Are Medical Care Prices Still Declining? A Systematic Examination of Quality-Adjusted Price Index Alternatives for Medical Care

by

Seidu Dauda, World Bank Group, (email: sdauda@worldbank.org)
Abe Dunn, Bureau of Economic Analysis, (email: abe.dunn@bea.gov)
Anne Hall, U.S. Department of the Treasury, (email: anne.hall@treasury.gov)

Draft April 2018
Comments welcome

Abstract: Health-care spending has grown rapidly over the past several decades but the societal benefits of these expenditures are not measured well. We address this topic within the framework of price measurement. Specifically, we construct quality-adjusted price indexes for three acute medical conditions for the period 2001-2014 using data on Medicare fee-for-service (FFS) beneficiaries. We find that quality adjustment has a large effect on price growth and that a substantial bias exists if one does not adjust for changes in quality. Measured through a benchmark cost-of-living index (COLI), conservative estimates show average prices for these conditions declining by around 5.1 percent per year, relative to economy-wide inflation. Applying this quality adjustment to BLS productivity estimates would increase annual multifactor productivity from the current annual rate of -0.3 percent to an annual rate of 2.9 percent. When comparing alternative methods of quality adjustment, we find a wide dispersion in the growth rates. According to our theoretical model, this dispersion results from the increases in benefits exceeding the increases in spending for these conditions. Only the benchmark COLI fully captures the value of the increases in benefits to patients.

1 We would like to thank Ben Bridgman, David Byrne, Dennis Fixler, Bronwyn Hall, Brett Matsumoto, John Romley, Louise Sheiner, Jon Samuels, Jon Skinner, and the experts on the Bureau of Economic Analysis Academic Advisory panel for comments, and Lasanthi Fernando for editing assistance.
1. Introduction

In a seminal work in the health economics literature, Cutler, McClellan, Newhouse and Remler (1998), asked the question: “Are medical prices declining?” The question of the true level of inflation in health care is one of the most challenging and important measurement problems in economics. Measuring the prices of treatments for heart attacks, Cutler et al. find that after accounting for quality improvement, the price of treatment did decline over their period of study, even while the unadjusted price of treatment rose. These findings suggest that the mismeasurement of health care output and productivity could be substantial. It has major implications for our beliefs about individual welfare and economy-wide real output given that health care is such a large share of the economy.

The topic has only grown in importance in the past two decades, as the share of GDP devoted to health care rose from 13 percent of GDP in 1998 to nearly 18 percent in 2016 (Hartman et al., 2017). Meanwhile, life expectancy at birth has increased by nearly two years over the same period with medical innovations possibly playing an important role. It is widely accepted, however, that the current methods for measuring inflation in the health-care sector used by the major US statistical agencies such as the Bureau of Labor Statistics (BLS) and the Bureau of Economic Analysis (BEA) are unsatisfactory, in large part because they fail to adjust for the value of the improvements in treatments over time. They therefore may overstate inflation for many conditions.

The statistical agencies face a formidable challenge with health care because their normal methods for quality adjustment when measuring prices do not work for health care for a number of reasons. The rapid pace of technological change in the sector, the widespread use of third-
party reimbursement, and the particular characteristic of health care that spending to extend life
can imply that the marginal utility of additional health technology greatly exceeds its marginal
cost preclude the use of standard methods for quality adjustment of price indexes. Research on
novel methods for quality adjustment specific to health care has been sporadic, however. As
detailed by Hall (2016) and Sheiner and Malinovskaya (2016), no clear consensus on the most
appropriate method as papers following Cutler et al. (1998) all used different methods from
Cutler et al. and from each other with no discussion of the connections among them.

The goal of this paper therefore is to provide a better understanding of the theoretical, practical,
and empirical differences across alternative methods of quality adjustment for health care price
indexes that have been applied in the literature. We first discuss potential theoretical differences
among the methods with a stylized model of health care quality change and spending. We
consider four methods: a cost-of-living (COLI) index based on consumer utility theory and the
monetary value of extended life (Cutler et al., 1998), a quality-constant price index based on the
methods set out in Berndt et al. (2002), a hedonic fixed-technology index based on Frank et al.
(2004), and an index with the quality adjustment based on the change in costs to produce a new
technology.

