking care), the conventional theory clearly breaks down. See below.

⁶⁷ See (Jensen, 1990) for an effort to implement this approach.

medical care there are problems of agency, implying that observed purchases may not always be an accurate measure of consumer welfare (McGuire, 2000). Because of possible non-price rationing, managed care and public production of services heighten these concerns.

For all these reasons I think it is better to write the utility function with health as one argument and non-medical related consumption as another argument.⁶⁸ For simplicity I assume in what follows that consumers do not value medical services other than for their direct effect on health. This assumption can be relaxed without changing the nature of the results below, but to do so would needlessly complicate the mathematics.⁶⁹

Thus, for the i^{th} consumer who might contract one of several diseases d (perfect health could be $d = 1$) one can write expected utility at time t as:

$$U_i(t) = \sum_{d=1}^D \pi_d(t) \cdot U_i(H_i(d, m_d(t)), Y_i - p_i(t) \cdot m_d(t) - T_i(t)), \quad (1)$$

where $\pi_d(t)$ is the probability that the person has disease d at time t , H is health, m_d is the vector of medical treatments received for disease d , Y is income, p is the unit price paid for the medical services (net of insurance), and T represents any taxes or premiums paid to finance a medical care insurance policy. The second argument is thus non-medical consumption. In the standard theory only the second argument, including medical consumption, appears in the utility function.

In this section I wish to focus on the value of medical advances to the consumer. To bring out the nature of the problem for price indices, I simplify by assuming first a representative consumer and second that there is only one disease which everyone contracts. I thus drop the i and d subscripts and rewrite the utility function as:

$$U(t) = U(H(m(t)), Y - p(t) \cdot m(t) - T(t)). \quad (2)$$

⁶⁸ I place health as an argument in the utility function in order to be able to value directly the effects of changes in medical technology on health. Using health as an argument in the utility function, however, raises two other issues. First, should exogenous changes in disease (e.g., the appearance of AIDS) change the cost of living? Second, suppose the consumption of non-medical goods that affect health, such as tobacco, changes. Should the consequences for health affect a cost-of-living index or should one assume that the consumer has accounted for those consequences in deciding to smoke? These are difficult issues that are outside the scope of this paper. For some discussion, however, see (Boskin, et al., 1996) and on the specific issue of tobacco see (Cutler et al., 2000a).

⁶⁹ In the more general case in which one allowed medical goods and services to affect welfare other than through their direct effect on health, one would simply keep track of medical spending throughout the indices and would allow for a term in which medical spending affected the utility of the consumption of other goods and services. In some cases medical services will directly affect other aspects of utility. Examples include some cosmetic surgery and some uses of “lifestyle” drugs such as human growth hormone and Viagra (but see (Keith, 2000)).

With this assumption and the assumption that income and other prices are constant, changes in welfare are a function only of changes in medical treatments over time and their costs.

We wish to find a value C that reflects the change in the cost-of-living between two periods when the vector of medical treatments m has changed. Letting subscripts reflect time periods zero and one, C will equalize utility between the two periods if it satisfies the following equality:

$$U(H(m(t_1)), Y - p(t_1) \cdot m(t_1) - T(t_1) - C) = U(H(m(t_0)), Y - p(t_0) \cdot m(t_0) - T(t_0)). \quad (3)$$

Taking a second-order Taylor Series expansion around t_0 and using x to represent the second argument or non-medical consumption, one derives the following expression for C :

$$C = \frac{U_H H_m}{U_x} dm - d(p \cdot m + T) + (U_{xx}/2)d(p \cdot m + T)^2 \quad (4)$$

The first term on the right hand side of this equation represents the health benefits of the change in the vector of medical services, normalized by the marginal utility of consumption. The second term is simply the change in medical spending, and the third term reflects the change in absolute risk aversion from any changes in the variance of out-of-pocket spending. In what follows I will ignore the last term because widespread health insurance will often make it negligible.

Relative to the standard theory behind the official indices, equation (4) has two important changes. First, any change in health status from a new vector of medical goods or services now results in a changed standard of living. Second, the change in the standard of living is a function of the change in the amount spent on medical services, not just the change in the unit price of medical services. The reason is that it is the total spending on medical services and not just the price of medical services that affects both health status and non-medical related consumption.

The first term in equation 4 implies that in a cost-of-living index based on this utility function one must value the health benefits from improved medical care. Three approaches to valuation have been taken in the literature. One is the hedonic approach described above, which I believe has limited applicability to medical care because of widespread insurance.

A second approach is to assume a structure of preferences for consumers among different goods, an approach that (Griliches and Cockburn, 1994) used in their study of generic drugs referred to above.⁷⁰ The results of this approach, however, obviously

⁷⁰ See also (Fisher and Griliches, 1995).

depend upon how good an approximation the assumption about preferences is, and there is no way to test that assumption.⁷¹ In effect, one is decomposing any price change into a portion that satisfies the assumed preferences and a pure price change. I therefore do not find this approach appealing, but in any event the lack of an empirical basis for the assumption about preferences clearly poses problems for its adoption in an official index.

The final approach, direct valuation of the improvement, was used by (Cutler, et al., 1998), when they first estimated and then valued the lifesaving gains from improved heart attack treatment. This study illustrates both the difficulties and the opportunities in directly valuing gains in improved health from medical advances.

Attribution of a change in mortality to improved medical care is difficult because one observes only the overall decline in mortality, and it may fall for many reasons other than medical care. In the case of heart attacks, for example, mortality may fall because of better diet or improved response times of emergency services, perhaps because of a better transportation network.⁷² In short, one must estimate how much of any observed overall decline is attributable to improved medical treatment. Second, one must estimate how long victims who would have died but for the improved treatment will survive. Heart attack patients have lost heart muscle, which impairs the pumping function of the heart and thus shortens their life expectancy relative to a person who has not had a heart attack. Thus, if one simply used life table values for mortality to calculate life expectancy following a heart attack, one will overestimate the gains from the improved treatment.⁷³ Finally, one must attach a dollar value to the improvements in life expectancy. In the lifesaving case, one is helped by the literature on the value of life (Viscusi, 1992), (Viscusi, 1993). Even here, however, there are difficulties if the quality of life that is added is compromised in some fashion by the illness, meaning that medical treatment can prolong life but cannot restore full functioning. This is an example of the much more general problem of valuing improvements that are not lifesaving, such as fewer side effects from a newer drug or improved mobility from artificial joints. Furthermore, the value may depend upon whether an individual has other diseases (comorbidities); that is, improvements across various diseases may not simply be additive. For this range of issues the literature is less developed and sometimes non-existent (Tolley et al., 1994). Despite the sparse literature, it is clear from everyday observation and their widespread adoption in all developed societies that the nature of medical advances in the last several decades has been of enormous value.

⁷¹ The difficulty with hedonic analysis in analyzing drug prices is that generic and brand name drugs are therapeutically equivalent so it is not obvious what the consumer is buying when paying the higher price of the branded drug. Branding might reduce search costs for consumers, but even if that were the case, there is no observed price for varying search costs that is analogous to the prices of computer chips with varying processing speeds. In short, it is not clear how one can value the characteristics that those who buy the brand name drug are buying using prices of observed characteristics.

⁷² Or mortality may fall in spite of clogged highways that worsen response times. See (Athey and Stern, 2000) for evidence that improved emergency response systems lowered cardiovascular mortality in Pennsylvania.

⁷³ An additional problem is that mortality rates have been improving over time, but lifetable values often assume the current values for each cohort will apply to future cohorts when they reach that age.

(Cutler, et al., 1998), in their study of heart attack care among the elderly, using a conservative \$25,000 figure as the net (of consumption) value of a life year, estimated that a cost-of-living index for heart attack treatment fell 1.1 percent from 1984 to 1991, although a calculation that mimicked the methodology of the official index showed the index rising by 2.8 percent, nearly a 4 percentage point difference from accounting for the mortality gains.⁷⁴ If one used a somewhat less conservative value of \$50,000 per life year, the estimated fall in the cost-of-living index was 3.1 percent annually, a full 6 percentage point annual bias in the index.

