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

Anne Hall (Bureau of Economic Analysis, United States)

Paper prepared for the 34th IARIW General Conference

Dresden, Germany, August 21-27, 2016

Session 2F: Meeting the Measurement Challenges of Official Statistics Offices I

Time: Monday, August 22, 2016 [Afternoon]

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

Anne E. Hall[1]
Bureau of Economic Analysis
Author e-mail: anne.hall@bea.gov

June 2016

[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, Anna Malinovskaya, Joe Newhouse, Louise Sheiner, Jon Skinner and two anonymous referees for their helpful comments. All mistakes are my own.

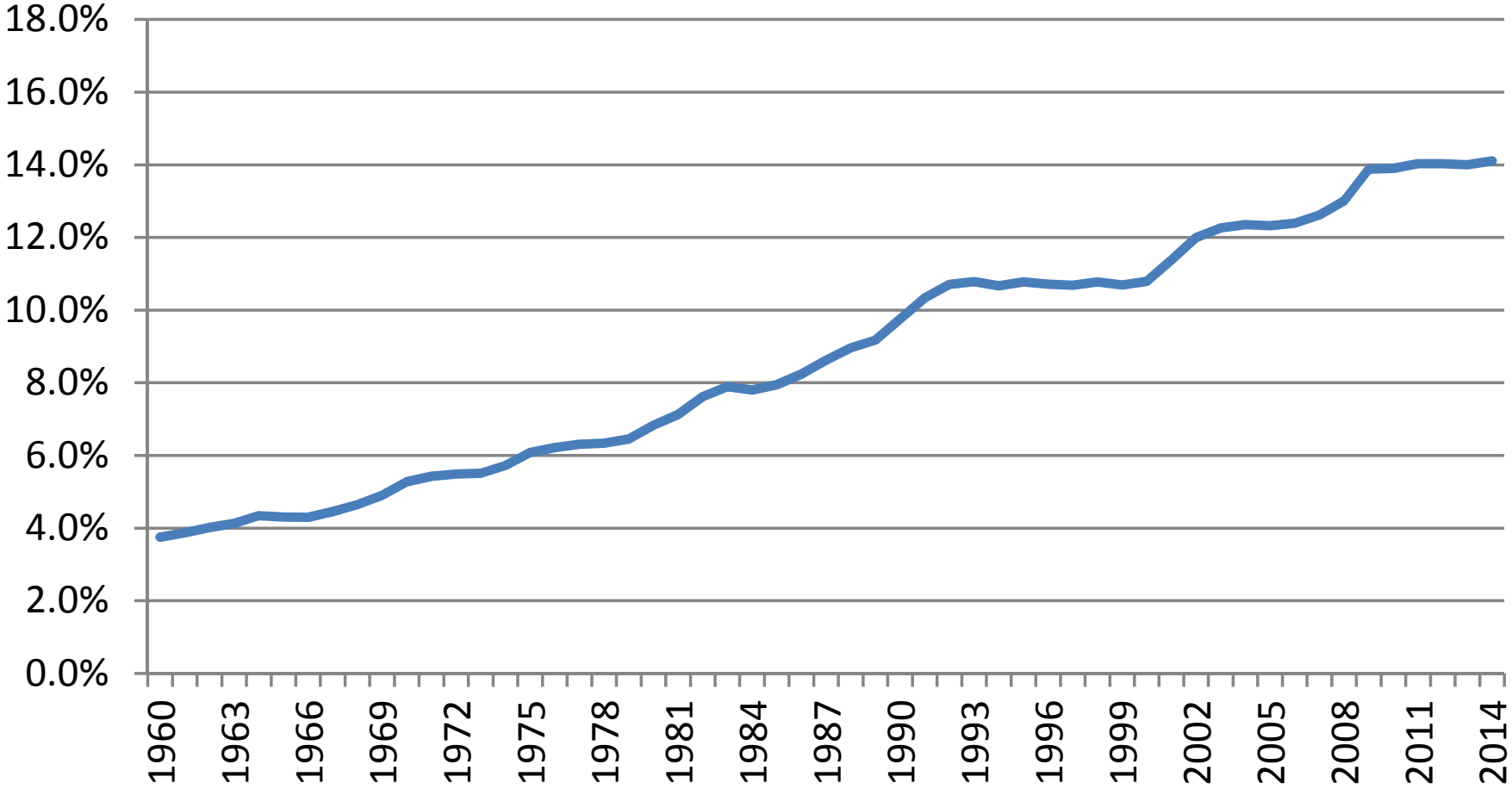
I. Introduction

Accurate economic measurement is important for the conduct of economic policy. The Federal Reserve relies extensively on the inflation, output and other measures constructed by statistical agencies to inform its decisions on monetary policy while the Congress and the executive branch make use of those measures to design and implement fiscal policy. Statistical agencies in the United States such as the Census Bureau, the Bureau of Labor Statistics (BLS), and the Bureau of Economic Analysis (BEA), have well-established frameworks and methods for collecting data and for measuring inflation, output, employment, productivity, and other measures of economic activity. The BEA, in particular, when constructing measures of gross domestic product, generally follows internationally established national accounting standards outlined in the System of National Accounts (Landefeld et al. 2008).

These agencies continually improve and update their measures within these frameworks as they receive feedback from users and experts. One sector that has come under attention at agencies in recent years due to its size, rapid rate of growth, and measurement challenges is the medical sector. In the National Income and Product Accounts (NIPAs) for 2014, the share of the health component of personal consumption expenditure (PCE) of nominal GDP is currently estimated to be 14.1 percent.[2] This share has grown fairly steadily over the past half-century (Figure 1), although the rate of growth has declined since the early 1990s.

