

Adjusting the Measurement of the Output of the Medical Sector for Quality: A Review of the Literature

Anne E. Hall¹

Bureau of Economic Analysis

October 2015

Comments welcome.

Author e-mail: anne.hall@bea.gov

¹ The views expressed in this paper are solely those of the author and do not necessarily reflect the views of the Bureau of Economic Analysis. I would like to thank Ernie Berndt, David Cutler, Abe Dunn, and Joe Newhouse for their helpful comments. All mistakes are my own.

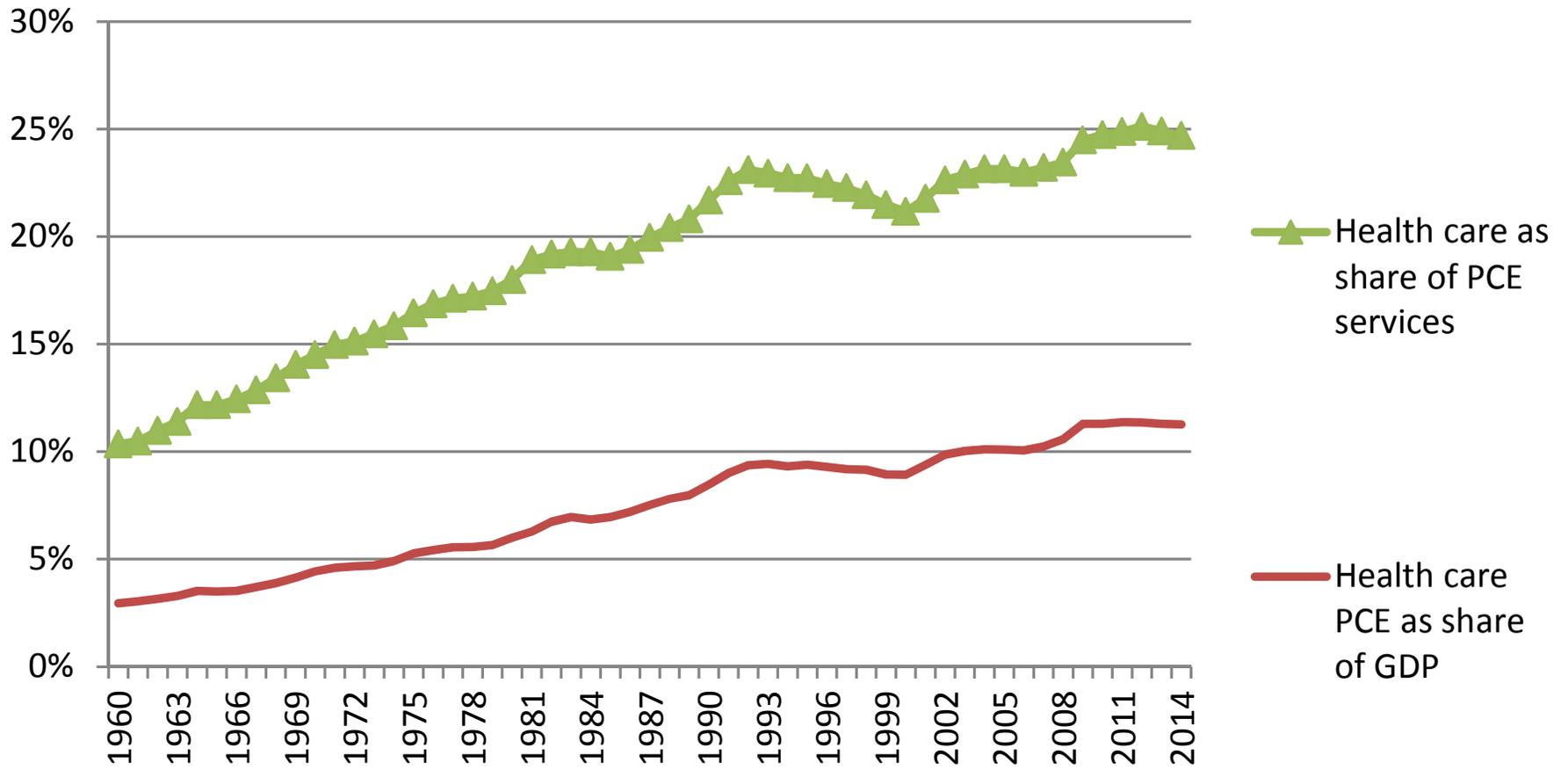
I. Introduction

The medical sector comprises a large and growing share of the economy. In the National Income and Product Accounts for 2014, the share of personal consumption expenditure (PCE) for health care of nominal GDP is currently estimated to be 11.3 percent and its share of PCE for services to be 24.7 percent. Both these shares have grown fairly steadily over the past half-century (Figure 1), although the rate of growth has declined since the early 1990s.

The medical sector also has a unique combination of features. A large fraction of health care spending is subsidized by insurance, whether public or private. According to the National Health Expenditure Accounts, in 2013, only 12.3 percent of health consumption expenditures were paid for out of pocket; of the rest, 34.9 percent was paid for by private insurance and 52.8 percent were paid for by public insurance. Rapid technological change is viewed as a major contributing factor to the rise in the medical sector's share of GDP (Newhouse 1992, Weisbrod 1991). Health care has potentially enormous benefits in terms of increasing length of and quality of life (Cutler 2004, Cutler et al. 2006, Hall and Jones 2007). Health care is also, however, subject to market failures leading to possible overuse: because a large share of spending on it is subsidized by insurance, its use is subject to moral hazard and health care has a substantial principal-agent problem as health care treatments are typically ordered by physicians who are paid on the basis of those same treatments. As much as 30 percent of health care spending in the United States has been estimated to be wasteful and not lead to improved health outcomes (Fisher et al. 2003, Skinner et al. 2005).

Because of these characteristics, a number of recommendations on how to measure economic activity in the medical sector in national accounts have been made, both internationally and specifically for the Bureau of Economic Analysis (BEA) in the US. The consensus is that the

Figure 1



Source: Bureau of Economic Analysis

unit of output of medical care should be an episode of completed treatment. This recommendation is made by the Eurostat "Handbook on price and volume measures in national accounts" (Eurostat 2001), the Atkinson report on government report in the UK (Atkinson 2005), "Accounting for Health and Health Care" (CNSTAT 2010), and the OECD, Eurostat, and the WHO in "A System of Health Accounts" (OECD/Eurostat/WHO 2011). (Current practice in the US national accounts is to treat a single service, such as a doctor's visit, as the unit of output.) The reasoning behind the recommendation is summed up by "Accounting for Health and Health Care" (AHHC):

Ideally, medical sector goods and services would be defined in such a way that: (1) expenditures could be estimated for each period for every good or service produced by the industry, (2) meaningful quantities and prices (nominal and real) could be tracked, and (3) quality change of the goods and services could be monitored. One way to proceed that embodies these three goals is to identify the output of the medical sector as completed treatments or procedures.