And even \$50,000 is a conservative valuation of a life-year. A 40 year old person in a developed country has roughly a 40 year life expectancy. At the \$25,000 per life year value that Cutler et al. used, this would translate into a value of life of \$1 million if undiscounted. Discounted at 3 percent, the value would be only about \$578,000, and it would be \$1.156 million if the \$50,000 per year figure is used. These figures are a factor of 3 to 10 below the values of \$3 to \$7 million in the value-of-life literature (Viscusi, 1992), (Viscusi, 1993).

To my knowledge no other calculations have been undertaken that attempt to quantify the bias in the medical price index by directly valuing changes in outcomes.⁷⁵ As a result, it is impossible to know to what degree the values for the heart attack case generalize to other diseases. Because of the dramatic fall in cardiovascular disease mortality rates referred to above, which has not been duplicated for other common causes of death, it may well be that the adjustment for cardiovascular disease is high relative to many other diseases. Nonetheless, improvements in the quality of life for diseases that do not kill but do incapacitate, such as major depression, may have similar value.

In my view incorporation of the gains in health from improved medical care into the price indices is an opportunity for substantial improvement in the price indices. Realizing this improvement, however, will require input from disciplines not commonly represented in the government agencies responsible for calculating the official price indices or will require that these agencies co-operate with other agencies where the appropriate expertise can be found. Moreover, given the room for judgment in implementing such an approach, it is best carried out initially as an experimental index; that is, the methods of the current official indices should be maintained. Over time one would hope that techniques for estimating the value of technological change in medicine would improve; if so, perhaps at some point the experimental index could be incorporated into the official index.

As (Triplett, 2001) points out, however, this approach to medical price indices will also require a change in the system of national accounts or market basket that underlies the calculation of medical care price indices. Rather than price indices for specific medical care goods or services, such as hospital services, there would be a price

⁷⁴ Cutler, et al. attributed about half the gain in mortality among the elderly to improved medical treatment, an independent confirmation of the Hunink et al. and the other Cutler et al. calculations discussed above.

⁷⁵(Frank et al., 1999) discussed in the next section, attempt to hold outcome constant in and adjust for changes in the cost of inputs.

index for the medical care sector based on the costs and consequences of the treatments for various diseases. For example, instead of hospital stays and physician visits entering the medical care index, the treatment of a heart attack or a middle ear infection would enter the index with weights proportional to the money expended on that disease. Adopting this approach would either require the assumption that medical care goods and services were only valued for their effect on health or, alternatively, would require valuation of any other effects on welfare. I return to this proposal in the conclusion.

Off-frontier production. Although medical care shares the characteristic of rapid technological change with many other sectors, for example telecommunications and computing, it appears to have much more production that is inside the production-possibility frontier than economists would commonly suppose competitive industries would exhibit. Off-frontier production can take the form of overuse, underuse, or misuse of medical services (Chassin, et al., 1998). Although many studies have documented the existence of off-frontier production in medical care, two types of studies can serve to illustrate the magnitude of the problem.

The first are studies by clinicians of appropriate and inappropriate care in the United States, beginning with (Chassin, et al., 1987).⁷⁶ The clinicians carrying out these studies defined a procedure as appropriate if the expected benefit of the procedure exceeded its clinical risk by a sufficient margin to make the procedure worth doing. Cost was not explicitly part of this definition of appropriate, although it potentially creeps in implicitly through the words “sufficient margin.” Because cost was not explicitly mentioned in the definition, it seems likely that clinicians employing this definition would categorize many procedures as appropriate where the marginal benefit is positive but less than marginal cost, whereas an economist would consider such procedures inefficient. Hence, it is likely that economists would find this definition would understate the number of economically inefficient procedures. Nonetheless, for several common procedures the proportion the clinicians considered inappropriate in the initial American studies varied between a sixth and a third (Table 4).⁷⁷

These findings are not limited to the United States. In the Trent region of England, the indications for 51 percent of coronary angiographies and 42 percent of coronary bypass operations were considered equivocal or inappropriate, and in four Israeli hospitals the indications for 29 percent of cholecystectomies (removal of the gall bladder) were considered equivocal or inappropriate (Brook, 1993).⁷⁸ Such findings, first brought forward in the late 1980s, gave birth to a worldwide movement to develop guidelines for care. By illustrating the potential savings from greater efficiency, they also spurred the managed care revolution in the United States.

⁷⁶ See also (Leape et al., 1990) and (Newhouse, et al., 1993), chapter 5, for additional material.

⁷⁷ Coronary angiography is an imaging procedure to discover which, if any, of the coronary arteries may be occluded (blocked). Coronary artery bypass graft is the splicing of (usually) a piece of a vein taken from some other part of the body around an occluded coronary artery in order to restore the flow of blood to the heart. Carotid endarterectomy is a procedure to open the artery or arteries supplying blood to the brain to prevent a stroke.

⁷⁸ Brook cites (Pilpel et al., 1992).

A second genre of literature quantifies rates of medical error. An initial study carried out in New York State showed that about one percent of all hospital patients in 1984 suffered a negligently caused “adverse event” (Brennan et al., 1991). An adverse event is a medically caused error that that prolonged the hospital stay, or that resulted in disability persisting beyond the hospital stay or in death. That it was negligent meant it resulted from care that fell below professional standards. About a quarter of the negligently caused errors resulted in death in the study in New York State (Weiler et al., 1993). A subsequent study using the same methodology was carried out in 1992 in Colorado and Utah (Thomas et al., 1999). It found similar, though somewhat lower rates of error. Summarizing these studies, the Institute of Medicine estimated that about 2 to 4 percent of all deaths in the United States appear to be caused by medical errors in hospitals (Institute of Medicine, 1999).⁷⁹

These studies have not attempted to define the efficient level of error (it almost certainly is not zero), but in light of the frequent serious consequences of error, it seems likely that the efficient level is lower – probably much lower – than the present level.⁸⁰ This view is supported by the current movement to install computerized drug ordering systems to reduce the rates of drug errors that the literature has documented (Bates et al., 1995; Leape et al., 1995) and by a special appropriation by the American Congress following the publication of (Institute of Medicine, 1999) to find ways to reduce error.⁸¹

The current methods of the official price indices will not capture the value of improvements in the reduction of inappropriate procedures or the rate of medical error because they assume that all care delivered is on the production-possibility frontier. It follows that if the proportion of off-frontier production is falling, price indices are biased up.

Ernst Berndt, Richard Frank, and Susan Busch have attempted to construct price indices for the treatment of acute depression that account for reductions in the rate of off-frontier production (Berndt, et al., 2000a), (Berndt, et al., 2001b), (Busch, et al., 2000), (Frank, et al., 1999). In an initial effort they used treatment guidelines published by the Agency for Health Care Policy and Research (now the Agency for Healthcare Research and Quality) and the American Psychiatric Association to identify treatments that were on an efficient frontier. They interpreted treatment that met the guidelines as efficacious and treatment that did not as inefficacious. They then determined how the price of

⁷⁹ See also (Leape, 1994).

⁸⁰ By contrast with the one percent negligent adverse event rate (and the four percent overall adverse event rate) among hospitalized patients, there are on the order of 10,000,000 aircraft takeoffs and landings each year, but only about 4 crashes. Air travel was not always so safe, however; by 1910 there were 2,000 pilots in the world, and 32 had already died. In an effort to meet delivery schedules in all kinds of weather, 31 of the first 40 Air Mail Service pilots were killed when the American air mail service was founded in 1918 (Leape, 1994).

⁸¹ Although it is likely socially efficient to install computerized drug ordering systems, it may well not be privately efficient for hospitals or other providers reimbursed on a fee-for-service basis to do so, because with poor information about the rate of errors, patients do not necessarily avoid hospitals with higher error rates (and hospitals lose the revenue from the additional admissions to deal with the consequences of the errors). As a result, several large American employers who are members of the Leapfrog Group are now giving employees incentives not to use hospitals without such ordering systems.

treatment that met the guidelines varied between 1991 and 1996. In other words, they determined how the price of treatments that were on the frontier changed over time, as well as the change in the proportion of treatments on the frontier.