[2] This figure is lower than the number commonly cited as the share of health care spent on GDP, namely the total of the National Health Expenditure Accounts divided by total GDP. In 2014, this share was 17.5 percent (Martin et al. 2016). The figures differ because personal consumption expenditures for health care only include household spending on health care goods, health care services, and health insurance premiums net of benefits; government payments to households by public insurance programs such as Medicare and Medicaid; and final consumption expenditures of nonprofit health services serving households. They do not include other health-care expenditures such as capital investment by health-care businesses and government public health activities. (In addition, there are

Figure 1: Health component of personal consumption expenditures as a share of gross domestic product



Source: Bureau of Economic Analysis National Income and Product Accounts

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 2014, only 11.5 percent of health consumption expenditures were paid for out of pocket; of the rest, 34.4 percent was paid for by private insurance and 38.7 percent were paid for by public insurance. Rapid technological change has been cited as a major contributing factor to the historical rise in the medical sector's share of GDP (Newhouse 1992). Dunn et al. (2015)'s finding that increases in costs from 2000 to 2010 are being primarily driven by cost per case suggests that technological change is still an important factor for growth in health-care spending. 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 its size and because a good portion of the medical sector is subsidized by government, there exists great demand for accurate measurement of inflation, output and productivity of this sector. The current practice in the Consumer Price and Producer Price Indexes and in the NIPAs is to treat a single service, such as a doctor's visit, as the unit of output. However, the consensus among health economists and national accounting authorities is that the

---

other data and methodological differences between the NHEA and the NIPAs; for a review see Kornfeld et al. 2010). The Health Care Satellite Account and the literature reviewed in this paper focus solely on the question of how to deflate consumer spending on health care.

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 output in the UK (Atkinson 2005), the Committee on National Statistics of the National Academies in the US (National Research Council 2010), and the OECD, Eurostat, and the WHO in "A System of Health Accounts" (OECD 2011). The reasoning behind the recommendation is summed up by the 2010 National Research Council report "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. (National Research Council 2010)

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 (HCSA) (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 (Eurostat 2001, Atkinson 2005).

II. New contributions

As discussed in the introduction, the question of how to adjust for quality when measuring inflation and output in the medical sector is of great importance to anyone interested in the accurate measurement of the economy. Techniques for quality adjustment of health-care output are still under development, however. This paper is the first paper to collect, summarize, and evaluate the diverse literature that has been published in this area. It divides the research on this topic into two groups: the first adjusts output primarily based on observed health outcomes and the second adjusts output based on observed treatments or processes. As I discuss, implementing outcomes-based adjustments is in some ways easier since it only requires a comparison of spending and observed health outcomes. However, assumptions may need to be made about how to monetize the outcomes, how to measure morbidity and/or disability, 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 alone.

III. Inflation measurement

In the United States, the main statistical agency charged with measuring inflation, the Bureau of Labor Statistics, constructs their price indexes based on the concept of a cost-of-living index (Bureau of Labor Statistics 2015). A cost-of-living index (COLI) measures how much more or less income do consumers need to be just as well off with a new set of prices as with the old. The Consumer Price Index as actually constructed by the BLS follows a Laspeyres formula, which measures the change between the base period and period *t* in the cost of a basket of goods chosen in the base period where quantities of the individual goods are held constant:

$$I_L = \frac{\sum_i p_{it} q_{i0}}{\sum_i p_{i0} q_{i0}}$$

As noted by the BLS and others (Bureau of Labor Statistics 2015, National Research Council 2002), when it is constructed for normal market goods, the CPI overstates a true COLI because it does not allow for substitutions of goods in response to price changes. Typically as prices of some goods rise relative to those of others, consumers will substitute towards the relatively cheaper goods. The CPI holds quantities constant, and utility will therefore by definition be lower from its market basket than from a basket that allows substitution; it is therefore higher than a true COLI.

An alternative price index formulation, the Paasche index measures the expenditure on a basket of goods chosen in period *t* relative to the expenditure on the same basket in the base period. It is a floor on the COLI, as the Laspeyres index is a ceiling to it.

$$I_P = \frac{\sum_i p_{it} q_{it}}{\sum_i p_{i0} q_{it}}$$

The true COLI therefore lies somewhere between the Laspeyres index and the Paasche index. A class of indexes known as superlative price indexes offer approximations that are likely closer to the true COLI than either the Laspeyres or the Paasche index. The Bureau of Economic Analysis, when constructing deflators to measure real output of the economy, uses such an index, the Fisher price index, which is the geometric mean of the Laspeyres index and the Paasche index:

$$I_F = \sqrt{I_L I_P}$$

In order for the indexes to be accurate, the goods whose prices are being measured must be identical in terms of characteristics important to the consumer. If the characteristics change over time, the indexes must adjust for change in quality received by the consumer. The BLS and the BEA both use a variety of methods to adjust for quality when measuring inflation in goods whose characteristics change rapidly over time such as housing and high-technology goods. These methods include hedonic regressions, matched-model indexes, and attribute-cost adjustment (Bureau of Labor Statistics 2015, Bureau of Economic Analysis 2015).

As many have noted, however, these methods do not apply well when measuring the quality of medical care (Cutler et al. 1998, Berndt et al. 2001). Hedonic regressions, for example, rely on consumers' full marginal valuation of quality changes being reflected in the price of the good. A substantial portion of health care, however, is subsidized by insurance so this assumption does

not hold. In 2014, for example, only 11.5 percent of health consumption expenditures were paid for out of pocket (Martin et al. 2016).

Another issue the constructors of quality-adjusted price indexes must confront is how to introduce innovations into price indexes. These innovations can range from a new breakfast cereal which is entering a market already populated by other cereals and other types of breakfast food to major inventions such as electricity or airplane travel, which allow the consumption of something that was never consumed before. There is an extensive literature on the "new goods" problem in price measurement of non-health-care goods and services but as Bresnahan and Gordon (1996) note, empirical methods in the economic literature for measuring the value of new goods all rely on demand-based assessments of willingness to pay. These methods are extensions of the methods mentioned above used to measure quality changes in the same products over time so they cannot be applied to health care for the same reason, that much of health care is subsidized by insurance and is not paid for on the margin by consumers. The issue of measuring the value of innovative new treatment is particularly pressing for health care, however, a sector in which technological change has historically been rapid.