A treatments-based organizing framework coordinates logically with a broader health data system because, in principle, it creates a unit of analysis for which changes in the effectiveness of various medical services can be monitored. It provides a mechanism whereby prices can be adjusted to reflect changing quality, the substitution of inputs can be handled better than they are currently, and the introduction of new treatments can be dealt with on a disease-by-disease basis.

In brief, changes in the productivity of the medical sector over time are better measured on a disease-by-disease basis. Other analysts have noted that, fundamentally, patients care about how much it will cost in total to treat their illness, not about the prices of individual services (Newhouse 1989).

The BEA made the first step towards redefining the unit of output in medical care as an episode of treatment in its release of the first version of the health-care satellite account (Dunn et al. 2015). In this version, output is defined as average spending per patient per condition per calendar year, a definition which is not quite the same as completed treatments, since it includes incomplete treatments. As will be discussed below, the issue of counting incomplete treatments in output is potentially a substantial one.

The same sources cited above also agree that the measurement of medical output should be adjusted for quality of the treatment, although they do not generally make specific recommendations about what method to use for quality adjustment. AHHC notes "Improvement in medical procedures creates a major measurement issue, and any price index that does not confront it will ultimately be less than satisfactory." Similarly, the OECD report (OECD 2011) states "Price and volume measures of output should reflect quality changes in the health services provided." The Eurostat manual and the Atkinson report both also note the importance of measuring quality change when measuring output of the medical sector.

Techniques for quality adjustment of health-care output are still under development, however. In general, the research on this topic may be divided into two groups: the first adjusts output primarily based on observed health outcomes and the second adjusts output based on observed treatments or processes. Implementing outcomes-based adjustments is in some ways easier since it only requires a comparison of spending and observed health outcomes although assumptions need to be made about how to monetize the outcomes, how to measure morbidity, and how much the health care contributed to the outcomes. Implementing process-based adjustments requires first, identifying individual treatments in the data, and then quantifying the expected effect of those individual treatments on health outcomes. This latter task requires thorough knowledge of the medical literature for each treatment which makes it difficult for economists to implement. Process-based adjustments are probably more technically correct for the national accounts, however, since they are closer to how we adjust quality in other services. As Triplett (2001) notes of outcomes-based adjustments, "no national statistical agency computes in national accounts the increment that car repair makes to the stock of functioning cars, nor calculates explicitly the benefit of the repair to the car owner."

II. Measuring output

In this section, I will discuss in greater detail the recommendation that the output of the medical sector be measured in "completed treatments." Measuring output correctly is revealing in itself and is a necessary first step when implementing a process-based quality adjustment.

As outlined in Dunn et al. (2015), the first version of the Health Care Satellite Account (HCSA) calculates the disease-based price indexes or medical care expenditure (MCE) indexes, as the average expenditure per patient c for condition d in time t divided by the average expenditure per patient c for condition d in the base period 0:

$$MCE_{d,t} = \left(\frac{c_{d,t}}{c_{d,0}} \right)$$

The HCSA weights all health-care spending equally. However, the OECD/Eurostat/WHO report defines a complete treatment as "the pathway that an individual takes through different health providers in order to receive full and final treatment for a disease or condition." The Eurostat handbook and the Atkinson report define them similarly.²

The difference between that definition and the definition used in the HCSA leads to potentially substantive differences in measurement. First, the definition used in the HCSA does not allow for multiple episodes in a year; if an individual has two strokes during a calendar year, their total expenditure for the year will be counted as one observation contributing to the average expenditure for strokes in that year. Conversely, a single episode of treatment that lasts only a

² In general, the international sources focus on the difficulties of measuring "complete treatments" across providers because health-care utilization data in European countries often do not allow for following individual patients. The UK National Health Service, for example, apparently has no way to track patients administratively across different providers (Dawson et al. 2005). This issue is less of a problem in the US where the main data sources used in the HCSA are the Medical Expenditure Panel Survey and medical claims data, both of which are patient-centered data sources and allow for measurement of complete treatments relatively easily.

few months but whose duration crosses over New Year's will be counted as two episodes in the HCSA and the spending for the single episode will be split across those two episodes.

Dunn et al. (2014) compared an encounter-based method that averaged spending per patient per condition per year (similar to what was ultimately used in the HCSA) to a commercial grouper that classified medical claims into episodes of treatment. They found no significant difference in overall medical inflation between the methods, although they did find differences in the sub-indexes created for major practice categories. They also found that restricting the analysis to episodes within a calendar year made no difference to inflation, suggesting that this restriction only makes trivial differences in practice in the HCSA.

Another potential problem with the approach taken by the HCSA, however, is whether every health-care dollar represents "full and final treatment for a disease or condition." Rosen et al. (2012), while evaluating two commercial groupers, note that one of them leaves 14 percent of spending ungrouped (not part of any treatment episode) and the other leaves 18 percent of spending ungrouped. If we adopt a strict definition where a complete treatment is one consistent with guidelines, then even more spending is not part of a complete treatment and therefore not leading to output. Berndt et al. (2001), for example, find that a full 50 percent of episodes of major depression involve treatment that is not consistent with guidelines for the treatment of that condition. Colla et al. (2015) find relatively high prevalence of low-value treatments inconsistent with provider's society guidelines among Medicare patients; 22 percent of low-risk back pain patients receive imaging against guidelines, for example, and 47 percent of surgery patients receive preoperative cardiac testing that is inconsistent with guidelines. More generally, it has been estimated that as much as 20 to 30 percent of US health-care spending is wasteful and does not lead to any improvement in health outcomes (Fisher et al. 2003, Skinner et al. 2005).

Refining the measure of output by establishing which health-care spending is part of an effective treatment and which is wasteful may therefore in itself substantially change measured medical inflation before the addition of overt quality adjustment.

III. Quality adjustment of output

The first version of the HCSA did not account for any changes in the quality of health care such as improved health outcomes, fewer side effects, or less risky and invasive procedures. The unique characteristics of the health care system described in the introduction make measuring quality change in health care both more difficult and more important than quality change in other sectors. Measuring quality change in health care is not straightforward since so much of health care is paid for through insurance rather than on the margin so standard hedonic analysis techniques (such as are used for housing, the largest category of PCE for services in the US national accounts) are not applicable.³ However, as will be discussed in more detail below, some research on specific conditions such as heart disease, depression and cancer show that adjusting for quality can make a significant difference to measured inflation in medical care (Cutler et al. 1998, Berndt et al. 2006, Lucarelli and Nicholson 2009).