Over this five year period, the proportion of observations on the frontier rose from 35 to 55 percent (Busch, et al., 2000), (Berndt, et al., 2001b). Although a large amount of treatment remained off the frontier, the proportion off the frontier had markedly fallen, a fall the authors attributed principally to the spread of managed care. Using a strict definition of treatments that met the guidelines, they found a decline in the nominal price of treating the disease of over 30 percent, although using a looser definition and making other methodological improvements the decline was only 3 percent. Even the nominal 3 percent figure, however, meant a real price decline of about 3.3 percent per year.⁸² This finding is analogous to the (Cutler, et al., 1998) finding for heart attacks cited above and stands in stark contrast to the substantial increases in the official indices for medical services in this period.

The finding of large amounts of off-frontier production in all of these studies means there is potentially enormous scope for improved efficiency in medical care delivery and hence reduced cost-of-living. Berndt, et al. attempted to account for this kind of improvement in their most recent study of the treatment of depression (Berndt, et al., 2000a). Instead of comparing treatment for depression against guidelines, they had a panel of ten physicians and mental health professionals determine the likely distribution of outcomes for 120 patient-treatment cells.⁸³ The panel was asked the following question:

“Using your best clinical judgment, consider an average group of 100 patients presenting with a major depressive disorder...After 16 weeks of the indicated treatment, how many patients will most likely fall into each of the following categories: depression-free, mildly depressed, moderately depressed, or no change....”⁸⁴

The researchers were thereby able to obtain an estimate of both the likelihood of full remission (“depression free”), as well as the likelihood of a roughly 50 percent improvement (“mildly depressed”).⁸⁵

The next step was to cost the various treatment types using a large claims database drawn from the 1991-1996 period and then apply the panel’s probability estimates to the same database. This method accounts for both changes over time in the

⁸² This calculation uses the change in the GDP deflator between 1991 and 1995.

⁸³ For example, one cell was a woman between the age of 18 and 45 with no comorbid condition, no recent substance abuse, treated with a Selective Serotonin Reuptake Inhibitor (SSRI) for at least 60 days plus 3 or more psychotherapy visits in a specialty mental health setting over a 16 week period.

⁸⁴ The clinical states such as mildly depressed were defined in terms of a specific scale to measure depression, which is the standard commonly used in clinical trials to define success. See (Berndt, et al., 2000a) for details.

⁸⁵ About a third of the treatments in the data base were sufficiently uncommon that the panel of psychiatrists was not asked to give probability estimates in order to lighten their task.

mix of patients, changes in the proportion of treatments that are off the frontier, and changes in the costliness of treatments.

Depending on whether a Laspeyres or Paasche index is calculated, the results show either a small decrease or a small increase in the nominal price index. By comparison, Producer Price Indices for mental health related services were increasing 2 to 6 percent per year in this period. This result for depression therefore also accords with the Cutler et al. result for heart attacks: If changes in outcomes and medical technology are properly accounted for, real medical prices have been falling, suggesting substantial bias in the official indices.

Conclusions

I have pointed to many sources of bias in the American medical care price indices. Although I am much less familiar with the price indices of other countries, several of these sources of bias would surely be important there as well. These sources include:

- The rapidity of technological change and the introduction of new goods with concomitant gains in health. Technological change has two consequences. New goods presumptively increase welfare by improving health (although because of insurance this is less obviously true in the medical marketplace than in markets for other goods and services), but the change in welfare is generally not accounted for in the official price indices. Second, because of technological change, weights quickly become out of date; in the case of heart attack treatment, annual weighting reduced the rate of real price increase by a factor of four relative to keeping the weights fixed for eleven years.
- The apparent large amount of off-frontier production of medical care and the consequent possibilities for gains in welfare from movements toward the frontier, which are ignored in the official indices. The off-frontier production may also relate in part to technological change and difficulties of medical professionals in keeping up with the literature.
- The definition of a market basket as a narrow group of goods and services, as opposed to the treatment of an episode of illness. This could in principle cause a bias in either direction, but in recent years has probably biased the American index upward. In recognition of this problem, American indices now make much greater use of episode-like pricing for hospital services than heretofore.
- Two other issues may be more specific to the United States: the difficulties in obtaining transaction prices and the resultant use of list prices, and the variation from cost shifting to or from government programs in the case of the Consumer Price Index.

Obtaining precise estimates of the bias in the official medical price indices is not possible at this time, but I have tried to summarize the sketchy evidence on the size of the various biases in Table 5. Some of these biases, such as the reduction in rents in medical prices from the spread of managed care and the reduction in off-frontier production, are once-and-for all type gains. The measures shown in Table 5, especially the 7 percent figure for the use of transaction prices, therefore overstate the steady-state bias if the official indices were to maintain their historical practices. Nonetheless, the figures in Table 5 certainly suggest that the medical care price index could have a large constant upward bias.

Because medical care carries a large and increasing weight in most countries' GDP deflator, even biases much less than those shown in Table 5 would be enough to cause a non-trivial bias in overall price indices, as I showed in the introduction using only an assumed 3 percentage point bias. Moreover, several of the biases shown in Table 5 could substantially change from year-to-year, confounding interpretation of price index changes for the purpose of monetary and fiscal policy.

Given what seems to me to be the irremediable problems of accounting for improved productivity and quality within the traditional framework of the price indices, I reach a similar conclusion to (Triplett, 2001) – that health care spending should have its own deflator that is built up from a sample of disease-specific deflators similar to those described above for heart attacks and depression. As described above, ideally the sample would be chosen randomly with probability proportional to spending on the disease, and the weight of health care in the overall index would reflect the proportion of health care spending in the relevant aggregate (e.g., GDP). The health care deflator should, at least initially, be carried alongside the official indices as an experimental or satellite index, although of course it could be used to calculate an alternative overall index (e.g., an alternative GDP deflator). Such a course holds the promise of substantial improvement over the current medical care price indices, but implementation will surely require additional resources, probably substantial additional resources, for the official statistical agencies charged with producing the improved indices.

In the case of the CPI, the new index would in principle replace the index for medical care and the current subindices, such as for hospital and related services, would be abandoned. Unlike the CPI, the PPI is industry based, which creates a problem insofar as treatment of illnesses and industries do not have one-to-one mappings. I have not determined the degree to which this is an important problem. Irrespective of whether the PPI can be treated in this fashion, the GDP deflator could be insofar as medical care has a given weight in the consumption deflator.

I have cited many examples in the text of how the American medical care price indices have been improved in recent years. Nonetheless, as Table 5 suggests, there are many opportunities for further improvement.

Table 1
Per Capita GDP Growth in the 1980s and 1990s, Japan and the United States

Years	Japan	United States
1990/1980	1.48	1.29
1998/1990	1.09	1.27

Source:(U.S. Bureau of the Census, 1999), page 842. Data shown are for constant 1990 dollars using exchange rates to convert yen to dollars. Data using purchasing power parity show a much less dramatic result; Japan outperforms the United States by 19 percent in the 1980s but only by 7 percent in the 1990s. The purchasing power parity adjustments, however, are themselves subject to potential anomalies from the treatment of medical care if they are interpreted as income needed to maintain a constant level of utility. Like the CPI, they are based on the market basket purchased from disposable income. Because of the smaller share purchased or produced by the government in the United States, the weight of medical care in the American index is substantially higher than in Japan and elsewhere. See <http://www.oecd.org/std/ppp/pps.htm> for other issues concerning the purchasing power parity index. These measures, of course, do not correct for differences between the countries in the change in real inputs.

Table 2
Percentage of GDP spent on Health Care, 1960 and 1997, Selected Countries

	1960	1997
Australia	4.9	8.4
Canada	5.5	9.0
France	4.2	9.6
Germany	4.8	10.4
Italy	3.6	7.6
Japan	3.0	7.3
Switzerland	3.3	10.1
United Kingdom	3.9	6.7
United States	5.2	13.5
OECD median	3.8	7.5

Source: (Anderson and Poullier, 1999), Table 1.