Cutler et al. (1998) lay out a theory of how to apply the COLI framework to health care while taking into account the incremental health benefits of improved medical technology. They write the COLI as:

$$COLI = 1 + \frac{C}{Y_0}$$

Where Y_0 is base period income and C is the amount required to make consumers just as well off in the current period as in the base period. The base period income provides utility given medical

spending and technology in the base period and C provides the additional amount to give utility equivalent to the base period given medical spending and medical technology in the current period.

Cutler et al. (1998) show that C can be split into two components, as it can be written as the difference between the monetized health benefit of improved medical care and the change in medical spending from the base period to the current period. The COLI is therefore closely related to the cost-benefit analysis of medical care, the field of health services research which compares the costs of treatments to their medical benefits.

This approach of measuring the health benefits of new medical technology, translating them into dollar terms, and comparing them to the change in costs is used by several papers discussed below. Another approach, however, is exemplified by Berndt et al. (2002). They identify a specific endpoint of treatment, the remission of major depression. They then measure how likely it is patients will achieve this endpoint, given what treatments they receive and the medical literature on the effectiveness of those treatments. In this way, the percent of cases that achieve the endpoint (such as remission) can be put into a price index formula (whether Laspeyres, Paasche or Fisher) as q_i and the cost per case can be put into the formula as p_i. A third approach, taken by Frank et al. (2004), measure the cost of treatment baskets over time, holding quality constant. It should be noted that neither Berndt et al. or Frank et al. offer no theoretical justifications for their approaches. However, as discussed in "At What Price", the CPI is not a pure COLI in practice so consideration of nontheoretical approaches is appropriate (National Research Council 2002). As will be discussed below, all of these approaches have been taken in the literature and have their advantages and disadvantages.

IV. 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:

$$MCE_{d,t} = (\frac{C_{d,t}}{C_{d,0}})$$

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" (OECD 2011). The Eurostat handbook and the Atkinson report define them similarly (Eurostat 2001, Atkinson 2005).[3]

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

[3] 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.

expenditure for strokes in that year. Conversely, a single episode of treatment that lasts only a 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.

V. 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. As discussed above, measuring quality change in health care is not straightforward since so much of health care is not paid for on the margin, limiting the use of hedonic analysis techniques.[4] 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. 2002, 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 (National Research Council

[4] 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.

2010). The first, which should be 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 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. (OECD 2011)

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" 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. (OECD 2011)

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.

While my discussion so far has focused on the issue of quality adjustment for price indexes, not all of the papers below construct full-fledged quality-adjusted price or expenditure indexes. Only four papers do so: Cutler et al. 1998, Berndt et al. 2002, Frank et al. 2004, and Howard et al.

2015. The papers, however, were chosen because they share a distinguishing characteristic, that they in some way relate changes in health-care spending and changes in treatment quality or health outcomes over a time period, rather than simply evaluating the cost-effectiveness of health-care spending at a point in time.

V.1 Outcomes-based quality adjustments. Table 1 lists the papers (or sets of papers) that calculate a quality-adjusted price index with an outcomes-based adjustment or otherwise quantify the value of a change in health-care spending based on a change in observed health outcomes. Papers that use an outcomes-based adjustment confront two major issues. First, they must establish what outcome it is appropriate to measure for the particular set of conditions they are considering and what units the outcomes should be measured in. 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 should therefore establish what proportion of those outcomes is due to medical care.

Cutler, McClellan, Newhouse, and Remler (1998) 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. As discussed above, in section III., a cost-of-living index for a single medical condition can be calculated as:

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

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

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)	Acute myocardial infarction (AMI)	Clinical data from a major teaching hospital and Medicare claims	Life expectancy post-AMI	100% after netting out improvements for non-AMI population
Cutler and Meara (2000)	Low-birthweight infants	Estimates from medical literature for spending; vital statistics for outcomes	Quality-adjusted life expectancy	100% conditional on birthweight
Cutler and McClellan (2001)	Breast cancer	SEER database linked to Medicare claims	Life expectancy	100% in case-based method; population-based method attributes rise in incidence to increased screening
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 population
Duggan and Evans (2008)	HIV/AIDS	Medical claims data linked to mortality files from California's Medicaid program	Mortality	100% (conditional on controls for demographics and HIV severity)
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%

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*.

As mentioned above, one of the key issues these papers must confront is establishing how much of the improvements in outcomes is due to medical care. Cutler et al. define the changes in benefits of medical care as the change in life expectancy post-AMI out to 5 years after the 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 a year 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. measure the benefits of medical care for AMI by taking the difference between the improvement in life expectancy for AMI patients and the improvement in life expectancy of the non-AMI population, adjusted for demographic differences. However, as they acknowledge, this method for isolating the effect of medical care on improvements in life expectancy post-AMI is potentially problematic. Insofar as other trends in behavior (such as reductions in smoking) or improvements in medical care for conditions that are correlated with AMI (such as diabetes, stroke or lung cancer) impact the AMI and the non-AMI population differentially, simply differencing out the improvement in life expectancy for the non-AMI population will not accurately capture the effect of medical care on improvements in post-AMI life expectancy.

Another potential difficulty with their analysis is that they only include spending out to a year after the AMI yet measure life expectancy out to 5 years after. This appears to preclude the possibility that there may be additional medical care and spending in the second through fifth years after the AMI that contributes to the improvements in outcomes, such as additional outpatient drugs or counseling on lifestyle changes. In general, in this literature, most papers compare spending and outcomes over the same period, a practice that seems advisable.

Cutler et al.'s work 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.