The various authorities that recommend adjusting for quality in health care in the national accounts do not come to a consensus on how exactly to undertake the task. "Accounting for Health and Health Care" (AHHC) envisions two separate accounts. The first, which should be

³ Hedonic analysis and related demand analysis techniques have been used quite successfully, however, to analyze the markets for goods that are inputs into the medical services sector. Multiple papers have conducted analyses of different markets for pharmaceuticals: Berndt et al. (1995) and Suslow (1996) for anti-ulcer drugs, Cockburn and Anis (2001) for rheumatoid arthritis drugs, Goldman et al. (2010) for cancer drugs, and Dunn (2012) for anti-cholesterol drugs. Trajtenberg (1990) conducted a hedonic analysis of the market for computed tomography scanners.

part of the National Income and Product Accounts (NIPAs), measures the output of the medical sector. This output is an input to the other account, which measures the stock of population health in quality-adjusted life-years (QALYs), and which also has non-medical inputs such as diet, exercise, and environment. AHHC emphasizes that the output of the medical care account, properly specified, is the medical treatment, not the associated health outcome, since the health outcome is the output of the health account. AHHC allows that an index based on treatments may be adjusted for quality based on outcomes but does not go into detail about how exactly to do that quality adjustment.

Similarly, the Eurostat handbook (Eurostat 2001), notes that the focus should be on treatments as outputs, not on health outcomes but information on specific aspects of outcomes might serve as proxies for changes in quality of output. The Atkinson report (Atkinson 2005) states that measuring quality change is important but does not make any recommendations on how to do it.

The OECD/Eurostat/WHO report "A System of Health Accounts" (OECD/Eurostat/WHO 2011) gives a set of guidelines on how to adjust for quality in health care:

- The quality measure should be aligned with the processes sought by consumers, which would generally be a complete treatment of the disease.
- The adjustment in output should reflect the marginal contribution of the health industry to an outcome. It should not be affected by any other factors that influence health outcomes such as genetic background, income or lifestyle.
- Consumers are ultimately concerned to achieve an improvement in their health outcome. Waiting times and comfort are secondary to improvements in health status. This points to the conclusion that different dimensions of quality should not be given the same weight.
- In many health treatments or processes, there is a time lag before the improvements in health status. Quality adjustment needs to address in a realistic manner the impact of lifetime effects of health expenditures.
- The quality measure should reflect as closely as possible the normal, average or expected effect of the activity on the state of health. Individual capacities to benefit from treatment, or what is known as co-production, should not be counted in the measure of quality-adjusted health volume output.
- International comparison is important, and the indicators and methods of output adjustment should be standardised across countries to facilitate comparisons.

The report concedes that techniques for quality adjustment are still under development and therefore it does not make an explicit recommendation for how to do it. It lists some possibilities, however, that are consistent with the guidelines set out above. One is to require an adjustment factor that reflects the rate of compliance with published treatment guidelines; it notes, however, that while treatment guidelines are readily available, data on rates of compliance are not. Treatment quality has multiple dimensions but the OECD report points to several ways in which the characteristics of treatment quality can be collapsed into one dimension: choose an endpoint (such as 30-day survival) to be measured, use multiple indicators and weight them equally, use multiple indicators and weight them based on expert opinion, or measure the effect of the treatments on QALYs. Among the papers discussed below, there will be examples of each of these techniques.

As discussed in the introduction, methods for adjusting quality of health-care output can be divided into those primarily based on health outcomes and those based on health-care processes.

"A System of Health Accounts" (OECD/Eurostat/WHO 2011) offers support for both approaches:

Obviously there is a strong connection between process and output, as treatment guidelines are based on medical evidence about what is efficacious. Health services researchers recommend using both process and outcome indicators for two reasons. First, there is a difference between evidence in research (efficacy) and outcomes in real life (effectiveness). Second, there is frequently a considerable time lag between a process and its impact on the outcome.

As mentioned above, it is slightly easier to implement outcomes-based adjustments, as reflected by the fact that there are more papers using outcomes-based adjustments than process-based adjustments.

III.1 Outcomes-based quality adjustments. Table 1 lists the papers (or sets of papers) that calculate a quality-adjusted price index or otherwise quantify the value of health-care outcomes

Table 1: Papers using primarily outcome-based quality adjustments

Paper	Condition(s)	Data	Outcome	Assumption about contribution of medical care to changes in outcomes
Cutler, McClellan, Newhouse and Remler (1998, 2001)	Acute myocardial infarction (AMI)	Clinical data from a major teaching hospital and Medicare claims	Life expectancy post-AMI (2001 also considers value of quality of life)	100%
Cutler, Rosen and Vijan (2006)	All conditions	Household surveys for spending; CDC life tables for outcomes	Life expectancy	50%
Rosen, Cutler, Norton, Hu, and Vijan (2007)	Coronary heart disease (CHD) and acute myocardial infarction (AMI)	MCBS for spending; Vital Statistics for CHD outcomes and Medicare claims for AMI outcomes	CHD: life expectancy. AMI: Difference between life expectancy for AMI patients and non-AMI patients	100% for CHD; 100% for AMI after netting out improvement for non-AMI patients
Eggleston, Shah, Smith, Berndt, and Newhouse (2009, 2011)	Diabetes	Spending and clinical data from Mayo Clinic's self-funded health plan	Modifiable cardiovascular risk (MCR)	100% (for MCR)
Highfill and Bernstein (2014)	30 chronic conditions	Household surveys for spending; Global Burden of Disease study for outcomes	Disability-adjusted life-years (DALYs) as found by GBD's survey of general population	100%
Romley, Goldman, and Sood (2015)	Acute myocardial infarction, heart failure, and pneumonia	Medicare hospital claims	30-day, 14-day, and one-year survival after hospitalization	100%

using an outcomes-based adjustment. Papers that use an outcomes-based adjustment confront two major issues. First, they must establish whether they are including only mortality as an outcome or whether they include morbidity or disability as well, and if they include the latter, they must quantify it in terms of QALYs or DALYs. Second, it is well-known that medical care is only one determinant of health. Papers that use actual health outcomes as a measure of the efficacy of the medical sector must therefore make an assumption about what proportion of those outcomes is due to medical care. Most of the papers below simplify these issues by considering one condition or a limited set of conditions.

Cutler, McClellan, Newhouse, and Remler (1998, 2001) calculate a cost-of-living index (COLI) for a single condition, acute myocardial infarction (AMI, or heart attack) from the mid-1980s through the mid-1990s using data for Medicare beneficiaries. A cost-of-living index for a single medical condition can be calculated as:

$$COLI_t = \frac{Y_0 - (\text{Net value of medical spending}_t)}{Y_0} = \frac{Y_0 - (\Delta \text{ in benefits} - \Delta \text{ in spending})}{Y_0}$$

Y_0 is income in the base period. The change in benefits is the monetized change in health outcomes. The numerator of the COLI is the amount of money it takes in period t to have the same utility as in the base period 0 , given period t 's technology. If the change in benefits is greater than the change in spending, the COLI will fall because it will be cheaper in period t to achieve the same level of utility as in period 0 .