The theoretical comparison shows that if the value of the outcome improvement is close to the
increased spending or cost of treatment, the quality-adjusted indexes deliver similar results
across methods. However, if the increase in value greatly departs from the increase in spending,
the cost-of-living index constructed with the method of Cutler et al. (1998) gives very different
results from the other indexes because it is the only methodology to incorporate the full value of
the consumer’s increased benefit.
After reviewing the theoretical differences, we apply the different methods to estimate price indexes for three acute conditions among FFS Medicare patients for the years 2001-2014: acute myocardial infarction (AMI), congestive heart failure (CHF), and pneumonia. The average expenditure per treatment for these conditions rise faster than general inflation, having an average excess growth rate of around 1 percent per year. However, these conditions also showed significant improvements in outcomes as measured by post-hospitalization life expectancy during this period. Not surprisingly, given the improvements in outcomes, the quality-adjusted indexes all give lower growth rates in prices with the COLIs showing the steepest declines. Applying the COLI quality adjustment method, and using the typical value of a life year used in the literature of $100,000, we find that the price of treating these conditions declines by 7.6 percent per year on average relative to aggregate inflation. Even using a more conservative estimate of a value of a life year of $50,000 shows these prices declining by 3.3 percent per year.

As expected from our simple theoretical model, we find that the COLI tends to fall much faster than other quality-adjustment methods that we explore. The divergence is caused by the benefits from the increase in life expectancies exceeding the increase in spending. For AMI, by one estimate, the increased benefits through the gains in life expectancy from 2001 to 2014 are worth more than one and a half times as much as the initial level of average spending per patient while that average declined slightly over the same period after adjusting for general inflation. Alternative methods and assumptions produce lower changes in quality-adjusted prices, however. All of the quality-adjusted indexes are sensitive to key assumptions but we find that even under the most conservative assumptions, the quality-adjusted indexes grow at about 1 percentage point less per year than the unadjusted indexes, highlighting the importance of quality adjustment.
These estimates have important implications for the measurement of productivity growth. The BLS estimates multifactor productivity growth for the hospital and nursing home sector to be negative over the 2001-2014 period, with an annual growth rate of -0.3 percent. Under the strong assumption that our conservative measure of quality adjustment is representative of the hospital sector more broadly, we find that it implies a multifactor productivity growth rate of 2.1 percent, holding inputs constant.

2. Background on price indexes in health care

While there are numerous measurement issues that arise in the construction of health care price indexes, there is general agreement among experts that these indexes should track the prices of treating specific conditions (National Research Council 2010, OECD/Eurostat/WHO 2011). Currently, the BLS measures the prices of individual medical services although it has moved towards condition-based measurement for the PPI for general medical and surgical hospitals. As noted in National Research Council (2010), however, with a treatments-based framework, analysts can better measure changes in practice patterns, technologies, outcomes, and associated expenditures on treatments relevant to a condition. Previous research on quality adjustment of price indexes in medical care has all been performed on condition-based indexes (Cutler et al., 1998; Berndt et al., 2002; Frank et al., 2004). Moreover, both the BLS and the BEA have already taken the initial experimental steps to redefine price as the expenditures to treat an episode of a condition (Bradley et al. 2015, Dunn et al., 2015). Throughout this paper therefore,
we will be discussing the quality adjustment of medical price indexes based on the expenditures to treat a condition.  

The particular characteristics of health care make it a difficult sector to construct quality-adjusted price indexes for, so it is not surprising that the practice of economic measurement as applied to health care has not advanced further than it has. When constructing the consumer price index (CPI) for series for which it constructs a quality-adjusted index, the BLS employs a quality adjustment based on hedonic regressions relating characteristics of products to prices. For the producer price index (PPI), it makes an adjustment based on the production cost of the quality change being implemented for a good on the assumption that production technologies are being held constant. For goods which see rapid change in production technologies (such as computers and other high-technology goods), the BLS also uses adjustments based on hedonic regressions (Bureau of Labor Statistics 2013).

Both the cost-based approach and hedonic regressions rest on the assumption that marginal utility of the good or service being purchased is equal to its marginal cost and therefore that the marginal valuation of a quality change can be inferred from the change in costs or in price related to change. As discussed in Triplett (1982), in a competitive equilibrium, the consumer-based (price change) and producer-based (production cost) approaches should give similar results.

The purchase of health care goods and services, however, does not meet this assumption for a number of reasons. First, it is also a sector which sees rapid changes in technology, with new

---

2 It is worth noting that there are still many practical challenges in how to allocate expenditures to conditions and how conditions should be defined, but there is general consensus on the fundamental idea of a condition-specific focus.

drugs, devices, and diagnostic and treatment technologies being introduced on a regular basis. As noted above, for industries like these, the BLS uses a hedonic-based adjustment rather than a cost-based one since it is impossible to hold production technologies constant in such a rapidly changing environment.