Cutler and Meara (2000) conduct a cost-effectiveness analysis of the changes in treatment over time of low-birthweight infants. In 1960, at the beginning of the period they study, there was no medical treatment available for low-birthweight infants. By 1990, hospital costs for these infants had risen substantially as treatment options expanded. Meanwhile mortality and the propensity to develop lifelong health problems conditional on birthweight had dropped considerably. Cutler and Meara value the additional years of life added using the standard estimate of $100,000 per year and discounting at 3 percent per year. They adjust for disability by assuming severe and moderate disabilities have a QALY weight similar to that of severe conditions like diabetes and heart disease. Conditional on birthweight, they assume all improvements in outcomes are due to medical care following the medical findings on this issue. For costs, they consider not just the costs of the hospitalization at birth but also additional lifetime

medical and special education costs. They find that the technological improvements in caring for low-birthweight infants have an outstanding rate of return of over 500 percent with the return for the subgroup of infants with birthweights from 1000 to 1500 grams being over 1800 percent. They find overall the increase in spending per QALY added to be $3,726, which is much lower than many other medical interventions.

Cutler and McClellan (2001) study productivity growth in the treatment of breast cancer from 1984 to 1991. During this period, incidence of breast cancer rose at the same time that the use of mammography increased. It is unclear therefore whether the rise in diagnoses reflected a true rise of breast cancer incidence in the population or increased detection of cases. They therefore conducted two analyses, one case-based and one population-based. The case-based analysis assumed that the increase in breast cancer cases was real and compared per-case expenditures and outcomes on the assumption that earlier patients were the same medically at the time of diagnosis as later patients. The population-based method assumes that the changes in incidence are entirely due to changes in screening and compares changes in spending and outcomes over the entire population. The case-based method suggests that the increased spending on breast cancer cases was well worth it while the population-based method finds that the increased costs and benefits are about equal. The paper therefore illustrates the difficulty of assessing health-care productivity when the use of diagnostic and screening services is increasing and conditions are being caught at an earlier stage.

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 in their base case 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.

One difficulty with their analysis has been noted by Garber and Skinner (2008) who pointed out that Cutler et al., when measuring life expectancy relative to future medical costs, discount future spending but not future life-years. As Garber and Skinner note, the standard practice is to discount both at the same rate to avoid the Keeler-Cretin paradox. The Keeler-Cretin paradox finds that if spending is discounted but life-years are not, it is optimal to delay all medical spending infinitely since it is better to save the money and buy more life-years in the future. Garber and Skinner recalculate Cutler et al.'s analysis while also discounting life-years and find that the cost per life-year gained for a 45-year-old is 2 – 2.5 times as much (depending on the decade) as what Cutler et al. found. Their revised results suggest that the increased medical spending between 1960 and 2000 may not actually have been worth the benefits.

Rosen, Cutler, Norton, Hu, and Vijan (2007) calculate the value of spending on coronary heart disease from 1987 to 2002, by considering both the value of spending on coronary heart disease from its effect on preventing heart attacks and the value of spending on heart attacks. They only include life expectancy as an outcome. 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), 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 the increased spending on coronary heart disease and on AMI generally was cost-effective in terms of incremental dollar per additional life-year saved over the period.

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. This method is similar to that of Cutler et al. (1998) and therefore has the same potential downfall; they fail to account for differential trends in behavior and the prevalence of other conditions in the AMI and non-AMI populations which could also affect the difference in trends in life expectancy for the two groups. In addition, their calculation of the value of spending on CHD appears to be fairly understated since they do not include the value of preventing strokes, whose incidence was also declining during this period.

Duggan and Evans (2008) estimate the cost per life-year extended of an important medical innovation of the last twenty-five years, the introduction of highly active antiretroviral therapy (HAART) for HIV/AIDS. Using medical claims from California's Medicaid program that are linked to mortality data, they exploit the rapid diffusion of HAART that followed the approvals and the introduction to the market of three antiretrovirals, Epivir and two drugs known as protease inhibitors, in 1995. The combination of two more antiretrovirals in HAART were much more effective at treating HIV than previous treatments and Duggan and Evans document a drop in quarterly mortality in HIV patients on Medicaid from 7 percent to 2 percent in the 18 months following the introduction of HAART as the patients took up the new therapy. In a regression-based framework conditioning on HIV severity, demographics and other health-care utilization, they find that HAART reduces mortality by 68 percent on average. In terms of spending, they found that higher quarterly pharmaceutical spending was more than offset by lower quarterly doctor and hospital spending and quarterly total medical spending for HIV patients was on average lower after the introduction of HAART. However, the increase in life expectancy among HIV patients was so dramatic that the present value of expected Medicaid spending for an HIV patient rose from $89,000 to $234,000. Duggan and Evans find that the cost per life-year was therefore around $19,000 and note that this value is well below standard valuations of a life-year. Since there were no other treatment innovations in this period and because the diffusion of HAART among patients was so rapid, Duggan and Evans are able to attribute the entire mortality improvement they see to HAART (conditional on the controls in their regression listed above). Duggan and Evans only consider mortality as an outcome; they

implicitly assume that HIV patients on HAART enjoy perfect quality of life even though HAART is accompanied by considerable side effects.

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 or improved quality of life. 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. One issue with their analysis, however, is that they count all of the medical spending of diabetes patients when measuring the change in spending. Since diabetics often have multiple comorbidities, it is likely that a good portion of that spending is on those other conditions. With Eggleston et al.'s approach, it is impossible to know how much of the change over time in patients' spending is attributable to diabetes and how much to the other conditions. If all of the increase in spending is due to spending associated with heart disease, for example, it is not accurate to label this analysis as an estimate of the value of diabetes treatment.