Cutler et al. define the changes in benefits of medical care as the change in life expectancy post-AMI. Life expectancy post-AMI rose by about a year between the mid-1980s and the mid-1990s. They convert this increase to monetary value using common

assumptions about the value of a life-year and find that the value ranges from \$8600 to \$86,000, depending on the assumption. They define spending on the heart attacks as all medical spending for the heart attack patient up to 90 days after the heart attack and find that these costs increased from \$11,500 to \$18,000. The net value of heart attack care is therefore positive and they find the COLI drops by 1.5 percent per year.

Cutler et al.'s work therefore illustrates the potential importance of adjusting for the benefit of health outcomes. They also calculate an episode-based fixed-basket index that is unadjusted for quality; it rises 2.3 percent annually. Adjusting for quality therefore greatly reduces medical inflation in the case of heart attacks during this period.

Their focus on a single acute medical event that has high mortality simplifies their analysis. They are able to assume that all medical spending within 90 days of the event is related to the event and that all mortality improvements after the event are related to medical treatment. This approach limits the applicability of their method to acute conditions such as heart attack or stroke.

While I have classified their paper as one using an outcomes-based adjustment since the final analysis only considers health outcomes, they have an extensive discussion of the treatments patients had, how these changed over the period, and the contribution of each treatment to the increases in life expectancy post-AMI and in costs. This discussion is necessary to ground their assumption that the improvement in outcomes is entirely due to medical care. Their accounting of treatments, however, uses privately obtained clinical data with details not available in claims data; this feature also limits the applicability of their work to other conditions in other populations.

Cutler, Rosen and Vijan (2006) measure the net value of all medical spending in the US on all conditions between 1960 and 2000 using household survey data on spending and life tables for outcomes. While they do not explicitly calculate a price index or a cost-of-living index, any analysis calculating a net value can be converted easily to a COLI using the formula above. They only consider life expectancy when measuring the benefits of medical care; they do not include morbidity or disability levels. They assume that medical care is responsible for 50 percent of the increase in life expectancy between 1960 and 2000. They base this assumption on the finding that 90 percent of the increase in life expectancy is due to reductions in two causes of death: cardiovascular disease and death in infancy, and on previous research showing that about half of the reduction in deaths from cardiovascular disease and nearly all of the reductions in infant mortality during this period are due to medical advances. They find that medical spending had a positive net value from 1960 to 2000, with variations among decades and among age groups. In particular, they find that the cost per life year gained has been much higher among the elderly in the last two decades of the period.

Rosen, Cutler, Norton, Hu, and Vijan (2007) calculate the net value of spending on coronary heart disease from 1987 to 2002. They take a similar approach to Cutler, Rosen and Vijan (2006) but focus on one related set of conditions. They also only consider life expectancy as an outcome. They assume that medical treatment for coronary heart disease is entirely responsible for the improvements in outcomes but, when calculating the improvement in life expectancy for AMI patients, they net out the improvement in life expectancy for non-AMI patients. They find from this calculation that nearly 90 percent of the improvement in life expectancy for AMI patient is due to improvement in

AMI care. When calculating the change in spending, they calculate the change in expected lifetime medical spending, not just the change in spending on the episode of treatment. Like Cutler et al. (1998, 2001), they document changes in treatments and their contributions to the improvements in outcomes based on the medical literature although their final calculation only depends on spending. They find that spending on coronary heart disease and on AMI generally had positive net value over the period.

Eggleston, Shah, Smith, Berndt, and Newhouse (2009, 2011) estimate the net value of spending on diabetes care between 1999 and 2009 using clinical and spending data from the Mayo Clinic's self-funded health plan. They measure outcomes with clinically modifiable risk, the part of cardiovascular risk that can be controlled by medical care and has been identified as such by previous medical research. The change in benefits is the change in life expectancy resulting from the reduction in the risk of a life-ending cardiovascular event; they do not consider other benefits of improved diabetes treatment (such as avoiding a foot amputation). They count all of the medical spending of diabetes patients when measuring the change in spending. They find that the net value of diabetes treatment over this period is positive, although the return on spending varies by diagnosis cohort. Earlier cohorts who were diagnosed later in the progression of the illness have a much higher return on spending on treatment than later cohorts who were diagnosed earlier.

Highfill and Bernstein (2014) measure the net value of spending of 30 chronic conditions from 1987 to 2010. They use spending data from household surveys (the 1987 National Medical Expenditure Survey and the 2010 Medical Expenditure Panel Survey) and outcomes data from the Global Burden of Disease study (US Burden of Disease

Collaborators 2013). Outcomes for each condition are measured in disability-adjusted life-years (DALYs) by the Global Burden of Disease study. Disability levels for each condition were not measured directly but were assigned by a survey of the general population using paired comparison questions about which person in different, random states of health are healthier (Salomon et al. 2012). Highfill and Bernstein, following other sources, assign a value of \$100,000 to a year of life in perfect health. Using that assumption, they find that only a few conditions have positive net value and some, such as Alzheimer's disease, have large negative net value over the period 1990 to 2010. The 30 conditions, aggregated together, had slightly positive net value. In their analysis, they assume that all improvements in outcomes are due to medical care, although they concede in the discussion that that assumption may be more appropriate for some conditions than others. Their paper is nearly unique, however, in studying so many conditions and in valuing outcomes other than mortality; their results indicate that the return on spending on conditions other than heart disease may not be very high and that the results found by Cutler and his co-authors may not be representative of the medical system as a whole.

Romley, Goldman and Sood (2015) estimate the productivity of hospitals in treating Medicare beneficiaries for three specific conditions with high mortality: AMI, heart failure, and pneumonia, over the period 2002-2011. The quality adjustment is in how they measure the outcome: as survival after 30 days without an unplanned readmission (they also use 14-day and 1-year survival rates). They do not value the outcome in monetized life-years but simply measure output as surviving to this endpoint and compare it to Medicare spending on the admission. They find that the productivity of hospitals

measured in this way grows, on net, over the period. The productivity of treating pneumonia increased more than productivity of treating the other two conditions. Their study is limited to hospital output and spending.

There are two main advantages of adjusting the measurement of medical output for quality based on observed outcomes. First, life expectancy is relatively easily measured and relatively easy to put a dollar value on. Mortality-based adjustments are easier to implement than process-based adjustments as reflected by the fact that there are more papers using mortality-based adjustments than process-based adjustments. Second, observed outcomes reflect the actual productivity of the medical sector.

There are several downsides to using outcomes-based adjustments, however. First, it is not easy to measure disability or quality of life, nor is it easy to measure them in dollar terms. The Global Burden of Disease data suggest measuring disability could be important as it finds that total years of life lost due to disability were nearly as high in 2010 as years of life lost due to mortality. Some conditions with low mortality but high prevalence, such as back pain and major depression, have higher years lost due to disability than all conditions other than ischemic heart disease do to mortality. It is also not easy to know how much of the change in observed life expectancy is due to medical care; three of the six papers discussed deal with this issue by only considering conditions with an acute onset and high mortality, so it is possible to attribute all improvement in outcomes to medical care.