The other features of health care, however, preclude the use of hedonics, which require the marginal utility derived from a product characteristic to equal its marginal price. Two major distinguishing characteristics of health care in fact generate reverse predictions about the relationship between marginal valuation and marginal spending. First, the widespread use of third-party reimbursement suggest that marginal utility could be less than marginal cost. Patients typically only pay a fraction of their health-care spending on the margin. It has been shown extensively in the literature that patients reduce health-care utilization in the face of higher cost-sharing, suggesting that for them the marginal utility of the reduced care was lower than that care’s total cost (Manning et al. 1987, Finkelstein et al. 2012).

Conversely, however, the marginal utility of health care may greatly exceed its marginal cost because of the unique purpose of health care spending: to extend life from which further utility may be derived. As Hall and Jones (2007) point out, the marginal utility of consumption goods may drop rapidly as consumption rises but “the marginal utility of life extension does not decline.” Sheiner and Malinovskaya (2016) note that consumers of health care are often up against a technological constraint that they would like to purchase more life but are simply not able to do so for any amount of money. Not only do two bypass surgeries not add twice as much health as one, the second surgery would probably add zero or negative health. Once the technological constraint is lifted, however, with the introduction of a new treatment technology, the new technology therefore likely raises spending but still delivers considerably more utility
than it costs. As Sheiner and Malinovskaya (2016) show and as we will also show below, because of this feature of health care, it will matter for which method of quality adjustment is used for health-care price indexes.

Finally, a third characteristic of health care makes the use of hedonic-based methods inapposite and that is patients’ limited understanding of the effects of medical care and providers’ incentives to provide more care in a fee-based reimbursement system. While it has been shown that patients respond to out-of-pocket costs, there is also evidence that this response is not always optimal. Brot-Goldberg et al. (2017), for example, found that beneficiaries who moved to an insurance plan with high cost-sharing reduced potentially high-value and low-value services at the same rate. This non-optimality is probably due to beneficiaries not understanding the medical value of different kinds of care. In such an environment, the relationship between marginal utility and consumer behavior is unclear and it is hard to infer value from consumer behavior.

There is evidence that this measurement problem may be widespread in the health care sector. New measures of spending by disease from the BEA show that expenditures per treatment, unadjusted for quality, exceed economy-wide inflation over the 2000-2014 period by about 2 percentage points per year.⁴ Innovative new treatments are an important contributing factor to this high growth rate. Examining BEA data, Dunn et al. (2018) finds technology to be a key factor behind the rapid growth in expenditures for many conditions, including treatments for rheumatoid arthritis, cancer, hepatitis, and HIV. For example, Sovaldi, approved by the FDA in 2013, and other new drugs to treat chronic hepatitis C are more effective than previous treatments, but they are also expensive, costing between $60,000 and $100,000 per patient

---

⁴ This figure was computed based on inflation calculated for the Blended Health Care Satellite Account.
The latest drug treatments for rheumatoid arthritis, a disease that causes pain, stiffness and loss of function in the joints, are effective at reducing pain and improving mobility, but are priced at over $30,000 per year. In addition, the use of effective non-drug treatments for rheumatoid arthritis, such as joint replacements, has increased rapidly in recent years despite the substantial costs (Fingar et al. 2014). New treatments for HIV and cancer have also been shown to be both costly and effective (Girouard et al. 2010 and Howard et al. 2015). The measurement problem likely goes beyond these specific examples, as many economists view the health care sector as an area of significant innovation where traditional measures of inflation greatly overstate actual inflation (Lebow and Rudd 2003, Groshen et al. 2017). While these new innovations contribute to the rising expenditures per treatment and the expanded share of GDP devoted to medical care, the quality improvements are currently not reflected in our national statistics.

Some evidence of quality change in the health care sector may be gleaned from national trends in mortality rates and elderly individuals living more disability-free years (Cutler et al. 2006; Cutler et al. 2017). However, as noted in reviews of the literature by Hall (2016) and Sheiner and Malinovskaya (2016), it can be extremely challenging to accurately attribute changes in the health of the population to the medical care sector, as non-medical factors may also play an important role. For this reason, solving this measurement problem for a broad range of conditions is an enormous hurdle. However, even for the conditions studied in this paper that are more amenable to the study of quality adjustment, Hall (2016) and Sheiner and Malinovskaya (2016) find there is no widely accepted method for creating quality-adjusted price indexes for

---

6 “Rheumatoid arthritis drug prices on the rise”, Modern Healthcare, April 1, 2016, by Adam Rubenfire.
medical care, even though the field of health economics is generally devoted to measuring the benefits of health care against its costs.