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 (GBD) study (US Burden of Disease Collaborative 2013). Outcomes for each condition are measured in disability-adjusted life-years (DALYs) by the GBD study. 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. Their analysis has two major downfalls. First, their paper is unusual among the outcomes-based papers in considering outcomes other than mortality. However, for disability outcomes, they are relying on the method used by the GBD study. The GBD study measured disability levels for each condition with 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). As noted by Cutler and Richardson (1997) and Dolan (2000), however, this approach for creating disability weights has been criticized since other research has shown that respondents actually in poorer states of health rate their own health higher than respondents with no personal experience of that state of health. The other problem with Highfill and Bernstein's approach is that 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

for heart attacks and coronary disease 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). In contrast to the other outcomes-based papers, they adopt the second approach described at the end of Section III of measuring the achievement of an endpoint of treatment rather than translating the benefits of medical care into QALYs and then monetizing them. They measured productivity as the ratio of outputs (number of survivors without a readmission) to inputs (estimated hospital treatment costs). They attempt to control for severity of the patients by taking into account basic patient demographics and comorbidities. They report productivity estimates both with and without controls and find that unadjusted productivity generally fell over the period but productivity adjusted for severity rose. However, as a published response to their paper noted, there is a potential problem with their adjustments for severity (Cram 2015). Romley et al. note themselves that the differences between unadjusted and adjusted productivity are driven by an increase in the number of comorbidities patients have been diagnosed with, apparently indicating that later patients are more ill upon admission. However, Cram (2015) pointed out that there exists considerable evidence of diagnosis upcoding during this period and that patients may not actually have been more ill in the later years of the period but simply have had more diagnoses coded. To this I will add that Romley et al.'s own summary statistics on the 2002 and 2011 cohorts (in

their Appendix) show that, other than the number of comorbidities reported, the 2011 cohorts for each condition look very similar to the 2002 cohorts. In particular, they have fairly equivalent levels of sensory, mental, and physical disabilities and rates of institutionalization. These similar levels of functioning suggest that the increased numbers of comorbidities reported may be an artifact of diagnosis practices and cast doubt on the positive results for hospital productivity growth that they find.

There are two main advantages of adjusting the measurement of medical output for quality based on observed outcomes. First, if mortality is the outcome being measured, life expectancy is relatively easily measured and relatively easy to put a dollar value on. Second, if sufficient controls are put in for other factors affecting health, observed outcomes reflect the actual productivity of the medical sector. As Duggan and Evans (2008) note, results from medical trials may not apply in the real world where treatment adherence may be at a lower level.

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 quantify 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 (US Burden of Disease Collaborative 2013). 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. Second, it is not easy to know how much of the change in observed life expectancy is due to medical care. Four of the nine papers discussed deal with this issue by only considering conditions with an acute onset and high mortality, which helps considerably with isolating the effect of medical care. Another paper, Duggan and Evans (2008), study the rapid diffusion of an extremely effective treatment of a

condition with very high mortality, a situation which simplified their analysis but which is not similar to the entrance of any other medical treatment of the past quarter century.

V.2 Process-based quality adjustment. The other basis for quality adjustment we observe in the literature is the value of medical processes or what treatments are given to patients. Adjusting the measurement of medical output for quality based on observed treatments has more data requirements than adjusting based on health outcomes. Adjusting based on outcomes only requires data on spending and on outcomes while adjusting based on treatments requires detailed data on which treatments were administered as well as medical knowledge of how effective the treatments are. Table 2 lists the papers that use process-based quality adjustments. Two papers studied mental health conditions and the other two studied cancer, with one focusing on a single cancer (colorectal) and the other studying all cancers.

Berndt, Bir, Busch, Frank, and Normand (2002) create a quality-adjusted price index for the outpatient 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 [selective serotonin reuptake 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

Table 2: Papers using process-based quality adjustments

Paper	Condition	Data	What is measured	Processes and how they are evaluated
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 treatment basket with quality held constant	Constructed treatment baskets based on treatment recommendations published by the Agency for Health Care Policy and Research and the National Institute of Mental Health
Lakdawalla, Shafrin, Lucarelli, Nicholson, Khan and Philipson (2015)	Colorectal cancer and multiple myeloma (drug treatments only)	IMS Health; National Comprehensive Cancer Network; IntrinsiQ; Medicare claims in SEER; Optum Touchstone claims; Citeline Trialtrove	Value of 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

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 an expenditure index measuring spending per remission of major depression and found that this price fell about 2 to 3 percent per year depending on specification.

With the use of an expert panel to measure the benefits of medical treatment, Berndt et al. take a unique approach to quality among the studies reviewed in this paper. However, the use of an expert panel is not without potential drawbacks since it is based on subjective judgments. Insofar as the panel consists of practicing clinicians, their estimates of the benefits of the treatments they provide may be upwardly biased across the panel. Another potential shortcoming of their paper is they only consider outpatient treatments and spending. Inpatient treatment, given the expense of individual stays, is likely to comprise a substantial share of spending on major depression, however, and including it in the study might have shed interesting light on the effect of the diffusion of SSRIs on inpatient spending by patients.

Frank, Berndt, Busch, and Lehman (2004) create a treatment-based price index for schizophrenia for the period 1994 to 2000 where treatment quality is held constant. They use medical claims data for Medicaid patients in two Florida counties for both treatments and spending. They place patients into treatment buckets and apply 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 control for quality by calculating expenditure indexes for treatment baskets where quality is held constant. They calculate a Laspeyres index where quality is held at the average level of

the base period, a Paasche index where quality is held at the average level of the final period, and a Fisher index which is the geometric mean of the two. Their method is unique in that, while they control for quality, the value of quality does not enter directly into the inflation measure. Rather, changes in the price level are only driven indirectly by treatment quality levels insofar as those levels affect the market shares of the different treatments. Their approach has the advantage that it does not rely on medical research to measure the value of treatments, other than for identifying the treatment baskets. It is therefore not subject to the potential biases of overly optimistic assessments of the value of medical treatments by clinicians and the values of the price indexes are not subject to revision when the results of medical research are revised. The disadvantage of their approach is that it is unclear how to deal with new treatment baskets, for whom the Paasche index cannot be constructed. In addition, if a new treatment basket entirely replaces the old treatment baskets, the Laspeyres index cannot be calculated for later periods either. Without the connection to some more universal metric such as QALY or achievement of a treatment endpoint, it is difficult to measure the relative value of a new treatment basket.

They found that, because of shifts from more expensive psychological treatments to less expensive pharmaceutical treatments, quality-constant treatment-based price indexes fell 5.5 percent annually.