Table 3
Trends in Discharge Rates, Length of Stay, and Outpatient Surgery,
United States, 1980-1995

Year	Discharges/ 1000	Length of Stay (days)	% Surgery performed on an outpatient basis
1980	159	7.1	16
1985	138	6.3	35
1990	113	6.2	51
1995	105	5.2	60 ^a

^aValue is for 1996.

Source: (U.S. Department of Health and Human Services, 1998), pp. 303, 316.
 Length of stay is age-adjusted.

Table 4
Appropriateness of Three Procedures,
United States, Medicare Patients, 1981

Procedure	% Appropriate (95% C.I.)	% Equivocal (95% C.I.)	% Inappropriate (95% C.I.)
Coronary Angiography	74.0 (71.9-76.1)	8.5 (7.2-9.9)	17.4 (15.6-19.3)
Carotid Endarterectomy	35.3 (32.7-37.9)	32.3 (29.8-34.9)	32.4 (29.8-34.0)
Upper GI Tract Endoscopy	72.0 (69.7-74.2)	10.8 (9.3-12.4)	17.2 (15.4-19.1)

Source: (Chassin, et al., 1987).

Table 5
Sources and Potential Sizes of Bias in the American Medical Care Price Indices

Source of Bias	Size of Bias (annual)
Changes in Public Payer Reimbursement and the Resulting Cost Shift to or from Private Payers	Variable, only applies to the CPI ^a
Pricing the Hospital Day Rather than the Stay	+2.0% for heart attacks, 1983-1994
Failure to Price the Episode of Treatment in the Market Basket	+4.6% for cataract treatment, 1969-1993 ^b
Use of List Prices Rather Than Transaction Prices	+7%, California, 1983-88 ^c ; +2%, US, 1993-1997
Substitution Bias from the Use of Arithmetic Rather than Geometric Weights	+0.7%, treatment of depression, 1991-1996
Reweighting the Market Basket Every Decade Instead of Every Year	+2.1% for heart attacks, 1983-1994 ^d
Overrepresentation of Old Drugs	Up to 3% in the PPI Drug Index, 1986-1991
Linking in Generic Drugs ^e	+13.4%, +7.6% for two antibiotics, +4% for anti-depressants, 1990s
Changes in the Quality of Treatment	+4-6% for heart attacks in the elderly, 1984-1991 ^f
Reduced Off-Frontier Production	+2-6% for treatment of depression, 1991-1996 ^g

^a 20 percentage point increase in private payer hospital margins, 1984-1991, 18 percentage point decrease, 1991-1998, associated with 10 point change in Medicare hospital margins over the same period.

^b Using data on other procedures, a negative bias in earlier years, but most likely positive in recent years; see text. Outpatient and inpatient procedures put in same category after 1997.

^c HMOs in the state of Massachusetts achieved a 42 percent discount in prices paid for heart attack treatment. Even apart from the generalizability of this figure to other regions and diseases, the consequences for the annual rate of price increase depend upon the dynamics of the spread of HMOs.

^d Market basket now to be reweighted every five years.

^e Practice of not linking in new generic drugs is longer used in the United States, but was important historically.

^f Life expectancy gains only.

^g Includes quality-of-care gains.

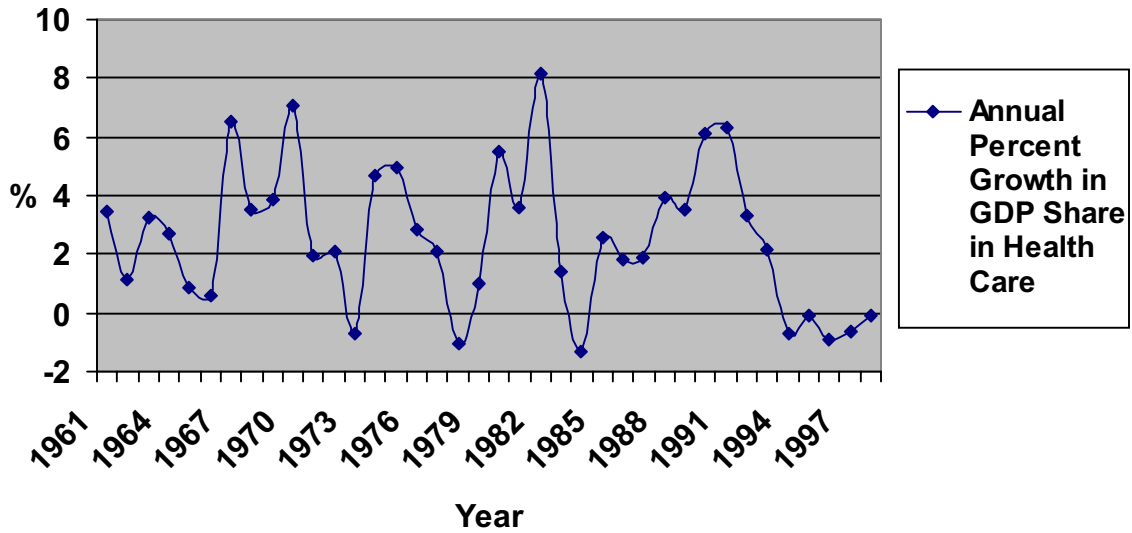


Figure 1
Annual Percentage Change in Share of American GDP in Health Care

Source: American National Health Accounts, Office of the Actuary, Health Care Financing Administration. The mean change is 2.56 percent, and the standard deviation is 2.45 percent.

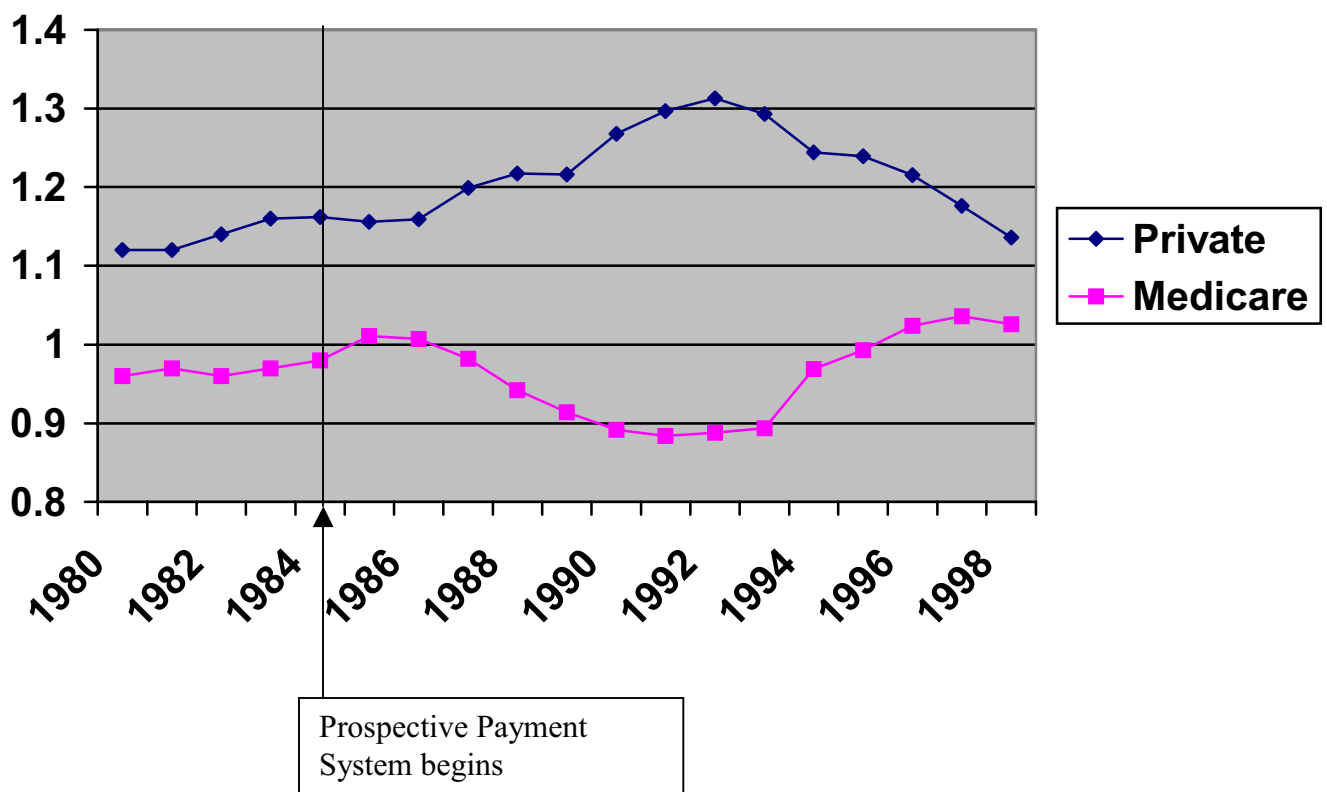


Figure 2

Ratio of Medicare and Private Payer Reimbursement to Accounting Cost, by year

Note: The R^2 between the two series for the values 1984 and after is 0.70 (for all years shown it is 0.47). (The Prospective Payment System was implemented in 1984.)