Table 2 presents an analysis that summarizes both the advantages and disadvantages of a mortality-based adjustment to the output of medical care. It calculates the net value of medical spending from 2000 to 2012, both for all medical spending and separately by ICD chapter. The

Table 2
Net value of medical spending by ICD chapter 2000-2012

	Spending						Mortality (Deaths per 100,000 population)			Value of life extended by medical care	
	2000		2012		Increase		2000	2012	Decrease		Net value
	Total (\$ billions)	Per capita (\$)	Total (\$ billions)	Per capita (\$)	Total (\$ billions)	Per capita (\$)					
Diseases of the circulatory system	\$153	\$554	\$243	\$775	\$90	\$221	334.6	250.8	83.8	\$3,625	\$3,405
Neoplasms	\$62	\$219	\$125	\$397	\$63	\$178	201.3	190.4	10.9	\$472	\$293
Certain conditions originating in the perinatal period	\$5	\$16	\$8	\$24	\$3	\$9	5	3.8	1.2	\$52	\$43
Congenital anomalies	\$5	\$18	\$7	\$23	\$2	\$5	3.8	3.1	0.7	\$30	\$25
Diseases of the respiratory system	\$93	\$327	\$158	\$503	\$65	\$176	82.1	78.5	3.6	\$156	-\$21
Complications of pregnancy; childbirth; and the puerperium	\$26	\$76	\$39	\$124	\$13	\$48	0.1	0.3	-0.2	-\$9	-\$56
Diseases of the skin and subcutaneous organs	\$21	\$77	\$45	\$142	\$23	\$66	1.3	1.3	0	\$0	-\$66
Diseases of the blood and blood-forming organs	\$9	\$31	\$37	\$119	\$29	\$88	3.3	3.3	0	\$0	-\$88
Diseases of the digestive system	\$56	\$201	\$108	\$344	\$52	\$143	29.9	30.4	-0.5	-\$22	-\$165
Infectious and parasitic diseases	\$23	\$83	\$68	\$215	\$44	\$133	21	21.8	-0.8	-\$35	-\$167
Diseases of the genitourinary system	\$65	\$193	\$114	\$362	\$49	\$170	19.4	19.8	-0.4	-\$17	-\$187
Endocrine; nutritional; and metabolic diseases and immunity disorders	\$53	\$192	\$139	\$444	\$87	\$252	33.5	33.7	-0.2	-\$9	-\$261
Diseases of the musculoskeletal system and connective tissue	\$77	\$275	\$188	\$598	\$111	\$323	4.9	4.4	0.5	\$22	-\$301
Injury and poisoning	\$65	\$231	\$119	\$379	\$53	\$148	53.8	61.5	-7.7	-\$333	-\$481
Symptoms; signs; and ill-defined conditions	\$73	\$295	\$250	\$795	\$177	\$500	11.3	13.4	-2.1	-\$91	-\$591
Diseases of the nervous system and sense organs	\$60	\$214	\$134	\$428	\$74	\$214	32.3	46.6	-14.3	-\$619	-\$833
Mental illness	\$43	\$140	\$80	\$256	\$37	\$116	16.4	47.1	-30.7	-\$1,328	-\$1,444
Residual codes; unclassified; all E codes	\$14	\$50	\$42	\$134	\$28	\$84	n/a	n/a	n/a		n/a
Total	\$901	\$3,192	\$1,903	\$6,064	\$1,003	\$2,872	854.0	810.2	43.8	\$1,895	-\$977

Notes:

1. Spending data are from the BEA Health Care Satellite "blended" account (<http://www.bea.gov/national/xls/HCSA/Blended.xlsx>).
2. Mortality data are from the CDC Compressed Mortality File (<http://wonder.cdc.gov/cmfi-icd10.htm>).
3. 50% of the improvements in mortality are attributed to medical care.
4. The value of a life saved is assumed to be \$8.7 million. This is the value used by the Department of Transportation (http://www.transportation.gov/sites/dot.gov/files/docs/VSL_Guidance_2014.pdf), deflated to 2009 levels.

spending data are per capita and come from the Health Care Satellite "blended" Account and are in nominal terms.⁴ The mortality data are from the CDC Compressed Mortality File. When calculating the monetized value of mortality improvements, I assume that 50 percent of the improvements are due to medical care and that the value of a life saved is approximately \$8.7 million. This value is based on the value used by the Department of Transportation which is \$9.1 million in 2012 dollars; I have deflated it to 2009 levels.

The results show that the net value of the increase in total medical spending from 2000 to 2012 was approximately -\$977. The breakup by ICD chapter, however, shows great differences among condition groups. "Diseases of the circulatory system" are an outlier at the top end, with a positive net value of \$3,405 which is more than ten times as large as that of the next ICD chapter "Neoplasms." The high net value of the treatment of circulatory disorders reflects the improvements in mortality from these conditions between 2000 and 2012; the death rate from these conditions fell by over 80 deaths per 100,000 population during this period, a decline of about a quarter. Excluding circulatory disorders, the net value of medical spending during this period was -\$4,298.

Only three other chapters have positive net value. "Neoplasms" has a net value of \$293, reflecting significant improvements in mortality. "Certain conditions originating in the perinatal period" and "Congenital anomalies" both had small improvements in mortality accompanied by tiny increases in per capita spending, leading to a slight positive net value for both.

The remaining thirteen chapters, accounting for over 80 percent of spending and about 45 percent of mortality in 2012, each all have negative net value. They have significant increases in

⁴ As outlined in Dunn et al. (2015), the "blended" account combines medical claims data for populations for whom it is available (Medicare and the privately insured) with data from the Medical Expenditure Panel Survey for the remaining populations.

spending accompanied by small declines or even increases in mortality. Two chapters, "Diseases of the nervous system" and "Mental illness" have increases in mortality of more than 10 per 100,000 population over the decade. As they also each have increases in per capita spending of more than \$100, they have the worst net values of -\$833 and -\$1,444 respectively.

The analysis in Table 2 has the advantage that it is relatively straightforward and feasible with publicly available data. As revealed by it, however, measurement of the benefits of medical care with mortality outcomes makes spending on circulatory system disorders appear hugely productive and spending on other conditions appear much less so. This difference leads to two not entirely mutually exclusive conclusions. First, it suggests that much of the research cited above is highly unrepresentative of medical spending and productivity as a whole. Four of the six papers using outcomes-based adjustments focus on the value of medical spending on cardiovascular-related conditions. These papers generally found the return on medical spending on these conditions to be quite high, in line with the results for 2000 to 2012 as shown in Table 2, although they generally covered earlier periods. However, Table 2 finds the net value of total medical spending during the period 2000-2012 to be well below zero.