Lakdawalla et al. (2015) calculate the net value of the improved health outcomes from innovations in cancer chemotherapy regimens.[6] The health benefits are measured in

[6] Lucarelli and Nicholson (2009) was the working paper version of the results for colorectal cancer presented in Lakdawalla et al. (2015). It calculated explicit price indexes for colorectal cancer, both unadjusted and adjusted for

QALYs and are based on the medical literature on the specific regimens. They find different results for different types of cancer and lines of treatment. For colorectal cancer, they find a slight negative net value from 1998 to 2005; the increased costs are slightly higher than the value of the added life. For first-line treatments for multiple myeloma, net value was substantially negative because there was no innovation in first-line treatments (so there were no improvements in health outcomes) but the price of the standard first-line treatment rose substantially. By contrast, there were innovations in second-line treatments for multiple myeloma, leading to large increases in net value as the improvements in health benefits were considerably larger than the increases in treatment costs. As they note themselves, however, their paper is limited because they only consider chemotherapy treatment costs, not other costs of cancer treatment. They point out that this omission is due to lack of nationally representative data on the costs of non-chemotherapy treatments for cancer.

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. They also only consider the costs of pharmaceutical treatments of cancer, not the costs of other treatments such as surgery.

Compared with outcomes-based adjustments, process-based adjustments have the advantage that they do not require data on actual outcomes. Analyses that use claims data, for example, usually

---

quality. It found that, while the unadjusted index rose enormously between 1998 and 2005, the adjusted index declined slightly.

do not have data on actual outcomes attached to the treatments patients are receiving. In addition, outcomes-based analyses generally have to consider the value of medical treatments in order to separate out the effect of medical care on outcomes, so process-based adjustments are not necessarily more work to implement.

The major disadvantage of quality adjustments based solely on processes is that they rely only on evaluations of the values of new treatments that are potentially subject to bias. One source of bias is that, if the adjustment relies on clinicians' opinions as does Berndt et al. (2002), clinicians can overvalue the benefits of the treatments they provide. Another source of bias, particularly relevant to the pharmaceutical treatments measured by Lakdawalla et al. (2015) and Howard et al. (2015), is that much of the research on the treatments' medical effectiveness is conducted by the treatment manufacturers or sellers themselves and is therefore subject to financial conflicts of interest (Bhandari et al. 2004, Goldacre 2013).

VI. 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 type of adjustment are summarized in Table 3.

Table 3: Summary of methods for quality adjustment of medical output

Type of adjustment	Advantages	Disadvantages
Outcomes-based	Data on mortality outcomes is available. Only need aggregate outcomes data. Reflects real-world productivity of medical care.	Require analysis of the contribution of medical care to health outcomes, which can require considering processes anyway.
Process-based	Do not require data on outcomes.	Requires more medical knowledge. Requires detailed data on treatments given. May be based on biased valuation of the benefits of medical treatment.

As Table 3 shows, in general, outcomes-based adjustments are perhaps easier to implement but require some strong assumptions, and how to measure disability and quality of life, which are both likely to be important, is unclear. Process-based adjustments, however, have higher data and medical knowledge requirements, and what medical knowledge is available may be biased.

The papers can be divided along a secondary axis by how they measure and value quality: in terms of QALYs, DALYs or simple life expectancy, by the achievement of a treatment endpoint, or by holding quality constant over time. Table 4 summarizes the advantages and disadvantages of each approach.

Table 4: Summary of methods for valuing quality		
	Advantages	**Disadvantages**
QALYs/DALYs/Life expectancy	Straightforward to compare the benefits of new treatments to those of old ones	Harder to construct Life expectancy requires looking forward Harder to translate impairments in terms of QALYs/DALYs
Achievement of endpoint	Do not have to value impairments in terms of QALYs or DALYs	Unclear how to handle if new treatment is so innovative that endpoint changes
Hold quality constant	Relies less on medical research for valuation of treatments. Results do not change if medical research revises benefits of treatments.	Unclear how to incorporate new treatment baskets.

The three options for valuing quality are ordered by how easily they incorporate new treatments. Valuations based on QALYs, DALYs or life expectancy are the most successful at measuring the relative value of new treatments since these are standards by which all treatments, new and old, can be judged. They are, however, the hardest to construct since there is currently no widely accepted method for translating impairments into QALYs or DALYs, a necessary step for comparing disability and mortality outcomes. The most ambitious effort to make those translations so far, the Global Burden of Disease study, uses a method described above which relies on the subjective views of the general population of what it is like to live in a certain health state and it is not clear how reliable those estimates are.

The second option, measuring the achievement of an endpoint of treatment, has the advantage of avoiding the necessity of valuing disabilities and impairments in QALYs or DALYs. This method of measuring quality also likely incorporates most new treatments well. However, it is unclear how to measure the relative value of a treatment that is so innovative that it changes the treatment endpoint of the condition. An example of such an innovation would be

the introduction of antiretrovirals for the treatment of HIV as detailed in Duggan and Evans (2008). Prior to the introduction of antiretrovirals, physicians were only able to manage the opportunistic infections arising secondarily to the primary HIV infection. With the approval of Retrovir (AZT) in 1987 and its followers such as Epivir and the other components of HAART in 1995, the goal of treatment turned into lowering viral load, a treatment goal that did not exist before. It would be unclear how to compare the costs of treatment before and after the introduction of Retrovir if only considering treatment endpoints.

The third method, used only by Frank et al. in the papers reviewed, has the advantage that it does not rely on valuations of quality at all but simply identifies treatment baskets and measures changes in the prices of episodes when the market shares of those treatment baskets are held constant, a method which holds average treatment quality constant. However, it has even more difficulty with incorporating new treatments than the second option of pricing an endpoint of treatment. While the second method only has problems with extremely innovative treatments, it is unclear with the third method how to introduce just about any new treatment or treatment basket. In Frank et al.'s study, they create treatment baskets by drug class (for example, one treatment basket is "any atypical antipsychotic"). If, however, a new member of the class is introduced that is of markedly higher effectiveness, the treatment basket's quality will no longer be comparable. In addition, if a new class of drugs is introduced creating a new treatment basket, it will be unclear how to construct the Paasche index with it since there is no price for it in the first period. These situations arise more often than an entirely innovative treatment that changes the endpoint to be measured.