Source: Prospective Payment Assessment Commission, “Medicare and the American Health Care System: Report to the Congress,” June 1996 (page 21) and June 1997 (page 23). Medicare Payment Advisory Commission, “Report to the Congress: Selected Medicare Issues,” June 2000 (Table C-12).

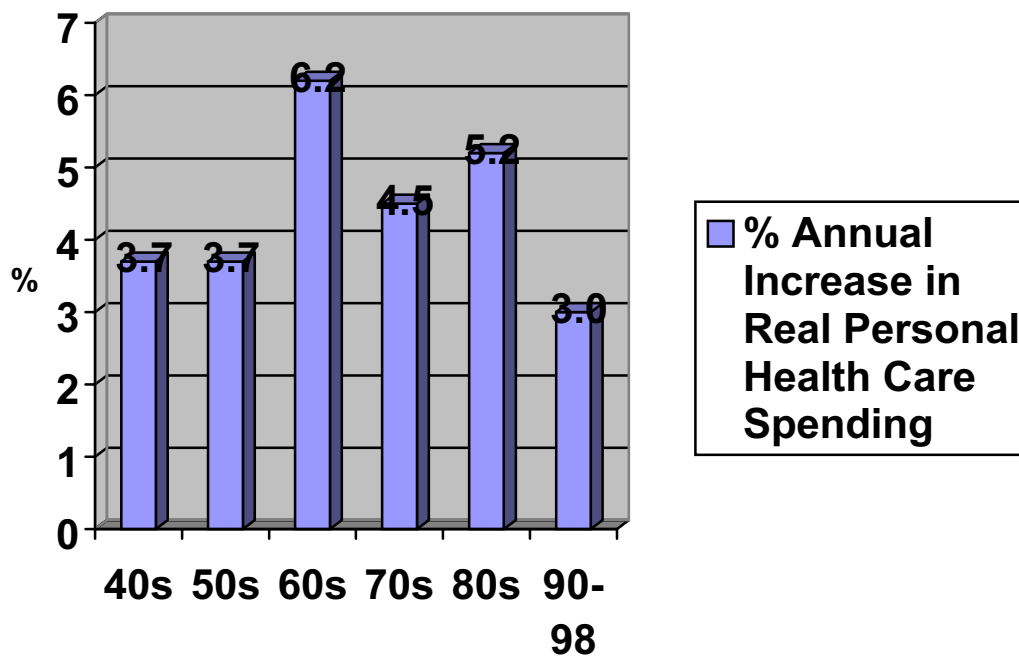


Figure 3
Growth in Real Annual per capita Increases in American Health Care Spending, by decade

Sources: (Levit, et al., 2000), (Newhouse, 1992). GDP deflator used to convert nominal dollars to real dollars.

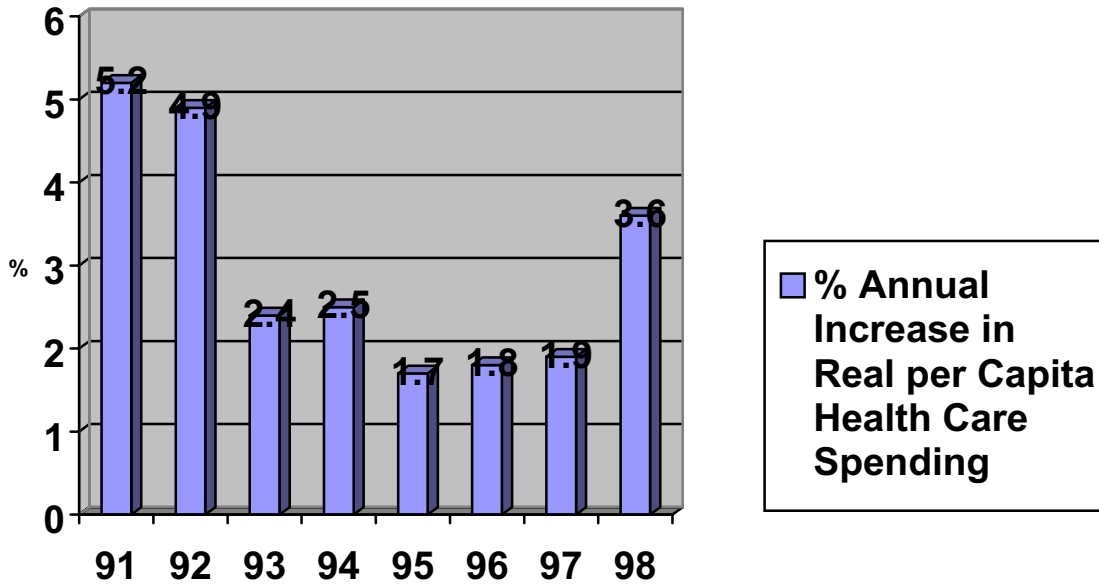


Figure 4
Real per capita Annual Rates of Spending Increase in the 1990s

Source: (Levit, et al., 2000). GDP deflator used to convert nominal to real dollars.

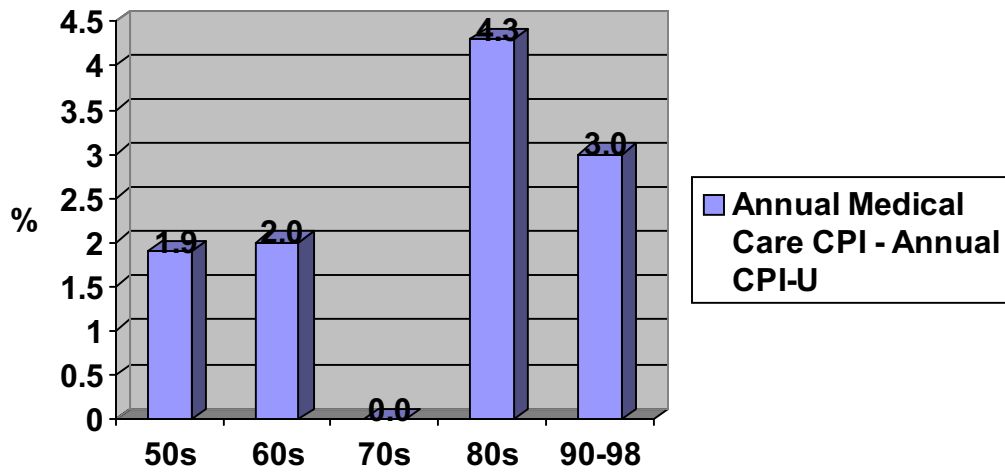


Figure 5
The Annual Excess of the American Medical Care Consumer Price Index Over the All Items Consumer Price Index in Percentage Points, by decade

Sources: (U.S. Bureau of the Census, 1998), page 489; (U.S. Council of Economic Advisers, 1999).