The second conclusion suggested, however, is that, due to the limitation of having only mortality as an outcome, the net value of other conditions is understated. For ICD chapters such as "Diseases of the musculoskeletal system and connective tissue", whose conditions induce a substantial amount of disability but not very much mortality, the increase in medical spending may very well accompany an unmeasured reduction in disability due to these conditions. The Agency for Healthcare Research and Quality has documented that, between 1997 and 2010, the per population rate of knee replacements nearly doubled and the per population rate of hip replacements rose by 38 percent (AHRQ 2013). This increase in utilization is included as part of

the nominal increase of \$323 per capita for this category in the HCSA. However, since joint replacements are intended to alleviate disability rather than reduce mortality, that return on spending is not captured in Table 2.

Both of these conclusions may simultaneously be true; the return on spending on conditions other than circulatory conditions may be higher than indicated by the mortality statistics but circulatory conditions may still be an outlier with respect to its return relative to other conditions. There is no question that the mortality improvements for circulatory conditions far outpace those of other ICD chapters. However, taking a full outcomes-based approach to adjusting the output of the medical sector would require more comprehensive data on disability levels for each condition and how those disability levels translate into quality-adjusted life-years than seem to be currently available.

III.2 Process-based quality adjustment. Adjusting the measurement of medical output for quality based on observed treatments rather than outcomes has more data requirements than adjusting for outcomes. Adjusting for outcomes only requires data on spending and on outcomes while adjusting for treatments requires detailed data on which treatments were administered as well as medical knowledge of how effective the treatments are. Table 3 lists the papers that use process-based quality adjustments. Three of them are grouped together as they were by approximately the same coauthors and studied mental health conditions. Berndt et al. 2006 aggregated the indexes in those papers into a single price index for mental health spending. The other two papers studied cancer, with one focusing on a single cancer (colorectal) and the other studying all cancers.

Table 3: Papers using process-based quality adjustments

Paper	Condition	Data	What is measured	Processes and how they are evaluated
Research on mental health conditions summarized and aggregated in Berndt et al. 2006				
Berndt, Bir, Busch, Frank, and Normand (2002)	Major depression	Claims data (Medstat)	Price per remission	Patients placed into treatment baskets and probabilities of remissions for each patient-treatment combination elicited from expert panel.
Frank, Berndt, Busch, and Lehman (2004)	Schizophrenia	Claims data (Medicaid)	Price per episode adjusted hedonically by quality measures	Constructed a set of quality measures in the claims data from treatment recommendations published by the Agency for Health Care Policy and Research and the National Institute of Mental Health
Ling, Busch, and Frank (2004)	Bipolar I disorder	Claims data (Medstat)	Price per episode adjusted hedonically by quality measures	Constructed a set of quality measures in the claims data from treatment guidelines published by the American Psychiatric Association
Lucarelli and Nicholson (2009)	Colorectal cancer	IMS Health (for drug prices); IntrinsicQ (for market shares of chemotherapy regimens 2002-2005); Medicare claims in SEER (1993-2001)	Price per QALY added by chemotherapy treatment	Chemotherapy regimens are evaluated from medical literature.
Howard, Bach, Berndt and Conti (2015)	All cancers (drug treatments only)	CenterWatch; FDA; Memorial Sloan-Kettering Cancer Center Center for Health Policy & Outcomes	Price per life-year added by treatment (with controls for side effects)	Effectiveness of drugs collected from FDA approvals

Berndt, Bir, Busch, Frank, and Normand (2002) create a quality-adjusted price index for the treatment of major depression for the period 1991 to 1996. They use medical claims data for the privately insured for both treatments and spending. They place patients into treatment buckets such as "1 psychotherapy visit", "SSRI [sustained serotonin release inhibitor] >30 days and 1-3 psychotherapy visits", or "TCA [tri-cyclic antidepressant] >30 days and no psychotherapy", to list a few. They then presented the treatment baskets to a panel of clinical experts and elicited from them probabilities that the particular treatment as applied to particular patients with particular comorbidities would result in a remission of the depression. This procedure allowed them to calculate a probability of remission for each treatment basket. When they combined these probabilities with average spending on each treatment, they were able to produce a price per expected remission for each treatment. Aggregating over treatments, they calculated a price index where the price is spending per remission of major depression and found that this price fell about 2 to 3 percent per year depending on specification.

Frank, Berndt, Busch, and Lehman (2004) create a quality-adjusted treatment-based price index for schizophrenia for the period 1994 to 2000. They use medical claims data for Medicaid patients in two Florida counties for both treatments and spending. They also place patients into treatment buckets. They applied quality measures that had been previously published by the Agency for Healthcare Policy and Research (AHCPR) and the National Institute for Mental Health (NIMH) to the claims data by coding each episode for whether or not they met the quality measures. They found that, because of dramatic changes in the compositions of treatments during this period, treatment-based price indexes fell 5.5 percent annually.

Ling, Busch, and Frank (2004) calculate a quality-adjusted price index for bipolar I disorder for the period 1991 to 1995. They use medical claims data for the privately insured for both treatments and spending. They adjust for quality with a method similar to that of Frank et al. (2004), by coding episodes for whether or not they met certain quality measures and creating a hedonic index, but in this paper they used measures developed by the American Psychiatric Association. Their treatment-based index falls 3.3 percent annually.

Lucarelli and Nicholson (2009) calculate several price indexes for the chemotherapy treatment of colon cancer (they do not include other treatments such as surgery) for the period 1998 to 2005. They combine data on prices from IMS Health and on market shares from IntrinsicQ. I will focus on the index that adjusts for quality of the treatments by measuring the price of incremental expected QALY of each treatment. The expected QALYs of treatments are taken from the medical literature; several studies had already calculated utility weights for the possible health states of colon cancer patients. They note that many of the drugs have an incremental cost per incremental QALY of \$100,000 to \$150,000 (relative to the reference treatment), close to the monetized value of life according to other sources. They also calculate an unadjusted index based on the average price of treatments. Their comparison shows the potential impact of adjusting for quality; their unadjusted index rises by 2600 percent over the period but their quality-adjusted index declines slightly.

Howard, Bach, Berndt, and Conti (2015) calculate a price index for cancer drugs where the price is treatment episode cost per life-year added by the treatment for the period 1995 to 2013. They do not use QALYs but control for side effects of the drugs in a

hedonic regression. They find that the index rises 10 percent annually and this result is robust to changes in specification.