Table 5			
	Method for valuing quality		
Basis for quality adjustments	**QALYs/DALYs/Life expectancy**	**Achievement of treatment endpoint**	**Hold quality constant**
Outcomes	Cutler et al. (1998) Cutler and Meara (2000) Cutler and McClellan (2001) Cutler et al. (2006) Rosen et al. (2007) Duggan and Evans (2008) Eggleston et al. (2009, 2011) Highfill and Bernstein (2014)	Romley et al. (2015)	n/a
Processes	Lakdawalla et al. (2015) Howard et al. (2015)	Berndt et al. (2002)	Frank et al. (2004)

Table 5 organizes the papers reviewed along the axes described above. It should be noted, of course, that all of these papers fulfil the guidelines given in the OECD/Eurostat/WHO report "A System of Health Accounts" (OECD 2011) and listed in Section V., with the exception of one, that quality adjustments should address the impact of effects on lifetime health expenditures. The only papers that address this issue are Cutler and Meara (2000), Cutler et al. (2006), Rosen et al. (2007), and Duggan and Evans (2008).

However, all of these papers still suffer from limitations relative to the ideal paper which would accurately measure the change in the entire marginal contribution of medical care for a particular condition and compare it to the change in spending for that condition. All but two of the papers using outcomes-based adjustments only include mortality as an outcome (or the risk of mortality in the case of Eggleston et al. 2009 and 2011) and not disability or quality of life. The adjustments likely understated or overstated the true value of the spending and correspondingly overstated or understated medical inflation. Highfill and Bernstein (2014), one of the two papers to consider non-mortality outcomes, is limited, however, in that it does not address the role of other factors besides medical care in improving health outcomes and therefore

its results likely overstate the net value of health care and understate inflation. The other significant shortcoming shared by many of the papers is measuring spending in some way that does not tie closely to the quality of treatment for the condition. The shortcoming is evinced by measuring spending and outcomes over different periods (Cutler et al. 1998), by measuring all spending rather than just spending on the condition (Eggleston et al. 2009, 2011; Duggan and Evans 2008), or by measuring spending in just one category rather than all spending for the condition (Romley et al. [2015] only measure hospital spending, Lakdawalla et al. [2015] and Howard et al. [2015] only measure pharmaceutical spending, and Berndt et al. [2002] only measure spending on outpatient care).

Where to proceed from here? While the literature so far has had its limitations, those limitations only reflect the intensity of resources required for this type of analysis. These resource requirements led to the recommendation that the statistical agencies such as BEA and BLS address this issue in the official statistics (National Research Council 2009, 2010). The data needs are considerable and I will now summarize them.

The first need is for data on spending by condition. The Health Care Satellite Account addresses this need although it should be noted that it is based on what spending data is available, and coverage of certain populations is better than others. In general, claims data for spending seem to be preferable to survey data. Survey data is prone to underreporting of spending; Zuvekas and Olin (2009) find significant underreporting in the MEPS when they compare the reported spending of Medicare beneficiaries to linked Medicare claims and the documentation for the MCBS notes that spending collected by survey is considerably lower than the total given in the linked claims. Survey data also has smaller sample sizes than claims data, an issue that is especially relevant when measuring spending by condition, and Dunn et al.

(2015) find that the version of the Health Care Satellite Account based solely on the MEPS has much higher volatility in the indexes and larger confidence intervals than the "blended" account which incorporates claims where available.

The second need is for either data on outcomes by condition or for data on treatments given and their medical value, depending which kind of quality adjustment is being made. In fact, using both may be preferable as the two methods are not necessarily mutually exclusive. Eggleston et al. (2009, 2011) arguably combine them since their measured outcome is derived from an established model of disease progression based on medical research. Several papers consider both outcomes and processes in their discussions although all of them only use one or the other in their calculations. Cutler et al. (1998), Rosen et al. (2007), and Duggan and Evans (2008) discuss extensively the value of the treatments they observe being given while ultimately basing their analyses on mortality outcomes. Berndt et al. (2002) meanwhile compare their process-based estimates with outcomes inferred from diagnoses later in the patients' history in the medical claims data. One could imagine an approach that measured both outcomes and the value of processes, and used the value added by the processes to directly adjust the outcomes-based measures. This approach would be quite resource-intensive and have very high demands for data on spending, on outcomes, and on treatments given and for knowledge of the medical effectiveness of those treatments. These data and knowledge requirements explain why the papers above all take approaches that are somewhat limited versions of the ideal.

There are strong arguments, however, for pursuing better measurement of the health-care sector. First, the results of the papers reviewed here show that quality adjustment of health care output could have substantial effects on measured inflation. 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.

Second, accurate economic measurement is important for setting economic policy. There are two major sets of policies that depend on the measurement of inflation, output, and productivity in health care. First, health care forms a large and growing part of the economy and inaccurate measurement of its output and productivity means output and productivity of the economy as a whole are significantly mismeasured. This mismeasurement hampers the conduct of monetary policy by the Federal Reserve. As part of its dual mandate to keep unemployment and inflation low, the Federal Reserve is interested in understanding all aspects of the labor market, including labor productivity. Health care is a labor-intensive sector and in 2014, health care and social assistance accounted for 12.0 percent of jobs in the United States (Henderson 2014). Short-term developments such as those contained in the GDP release or the monthly employment situation report must be interpreted by the Federal Reserve in light of what is known about long-term trends in labor productivity. In addition, the understanding of overall economic productivity underlies the formulation and evaluation of many economic policies by agencies such as the Council of Economic Advisers, the Office of Management and the Budget, the Treasury Department, and the Congressional Budget Office (CBO).