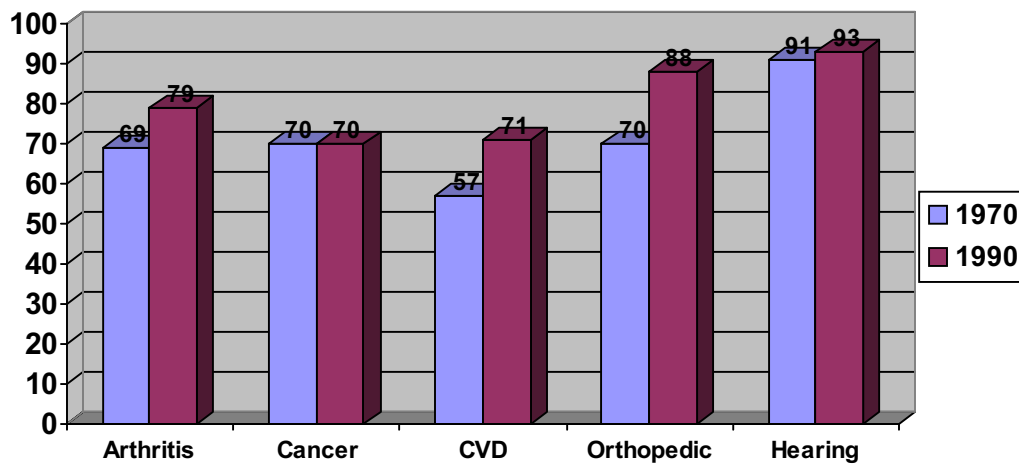


Figure 6
Quality Adjusted Life Year Weights for Various Chronic Diseases, 1970 and 1990

Note: CVD is cardiovascular disease.

Source: (Cutler and Richardson, 1997).

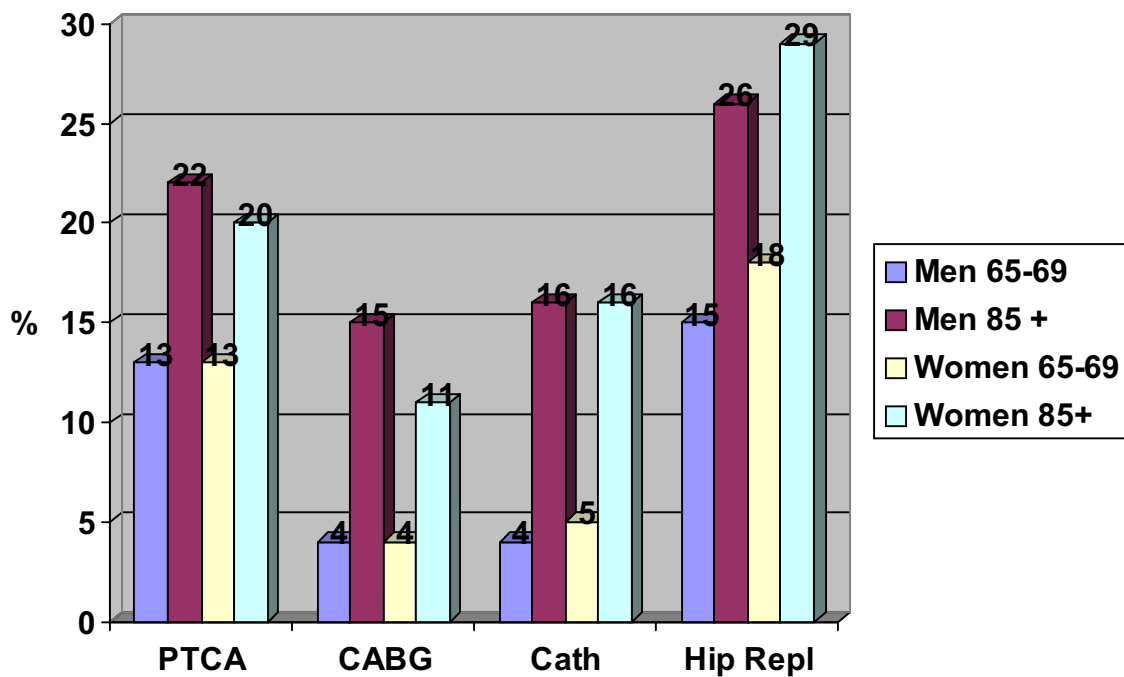


Figure 7
Annual Percentage Rates of Increase of Four Procedures in the US in Two Age Groups, 1987-1995

Note: PTCA is percutaneous transluminal coronary angioplasty. CABG is coronary artery bypass graft. Cath is cardiac catheterization. Hip Repl is hip replacement.

Source: (Fuchs, 1999).

REFERENCES

- Aaron, Henry J., and Robert Reischauer, (1995), "The Medicare Reform Debate: What Is the Next Step?," Health Affairs, **14**(4): 8-30.
- Anderson, Gerard F., and Jean-Pierre Poullier, (1999), "Health Spending, Access, and Outcomes: Trends in Industrialized Countries," Health Affairs, **18**(3): 178-192.
- Arrow, Kenneth J., (1963), "Uncertainty and the Welfare Economics of Medical Care," American Economic Review, **53**(5): 941-973.
- Athey, Susan, and Scott Stern, (2000), "The Impact of Information Technology on Emergency Health Care Outcomes"; Cambridge: National Bureau of Economic Research.
- Bates, David W., D.J. Cullen, Nan Laird, et al., (1995), "Incidence of Adverse Drug Events and Potential Adverse Drug Events," Journal of the American Medical Association, **274**(1): 29-34.
- Baumol, William J., (1967), "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis," American Economic Review, **57**(3): 415-426.
- Baumol, William J., (1993), "Health Care, Education, and the Cost Disease: A Looming Crisis for Public Choice," Public Choice, **77**(1): 17-28.
- Baumol, William J., (1997), Baumol's Cost Disease: The Arts and Other Victims; Northampton, MA: Elgar.
- Berndt, Ernst R., (1991), The Practice of Econometrics: Classic and Contemporary; Reading, MA: Addison-Wesley.
- Berndt, Ernst R., Anupa Bir, Susan H. Busch, et al., (2000a), "The Medical Treatment of Depression, 1991-1996: Productive Inefficiency, Expected Outcome Variations, and Price Indexes"; Cambridge: National Bureau of Economic Research.
- Berndt, Ernst R., David M. Cutler, Richard G. Frank, et al., (2000b), "Medical Care Prices and Output", Handbook of Health Economics, Anthony J. Culyer and Joseph P. Newhouse, ed.; Amsterdam: Elsevier, **1A**: 120-180.
- Berndt, Ernst R., David M. Cutler, Richard G. Frank, et al., (2001a), "Price Indexes for Medical Care Goods and Services: An Overview of Measurement Issues", Medical Care Output and Productivity, Ernst R. Berndt and David M. Cutler, ed.; Chicago: University of Chicago Press.
- Berndt, Ernst R., and Paul E. Greenberg, (1995), "An Updated and Extended Study of the Price Growth of Prescription Pharmaceutical Preparations", Competitive Strategies in the Pharmaceutical Industry, Robert B. Helms, ed.; Washington, DC: American Enterprise Institute: 35-48.
- Berndt, Ernst R., Iain Cockburn, and Zvi Griliches, (1996), "Pharmaceutical Innovations and Market Dynamics: Tracking Effects on Price Indexes for Antidepressant Drugs", Brookings Papers on Economic Activity: Microeconomics 1996; Washington, DC: The Brookings Institution: 133-188.
- Berndt, Ernst R., Susan M. Busch, and Richard G. Frank, (2001b), "Price Indexes for Acute Phase Treatment of Major Depression", Medical Care Output and Productivity, Ernst R. Berndt and David M. Cutler, eds.; Chicago: University of Chicago Press.
- Berndt, Ernst R., Zvi Griliches, and Joshua G. Rosett, (1993), "Auditing the Producer Price Index: Micro Evidence from Prescription Pharmaceutical Preparations," Journal of Business and Economic Statistics, **11**(3): 251-264.