Compared with outcomes-based quality adjustments, process-based quality adjustments have the advantage that they do not require an assumption about the contribution of medical care; they are based on the medical literature on how the treatments contribute to health outcomes. They also give more information on the treatment of which conditions are improving in productivity; an approach like Cutler, Rosen and Vijan (2006) suggests that medical spending has had a positive return on the whole but does not give any information on which medical spending has been most productive. As mentioned above, several of the papers on individual conditions with outcomes-based adjustments (Cutler et al. 1998 and 2001, Rosen et al. 2007, and Eggleston et al. 2009 and 2011) also examine the trends in treatments to justify their assumptions on how medical care contributed to the observed health outcomes.

The major disadvantage of process-based quality adjustments is that they are more difficult to implement. They require knowledge of the medical effectiveness of every treatment and detailed data on what treatments were provided. It should be noted that the author teams of Frank et al. (2004), Ling et al. (2004), and Howard et al. (2015) all contained at least one MD. Cutler et al. (1998, 2001) and Eggleston et al. (2009, 2011), when examining trends in treatments, used clinical indicators that are not normally available in claims data which are the primary source of data BEA has been using for the HCSA. Berndt et al. (2002) used claims data but measured quality with an expert panel; Dawson et al. (2005), when evaluating quality adjustment methods for the UK National Health Service, noted of expert panels "Such groups are costly to convene, organize, and train." In addition, as payers move to bundled payments, treatment details may no longer be available in claims. Cutler et al. (1998, 2001) used clinical data from a teaching

hospital to track what treatments were being used because such detail is not available in Medicare claims since Medicare pays hospitals on a DRG basis.

On the up side, models of how treatments affect the outcomes of individual diseases are becoming more available. Eggleston et al. (2009, 2011) and Lucarelli and Nicholson (2009) employed such models taken from the medical literature. Applying process-based quality adjustment to measurement of the output of the medical sector in the national accounts, however, will require modeling every treatment for every condition separately and will require intensive interdisciplinary work.

IV. Conclusion

This paper has reviewed the recommendations from domestic and international sources for measuring the output of the medical sector in national accounts and for adjusting the measurement for changes in health-care quality. It then discussed how refining the measurement of output in the Health Care Satellite Account to completed treatments might contribute to improving the measurement of quality in the medical sector. It then went on to summarize and discuss research papers that have implemented quality adjustment in price indexes for health care or otherwise compared the benefits of health care with its costs. It divided these papers into two groups: one group based its quality adjustment mostly on observed health outcomes (while taking into account treatments used) while the other based the adjustment on health-care processes.

The advantages and disadvantages of each approach are summarized in Table 4.

Table 4: Summary of methods for quality adjustment of medical output		
Type of adjustment	Advantages	Disadvantages
Outcomes-based	<ul style="list-style-type: none"> - Easier to implement; fewer data demands. - Data on mortality outcomes is available. 	<ul style="list-style-type: none"> - Require assumptions about the contribution of medical care to health outcomes, which can require considering processes anyway. - Difficult to measure or value quality of life or morbidity. - Difficult to separate mortality outcomes by condition.
Process-based	<ul style="list-style-type: none"> - Closer to how other services are measured in the national accounts. 	<ul style="list-style-type: none"> - Requires more medical knowledge. - More stringent data requirements.

As Table 4 shows, in general, outcomes-based adjustments are perhaps easier to implement but require some strong assumptions, and how to measure quality of life, which is likely to be important, is unclear. Process-based adjustments, however, have higher data and knowledge requirements. The results of the papers, however, show that quality adjustment of health care output is quite important. After quality adjustment, the outcomes-based price index for heart attacks created by Cutler et al. (1998, 2001) and the process-based price index for colon cancer created by Lucarelli and Nicholson (2009) showed declines instead of increases. In the summary of the CNSTAT workshop "Strategies for a BEA Satellite Health Care Account", it is stated "Among participants, there seemed to be complete agreement that quality adjustment of price indexes for the satellite health care accounts is extremely important, and also that it is very hard to do" (CNSTAT 2009). The discussion here reflects that position.

References

- Agency for Healthcare Research and Quality (2013). "Healthcare Cost and Utilization Project Statistical Brief #149: Most Frequent Procedures Performed in U.S. Hospitals, 2010" (<http://www.hcup-us.ahrq.gov/reports/statbriefs/sb149.pdf>).
- Atkinson, T. (2005). *Atkinson Review: Final report. Measurement of Government Output and Productivity for the National Accounts*. Hampshire, England: Palgrave-Macmillan.
- Berndt, E. R., Bir, A., Busch, S. H., Frank, R. G., and Normand, S. T. (2002). The Medical Treatment of Depression, 1991-1996: Productivity Inefficiency, Expected Outcome Variations, and Price Indexes. *Journal of Health Economics* 21(3): 373-96.
- Berndt, E. R., Bui, L., Reiley, D. R., and Urban, G. L. (1995). Information, Marketing, and Pricing in the U.S. Antiulcer Drug Market." *American Economic Review* 85(2): 100-105.
- Berndt, E. R., Busch, A. B. and Frank, R. G. (2006). Real Output in Mental Health Care during the 1990s. In D. M. Cutler and A. M. Garber (eds.), *Frontiers in Health Policy Research, Volume 9*. Berkeley Electronic Press.
- Berndt, E. R., Busch, S. H. and Frank, R. G. (2001). Treatment Price Indexes for Acute Phase Major Depression. In D. M. Cutler and E. R. Berndt (eds.), *Medical Care Output and Productivity* (National Bureau of Economic Research Studies in Income and Wealth Series, vol. 62, pp. 463-505). Chicago, IL: University of Chicago Press.
- Cockburn, I. M. and Anis, A. H. (2001). Hedonic Analysis of Arthritis Drugs. In D. M. Cutler and E. R. Berndt (eds.), *Medical Care Output and Productivity* (National Bureau of Economic Research Studies in Income and Wealth Series, vol. 62, pp. 439-458). Chicago, IL: University of Chicago Press.
- Colla, C. H., Morden, N. E., Sequist, T. D., Schpero, W. L., and Rosenthal, M. B. (2015). Choosing Wisely: Prevalence and Correlates of Low-Value Health Care Services in the United States. *Journal of General Internal Medicine* 30(2): 221-8.
- Committee on National Statistics (CNSTAT) (2009). *Strategies for a BEA Satellite Health Care Account: Summary of a Workshop*. Washington, DC: National Academies Press.
- Committee on National Statistics (CNSTAT) (2010). *Accounting for Health and Health Care: Approaches to Measuring the Sources and Costs of Their Improvement*. Washington, DC: National Academies Press.
- Cutler, D. M. (2004). *Your Money or Your Life: Strong Medicine for America's Health Care System*. Cary, NC: Oxford University Press.
- Cutler, D. M., McClellan, M., Newhouse, J. P., and Remler, D. (1998). Are medical prices declining? Evidence for heart attack treatments. *Quarterly Journal of Economics* 113(4): 991-1024.