The second issue confronting decision-makers is the pressure exerted by rising health-care costs on the long-term fiscal situation of governments at the federal, state, and local levels. The Congressional Budget Office attributes much of its projected increasing federal deficits over the next ten years to the rising costs of Medicare (Congressional Budget Office 2016). For state and local governments, a report by the Government Accountability Office (GAO) forecasts that they will have considerably lower revenues than spending over the next half-century, largely

because of Medicaid costs and the costs of health insurance for their employees (Government Accountability Office 2015). At some point, these deficits at all levels of government will become unsustainable and policymakers will have to choose between cutting benefits or raising taxes. When that point arrives, it would be helpful to know more about the overall value of health-care spending. Moreover, it is hard for governments to decide on levels of investment across sectors without knowing the value of those investments.

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" (National Research Council 2009). The discussion here reflects that position.

References

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.

Berndt, E. R., Cutler, D. M., Frank, R. G., Griliches, Z., Newhouse, J. P. and Triplett, J. R. (2001). Price Indexes for Medical Care Goods and Services: An Overview of Measurement Issues. 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. 141-198). Chicago, IL: University of Chicago Press.

Bhandari, M., Busse, J. W., Jackowski, D. et al. (2004). Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized trials. *CMAJ* 170(4): 477-480.

Bresnahan, T. F. and Gordon, R. J. (1996). Introduction to "The Economics of New Goods." In T. F. Bresnahan and R. J. Gordon (eds.), *The Economics of New Goods* (National Bureau of Economic Research Studies in Income and Wealth Series, vol. 58, pp. 1-26). Chicago, IL: University of Chicago Press.

Bureau of Economic Analysis (2015). *NIPA Handbook: Concepts and Methods of the U.S. National Income and Product Accounts.* http://www.bea.gov/national/pdf/all-chapters.pdf. Last accessed 2/11/2016.

Bureau of Labor Statistics (2015), *Handbook of Methods*, Chapter 17: The Consumer Price Index. http://www.bls.gov/opub/hom/pdf/homch17.pdf. Last accessed 2/10/16.

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.

Congressional Budget Office (2016). The Budget and Economic Outloook: 2016 to 2026. January 25, 2016. https://www.cbo.gov/publication/51129. Last accessed 6/8/2016.

Cram, P. (2015). Assessing Hospital Productivity. *Health Affairs* 34(6): 1069. doi: 10.1377/hlthaff.2015.0454.

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. and McClellan, M. (2001). Productivity change in health care. *American Economic Review* 91 (2): 281-286.

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. and Meara, E. (2000). The Technology of Birth: Is It Worth It? In A. M. Garber (ed.), *Frontiers in Health Policy Research, Volume 3* (National Bureau of Economic Research), pp. 33-68. Cambridge, MA: MIT Press.

Cutler, D. M. and Richardson, E. (1997). Measuring the Health of the U.S. Population. *Brookings Papers on Economic Activity: Microeconomics* 1997: 217-282.

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).

Dolan, P. (2000). The Measurement of Health-Related Quality of Life. In A. J. Culyer and J. P. Newhouse (eds.), *Handbook of Health Economics: Volume 1B*, pp. 1723-1760. New York: Elsevier North Holland.

Duggan, M. G. and Evans, W. N. (2008). Estimating the Impact of Medical Innovation: A Case Study of HIV Antiretroviral Treatments. *Forum for Health Economics & Policy* 11(2): Article 1.

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.

Goldacre, B. (2013). *Bad Pharma: How Drug Companies Mislead Doctors and Harm Patients.* New York: Faber & Faber.

Government Accountability Office. (2015). State and Local Governments' Fiscal Outlook: 2015 Update. http://www.gao.gov/assets/680/674205.pdf. Last accessed 6/8/2016.

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.

Henderson, R. (2015). Industry employment and output projections to 2024. *Monthly Labor Review* December 2015. http://www.bls.gov/opub/mlr/2015/article/industry-employment-and-output-projections-to-2024.htm.

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.

Kornfeld, R., Hartman, M. and Catlin, A. (2010). Health Care Expenditures in the National Health Expenditures Accounts and in Gross Domestic Product: A Reconciliation. Bureau of Economic Analysis working paper (http://www.bea.gov/papers/pdf/healthrecon_workingpaper_Sep2010.pdf). Last accessed February 2, 2016.

Lakdawalla, D., Shafrin, J., Lucarelli, C., Nicholson, S., Khan, Z. and Philipson, T. (2015) Quality-Adjusted Cost of Care: A Meaningful Way To Measure Growth In Innovation Cost Versus The Value of Health Gains. *Health Affairs* 34(4): 555-561.

Landefeld, J. S., Seskin, E. P., and Fraumeni, B. M. (2008). Taking the Pulse of the Economy: Measuring GDP. *Journal of Economic Perspectives,* 22(2): 193-216.

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).

Martin, A. B., Hartman, M., Benson, J., Catlin, A., and the National Health Expenditure Accounts Team. (2016). National Health Spending in 2014: Faster Growth Driven By Coverage Expansion And Prescription Drug Spending. *Health Affairs* 35 (1): 150-60. DOI: 10.1377/hlthaff.2015.1194.

National Research Council (2002). *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes.* Panel on Conceptual, Measurement, and Other Statistical Issues in Developing Cost-of-Living Indexes, Charles L. Schultze and Christopher Mackie, Editors. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.

National Research Council (2009). *Strategies for a BEA Satellite Health Care Account: Summary of a Workshop*. Christopher Mackie, rapporteur. Steering Committee for the Workshop to Provide Guidance for Development of a Satellite Health Account at the Bureau of Economic Analysis. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.

National Research Council (2010). *Accounting for Health and Health Care: Approaches to Measuring the Sources and Costs of Their Improvement.* Panel to Advance a Research Program on the Design of National Health Accounts, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.

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.

Zuvekas, S. H., and G. L. Olin (2009), “Accuracy of Medicare Expenditures in the Medical Expenditure Panel Survey”, *Inquiry* 46 (Spring): 92-108.