- Boskin, Michael J., Ellen R. Dulberger, Zvi Griliches, et al., (1996), Toward a More Accurate Measure of the Cost of Living: Final Report from the Advisory Commission to Study the Consumer Price Index; Washington, DC: U.S. Senate Finance Committee.
- Boskin, Michael J., Ellen R. Dulberger, Zvi Griliches, et al., (1998), "Consumer Prices, the Consumer Price Index, and the Cost of Living," Journal of Economic Perspectives, **12**(1): 3-26.
- Boskin, Michael J., and Dale Jorgenson, (1997), "Implications of Overstating Inflation for Indexing Government Programs and Understanding Economic Progress," American Economic Review, **87**(2): 89-93.
- Brennan, Troyen A., Lucian L. Leape, Nan Laird, et al., (1991), "Incidence of Adverse Events and Negligence in Hospitalized Patients," New England Journal of Medicine, **324**: 370-376.
- Brook, Robert H., (1993), "Maintaining Hospital Quality: The Need for International Cooperation," Journal of the American Medical Association, **270**(8): 985-987.
- Busch, Susan H., Ernst R. Berndt, and Richard G. Frank, (2000), "Creating Price Indices for Measuring Productivity in Mental Health Care"; unpublished mimeo.
- Cardenas, Elaine M., (1996), "The CPI for Hospital Services: Concepts and Procedures," Monthly Labor Review, **119**(7): 32-42.
- Catron, Brian , and Bonnie Murphy, (1996), "Hospital Price Inflation: What Does the New PPI Tell Us?," Monthly Labor Review, **119**(7): 24-31.
- Chassin, Mark R., Robert W. Galvin, and the National Roundtable on Health Care Quality, (1998), "The Urgent Need to Improve Health Care Quality," Journal of the American Medical Association, **280**: 1000-1005.
- Chassin, Mark R., Jacquiline Kosecoff, Rolla Edward Park, et al., (1987), "Does Inappropriate Use Explain Geographic Variation in the Use of Health Care Services? A Study of Three Procedures," Journal of the American Medical Association, **258**(18): 2533-2537.
- Cowan, Cathy A., Bradley R. Braden, Patricia A. McDonnell, et al., (1996), "Business, Households, and Government: Health Spending, 1994," Health Care Financing Review, **17**(4): 157-178.
- Cowan, Cathy A., Helen C. Lazenby, Anne B. Martin, et al., (1999), "National Health Expenditures, 1998," Health Care Financing Review, **21**(2): 165-210.
- Cutler, David M., (1995), "Cutting Costs and Improving Health: Making Reform Work," Health Affairs, **14**(1): 161-172.
- Cutler, David M., and Elizabeth Richardson, (1998), "What We Get for Health-Care Spending," American Economic Review, **88**(2): 97-100.
- Cutler, David M., Jonathan Gruber, Raymond Hartman, et al., (2000a), "The Economic Impacts of the Tobacco Settlement"; Cambridge, MA: National Bureau of Economic Research.
- Cutler, David M., Mark B. McClellan, Joseph P. Newhouse, et al., (1998), "Are Medical Prices Declining?," Quarterly Journal of Economics, **113**(4): 991-1024.
- Cutler, David M., Mark B. McClellan, and Joseph P. Newhouse, (2000b), "How Does Managed Care Do It?," RAND Journal of Economics, **31**(3): 526-548.
- Cutler, David M., Mark McClellan, and Joseph P. Newhouse, (1999), "The Costs and Benefits of Intensive Treatment for Cardiovascular Disease", Measuring the

- Prices of Medical Treatments, Jack Triplett, ed.; Washington, DC: The Brookings Institution.
- Cutler, David M., Mark B. McClellan, Joseph P. Newhouse, et al., (2001), "Pricing Heart Attack Treatments", Medical Care Output and Productivity, Ernst R. Berndt and David M. Cutler, ed.; Chicago: University of Chicago Press.
- Cutler, David M., and Sarah J. Reber, (1998), "Paying for Health Insurance: The Tradeoff Between Competition and Adverse Selection," Quarterly Journal of Economics, **113**(2): 433-466.
- Cutler, David M., and Elizabeth Richardson, (1997), "Measuring the Health of the United States Population," Brookings Papers on Economic Activity, Microeconomics.
- Cutler, David M., and Richard J. Zeckhauser, (2000), "The Anatomy of Health Insurance", Handbook of Health Economics, Anthony J. Culyer and Joseph P. Newhouse, ed.; Amsterdam: Elsevier, **1A**: 563-644.
- Dranove, David, Mark Shanley, and William D. White, (1991), "Does the Consumer Price Index Overstate Hospital Price Inflation?," Medical Care, **29**(8): 690-696.
- Fisher, Franklin M., and Zvi Griliches, (1995), "Aggregate Price Indexes, New Goods, and Generics," Quarterly Journal of Economics, **110**(1): 229-244.
- Fisher, Franklin M., and Karl Shell, (1972), "The Pure Theory of the National Output Deflator", The Economic Theory of Price Indexes, Franklin M. Fisher and Karl Shell, eds.; New York: Academic Press: 49-113.
- Frank, Richard G., Ernst R. Berndt, and Susan H. Busch, (1999), "Price Indexes for the Treatment of Depression", Measuring the Prices of Medical Treatments, Jack E. Triplett, ed.; Washington, DC: Brookings Institution Press: 72-117.
- Friedman, Milton, and Simon Kuznets, (1945), Income from Independent Professional Practice; Cambridge: National Bureau of Economic Research.
- Fuchs, Victor R., (1999), "Health Care for the Elderly: How Much? Who Will Pay for It?," Health Affairs, **18**(1): 11-22.
- Gabel, Jon, Steven Di Carlo, Steven Fink, et al., (1989), "Employer-Sponsored Health Insurance in America," Health Affairs, **8**(2): 116-128.
- Goldman, Lee, and E. Fran Cook, (1984), "The Decline in Ischemic Heart Disease Mortality Rates: An Analysis of the Comparative Effects of Medical Interventions and Changes in Lifestyle," Annals of Internal Medicine, **101**: 825-836.
- Gordon, Robert J., (2000), "Does the 'New Economy' Measure Up to the Great Inventions of the Past?," Journal of Economic Perspectives, **14**(4): 49-74.
- Griliches, Zvi, (1988), Technology, Education, and Productivity; New York: Basil Blackwell.
- Griliches, Zvi, and Iain Cockburn, (1994), "Generics and New Goods in Pharmaceutical Price Indexes," American Economic Review, **84**(5): 1213-1232.
- Gruber, Jonathan, (2000), "Health Insurance and the Labor Market", Handbook of Health Economics, Anthony J. Culyer and Joseph P. Newhouse, eds.; Amsterdam: Elsevier, **1A**: 645-706.
- Hausman, Jerry A., (1999), "Cellular Telephones, New Products, and the CPI," Journal of Business and Economic Statistics, **17**(2): 188-194.
- Hunink, Maria G.M., Lee Goldman, Anna N.A. Tosteson, et al., (1997), "The Recent Decline in Mortality from Coronary Heart Disease, 1980-1990: The Effect of