Cutler, D. M., McClellan, M., Newhouse, J. P., and Remler, D. (2001). Pricing Heart Attack Treatments. In D. M. Cutler and E. R. Berndt (eds.), *Medical Care Output and Productivity* (National Bureau of Economic Research Studies in Income and Wealth Series, vol. 62, pp. 305-362). Chicago, IL: University of Chicago Press.

Cutler, D. M., Rosen, A. B., and Vijan, S. (2006). The Value of Medical Spending in the United States, 1960-2000. *The New England Journal of Medicine* 355(9): 920-27.

Dawson, D., Gravelle, H., O'Mahony, M., Street, A., Weale, M., Castelli, A., Jacobs, R., Kind, P., Loveridge, P., Martin, S., Stevens, P., and Stokes, L. (2005). Developing new approaches to measuring NHS outputs and productivity. CHE Research Paper 6 (http://www.york.ac.uk/media/che/documents/papers/researchpapers/rp6_Measuring_NHS_outputs_and_productivity.pdf).

Dunn, A. (2012). Drug Innovations and Welfare Measures Computed from Market Demand: The Case of Anti-cholesterol Drugs. *American Economic Journal: Applied Economics* 4(3): 167-89.

Dunn, A., Liebman, E., Rittmueller, L., and Shapiro, A. (2014). Defining Disease Episodes and the Effects on the Components of Expenditure Growth. BEA working paper 2014-4 (<http://www.bea.gov/papers/pdf/definingdisease.pdf>).

Dunn, A., Rittmueller, L., and Whitmire, B. (2015). Introducing the New BEA Health Care Satellite Account. *Survey of Current Business* 95(1) (http://www.bea.gov/scb/pdf/2015/01%20January/0115_bea_health_care_satellite_account.pdf).

Eggleston, K. N., Shah, N. D., Smith, S. A., Wagie, A. E., Williams, A. R., Grossman, J. H., Berndt, E. R., Long, K. H., Banerjee, R., and Newhouse, J. P. (2009). The Net Value of Health Care for Patients with Type 2 Diabetes, 1997 to 2005. *Annals of Internal Medicine* 151(6): 386-93.

Eggleston, K. N., Shah, N. D., Smith, S. A., Berndt, E. R., and Newhouse, J. P. (2011). Quality Adjustment for Health Care Spending on Chronic Disease: Evidence from Diabetes Treatment, 1999-2009. *American Economic Review: Papers & Proceedings*, 101(3): 206-11.

Eurostat (2001). *Handbook on Price and Volume Measures in National Accounts*. Luxembourg, Germany: European Communities.

Fisher, E. S., Wennberg, D. E., Stukel, T. A., Gottlieb, D. A., Lucas, F. L., and Pinder, E. L. (2003). The implications of regional variations in Medicare spending: Health outcomes and satisfaction with care. *Annals of Internal Medicine* 138(18): 288-98.

Frank, R. G., Berndt, E. R., Busch, A. B. and Lehman, A. F. (2004). Quality-constant "prices" for the ongoing treatment of schizophrenia: an exploratory study. *The Quarterly Review of Economics and Finance* 44: 390-409.

Goldman, D. P., Jena, A., Lakdawalla, D. N., Malin, J. L., Malkin, J. D. and Sun, E. (2010). The Value of Specialty Oncology Drugs. *Health Services Research* 45(1): 115-32.

Hall, R. E. and Jones, C. I. (2007). The Value of Life and the Rise in Health Spending. *The Quarterly Journal of Economics* 122(1): 39-72.

- Highfill, T. and Bernstein, E. (2014). Using Disability-Adjusted Life Years to Value the Treatment of Thirty Chronic Conditions in the U.S. from 1987-2010. BEA working paper 2014-9 (http://www.bea.gov/papers/pdf/Highfill_Bernstein%202015%20DALYS%20update.pdf).
- Howard, D. H. Bach, P. B., Berndt, E. R., and Conti, R. M. (2015). Pricing in the Market for Anticancer Drugs. *Journal of Economic Perspectives* 29(1): 139-62.
- Ling, D. Y. C., Busch, A. B., and Frank, R. G. (2004). Treatment price indexes for bipolar disorder. Working paper, Harvard Medical School.
- Lucarelli, C. and Nicholson, S. (2009). A Quality-Adjusted Price Index for Colorectal Cancer Drugs. NBER working paper 15174 (<http://www.nber.org/papers/w15174>).
- Newhouse, J. P. (1989). Measuring Medical Prices and Understanding Their Effects. *Journal of Health Administration Education* 7(1): 19-26.
- Newhouse, J. P. (1992). Medical Care Costs: How Much Welfare Lost? *Journal of Economic Perspectives* 6(3): 3-21.
- OECD, Eurostat, WHO (2011). *A System of Health Accounts*. OECD Publishing.
- Romley, J. A., Goldman, D. P. and Sood, N. (2015). US Hospitals Experienced Substantial Productivity Growth During 2002-11. *Health Affairs* 34(3) 511-58.
- Rosen, A. B., Cutler, D. M., Norton, D. M., Hu, H. M., and Vijan, S. (2007). The Value of Coronary Heart Disease Care For The Elderly: 1987-2002. *Health Affairs* 26(1): 111-23.
- Rosen, A. B., Liebman, E., Aizcorbe, A., and Cutler, D. M. (2012). Comparing Commercial Systems for Characterizing Episodes of Care. BEA working paper 2012-7 (http://www.bea.gov/papers/pdf/meg_vs_etg_paper.pdf).
- Salomon, J. A., Vos, T. et al. (2012). Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *The Lancet* 380: 2129-43.
- Skinner, J., Fisher, E., and Wennberg, J. E. (2005). The Efficiency of Medicare. In D. Wise (ed.) *Analyses of the Economics of Aging*. Chicago: University of Chicago Press and NBER.
- Suslow, V. Y. (1996). Measuring Quality Change in Pharmaceutical Markets: Hedonic Price Indexes for Anti-Ulcer Drugs. In R. B. Helms (ed.), *Competitive Strategies in the Pharmaceutical Industry*. Washington, DC: American Enterprise Institute.
- Trajtenberg, M. (1990). *Economic Analysis of Product Innovation*. Cambridge, MA: Harvard University Press.
- Triplett, J. E. (2001). What's Different about Health? Human Repair and Car Repair in National Accounts and National Health Accounts. In D. M. Cutler and E. R. Berndt (eds.), *Medical Care Output and Productivity* (National Bureau of Economic Research Studies in Income and Wealth Series, vol. 62, pp. 15-94). Chicago, IL: University of Chicago Press.

US Burden of Disease Collaborative (2013). The State of US Health, 1990-2010: Burden of Disease, Injuries, and Risk Factors. *Journal of the American Medical Association* 310(6): 591-606.

Weisbrod, B. A. (1991). The Health Care Quadrilemma: An Essay on Technological Change, Insurance, Quality of Care, and Cost Containment. *Journal of Economic Literature* 29(2): 523-52.