- Secular Trends in Risk Factors and Treatment,” Journal of the American Medical Association, **277**(7): 535-542.
- Hurley, Jeremiah, (2000), "An Overview of the Normative Economics of the Health Sector", Handbook of Health Economics, Anthony J. Culyer and Joseph P. Newhouse, eds.; Amsterdam: Elsevier, **1A**: 55-118.
- Huskamp, Haiden A., and Joseph P. Newhouse, (1994), “Is Health Spending Slowing Down?,” Health Affairs, **13**(5): 32-38.
- Institute of Medicine, (1999), To Err Is Human; Washington, DC: National Academy Press.
- Jensen, Gail A., and Michael A. Morrisey, (1990), “Group Health Insurance: A Hedonic Price Approach,” Review of Economics and Statistics, **72**(1): 38-44.
- Jensen, Gail A., Michael A. Morrisey, Shannon Gaffney, et al., (1997), “The New Dominance of Managed Care: Insurance Trends in the 1990s,” Health Affairs, **16**(1): 125-136.
- Kahn, Katherine L., Emmett B. Keeler, Marjorie J. Sherwood, et al., (1990), “Comparing Outcomes of Care Before and After Implementation of the DRG-Based Prospective Payment System,” Journal of the American Medical Association, **264**(15): 1984-88.
- Keith, Alison, (2000), “The Economics of Viagra,” Health Affairs, **19**(2): 147-157.
- Kelly, Gregory G., (1997), “Improving the PPI Samples for Prescription Pharmaceuticals,” Monthly Labor Review, **120**(10): 10-17.
- Leape, Lucian L., (1994), “Error in Medicine,” Journal of the American Medical Association, **272**: 1851-1857.
- Leape, Lucian L., David W. Bates, David J. Cullen, et al., (1995), “Systems Analysis of Adverse Drug Events,” Journal of the American Medical Association, **274**: 35-43.
- Leape, Lucian L., Rolla Edward Park, David H. Solomon, et al., (1990), “Does Inappropriate Use Explain Small-Area Variation in the Use of Health Care Services?,” Journal of the American Medical Association, **263**(5): 669-672.
- Levit, Katharine R., Cathy Cowan, Helen Lazenby, et al., (2000), “Health Spending in 1998: Signals of Change,” Health Affairs, **19**(1): 124-132.
- Lum, S.K.S., and R.E. Yuskavage, (1997), “Gross Product by Industry, 1947-96,” Survey of Current Business, **77**(11): 220-234.
- McClellan, Mark, Barbara J. McNeil, and Joseph P. Newhouse, (1994), “Does More Intensive Treatment of Acute Myocardial Infarction Reduce Mortality,” Journal of the American Medical Association, **272**(11): 859-866.
- McClellan, Mark, and Joseph P. Newhouse, (1997), “The Marginal Costs and Benefits of Medical Technology: A Panel Instrumental Variables Approach,” Journal of Econometrics, **77**(1): 39-64.
- McDonnell, Ken, and Paul Fronstin, (1999), EBRI Health Benefits Data Book; Washington, DC: Employee Benefits Research Institute.
- McGuire, Thomas G., (2000), "Physician Agency", Handbook of Health Economics, Anthony J. Culyer and Joseph P. Newhouse, ed.; Amsterdam: Elsevier, **1A**: 461-536.
- Medicare Payment Advisory Commission, (1999), Report to the Congress: Medicare Payment Policy; Washington, DC: The Commission.

- Newhouse, Joseph P., (1992), "Medical Care Costs: How Much Welfare Loss?," Journal of Economic Perspectives, **6**(3): 3-21.
- Newhouse, Joseph P., (1996), "Reimbursing Health Plans and Health Providers: Selection versus Efficiency in Production," Journal of Economic Literature, **34**(3): 1236-1263.
- Newhouse, Joseph P., and the Insurance Experiment Group, (Rae W. Archibald, et al., (1993), Free for All? Lessons from the RAND Health Insurance Experiment; Cambridge: Harvard University Press.
- Nordhaus, William, (1998), "Quality Change in Price Indices," Journal of Economic Perspectives, **12**(1): 59-68.
- Oliner, Stephen D., and Daniel E. Sichel, (2000), "The Resurgence of Growth in the Late 1990s: Is Information Technology the Story?," Journal of Economic Perspectives, **14**(4): 3-22.
- Pauly, Mark V., (1997), Health Benefits at Work: An Economic and Political Analysis of Employment-Based Health Insurance; Ann Arbor, MI: University of Michigan Press.
- Pauly, Mark V., (1999), "Medical Care Costs, Benefits, and Effects Conceptual Issues for Measuring Price Changes", Measuring the Prices of Medical Treatments, Jack E. Triplett, ed.; Washington, DC: The Brookings Institution.
- Physician Payment Review Commission, (1997), Annual Report to Congress, 1997; Washington, DC: The Commission.
- Pilpel, D, GM Fraser, Jacquiline Kosecoff, et al., (1992), "Regional Differences in Appropriateness of Cholecystectomy in a Prepaid Health Insurance System," Public Health Review, **93**(20): 61-74.
- Reder, Melvin W., (1969), "Some Problems in the Measurement of Productivity in the Medical Care Industry", Production and Productivity in the Service Industries, Victor Fuchs, ed.; New York: Columbia University Press.
- Rice, Thomas H., (1998), The Economics of Health Reconsidered; Chicago: Health Administration Press.
- Scitovsky, Anne A., (1967), "Changes in the Costs of Treatment of Selected Illnesses, 1951-65," American Economic Review, **57**(5): 1182-1195.
- Scitovsky, Anne A., (1985), "Changes in the Costs of Treatment of Selected Illnesses, 1971-1981," Medical Care, **23**(12): 1345-1357.
- Sensenig, Arthur, and Ernest Wilcox, (2001), "National Health Accounts and National Income and Product Accounts: Reconciliation", Price Indexes for Medical Care Goods and Services: An Overview of Measurement Issues, Ernst R. Berndt and David M. Cutler, eds.; Chicago: University of Chicago Press.
- Shapiro, Matthew D., and David W. Wilcox, (1996), "Mismeasurement in the Consumer Price Index: An Evaluation", NBER Macroeconomics Annual 1996.
- Summers, Lawrence H., (1989), "Some Simple Economics of Mandated Benefits," American Economic Review, **79**(2): 177-183.
- Thomas, Eric J., David Studdert, Joseph P. Newhouse, et al., (1999), "Costs of Medical Injuries in Utah and Colorado in 1992," Inquiry, **36**: 255-264.
- Thurow, Lester C., (1984), "The Need for Industrial Policies: The Case of the U.S.A.," Annals of Public and Cooperative Economics, **55**(1): 3-31.

- Thurow, Lester C., (1987), "The Case for Industrial Policies in America", Economic Policy and Development: New Perspectives, Toshio Shishido and Ryuzo Sato, eds.; Dover, MA: Auburn House: 225-259.
- Tolley, George, Donald Kenkel, and Robert Fabian, (1994), Valuing Health for Policy: An Economic Approach; Chicago: University of Chicago Press.
- Triplett, Jack E., (2001), "What's Different about Health? Human Repair and Car Repair in the National Accounts", Medical Care Output and Productivity, Ernst R. Berndt and David M. Cutler, eds.; Chicago: University of Chicago Press.
- Triplett, Jack E., ed., (1999), Measuring the Prices of Medical Treatments; Washington, DC: Brookings Institution Press.
- U.S. Bureau of the Census, (1998), Statistical Abstract of the United States; Washington, DC: Government Printing Office.
- U.S. Bureau of the Census, (1999), Statistical Abstract of the United States; Washington, DC: Government Printing Office.
- U.S. Council of Economic Advisers, (1999), Economic Report of the President; Washington, DC: Government Printing Office.
- U.S. Department of Health and Human Services, National Center for Health Statistics, (1998), Health, United States, 1998; Washington, DC: Government Printing Office.
- U.S. Department of Health and Human Services, National Center for Health Statistics, (2000), Health, United States, 2000; Washington, DC: Government Printing Office.
- U.S. Government Accounting Office, (1996), "Consumer Price Index: Cost-of-Living Concepts and the Housing and Medical Care Components, Report to the Ranking Minority Member, Committee on Banking and Financial Services, house of Representatives"; Washington, DC.
- van de Ven, Wynand P.M.M., and Randall P. Ellis, (2000), "Risk Adjustment in Competitive Health Plan Markets", Handbook of Health Economics, Anthony J. Culyer and Joseph P. Newhouse, eds.; Amsterdam: Elsevier, **1A**: 755-846.
- Viscusi, W. Kip, (1992), Fatal Tradeoffs; New York: Oxford University Press.
- Viscusi, W. Kip, (1993), "The Value of Risks to Life and Health," Journal of Economic Literature, **31**(4): 1912-1946.
- Weiler, Paul C., Howard H. Hiatt, Joseph P. Newhouse, et al., (1993), A Measure of Malpractice: A Study of Medical Injury, Malpractice Litigation, and Patient Compensation; Cambridge: Harvard University Press.
- Wilensky, Gail R., and Joseph P. Newhouse, (1999), "Medicare: What's Right? What's Wrong? What's Next?," Health Affairs, **18**(1): 92-106.
- Zweifel, Peter, and Willard G. Manning, Jr., (2000), "Moral Hazard and Consumer Incentives in Health Care", Handbook of Health Economics, Anthony J. Culyer and Joseph P. Newhouse, ed.; Amsterdam: Elsevier, **1a**: 409-460.