Adjusting condition-specific price indexes for quality is also necessary to improve the measurement of output and productivity in the health care sector and to fit this measurement into the broader framework of the system of national accounts. In general, macroeconomic literature attempting to understand the observed changes in our economy assumes that health care is measured in a similar fashion to the rest of the economy (Kongsamut, Rebelo, and Xie 2001, Ngai and Pissarides 2007, and Herrendorf, Rogerson and Velentinyi 2013). Further, because of its large share of the economy, the measurement of health care lies at the center of some recent macroeconomic puzzles. For example, one explanation for the rise in health spending is a scenario suggested by Baumol (1967), where more expenditures may shift to labor-intensive sectors with little technological improvement, when individuals are price-inelastic or income-elastic for goods and services in that sector. As discussed in Cutler et al. (2006), however, health care has seen significant technological change which have improved health and mortality outcomes over the past 60 years. The reverse scenario could therefore be the case, that resources are shifting to health care in response to the relative prices for health care falling.

A related economic puzzle is the current slowdown in measured productivity growth in the U.S., which has received considerable attention and one ongoing debate has been over whether or not the slowdown is real or an artifact of mismeasurement (Byrne et al., 2016; Syverson, 2017). Byrne et al. (2016) do find that, even if the weights of health care and other poorly measured industries were held fixed, the slowdown still exists. However, if productivity in health care were growing faster than measured, aggregate productivity would also be growing faster whether the weights of health care are held fixed or not. More generally, from society’s perspective, how the
increasing expenditures flowing to this sector compare to benefits is a topic of broad interest to researchers, policy-makers and individuals.

While this question may seem challenging, this issue is commonly addressed for the more traditional sectors where we also observe substantial innovation. We argue that the guiding theories and principles behind price measurement in the health care sector should have theoretical foundations shared by the rest of the measurement system. For this reason, we view the ideal quality-adjusted index as a cost-of-living index (COLI), in accordance with the guidelines laid out in “At What Price?” (National Research Council, 2002). The COLI captures the change in the value of expenditures necessary for a consumer to be equally satisfied across periods. Adjusting for inflation using a COLI-based measure of price growth, translates nominal expenditures into a welfare-based change in output. The COLI approach to quality adjustment is directly applicable to health care, as health is a component of personal consumption expenditures (PCE), or consumer spending.

**Previous research on quality-adjusted price indexes in health care**

Because of the price measurement issues specific to health care discussed in the previous section, other methods have been proposed in the literature, as has been reviewed by Hall (2016) and Sheiner and Malinovskaya (2016). Cutler et al. (1998) constructed an index for heart attack treatment theoretically grounded on a COLI with a quality adjustment based on observed mortality outcomes, as measured by the value of life extended by medical treatment. Between 1984 and 1991, post-incident life expectancy for heart attack patients rose by four months relative to non-heart attack patients, an increase which, even at conservative estimates of the

---

7 More formally, they recommend a conditional COLI, which is “conditional” in the sense that it ignores particular factors that are outside of a pre-defined scope, such as public goods or the weather.
value of life, represents a considerable amount of value. They calculate that the value of life extended by improved heart attack treatment was around $17,000 per patient but that treatment costs for those patients only increased by about $3,200, resulting in a price index which drops 1.5 percent annually.

Cutler’s approach is grounded in a COLI but an intuitively appealing alternative is to hold quality constant by pricing the same service or outcome over time. In an application to health markets, Berndt et al. (2002), for example, define the output as the number of successful treatments rather than as the number of treatments. In this case, the price to be measured is the incremental price of achieving a certain endpoint or goal of treatment relative to receiving no treatment at all. Berndt et al. measure the price per remission of major depression and this approach is also taken by Romley et al. (2015) who define output as survival to a certain endpoint following a hospitalization. A similar approach is to measure the price over time of a quality-adjusted life-year (QALY) added by medical treatment (Lucarelli and Nicholson 2009, Howard et al. 2015, Hult et al. 2017).

Frank et al. (2004) create a price index for schizophrenia by defining treatment baskets for schizophrenia based on treatment guidelines issued by medical experts. Rather than directly holding quality constant, this approach holds constant the relevant technologies or treatments. Quality is thus held constant on the assumption that the quality change is entirely reflected in the price change due to the observed technologies being applied. This index is very similar to a hedonic regression model, where researchers must carefully select the appropriate technologies to be held constant. In this case, the hedonic model controls for particular technologies, while allowing for shifts in all other treatments and services, where quality is implicitly assumed to be unchanged.
A final approach for measuring quality is to adjust for the “resource cost” of the producer (e.g., hospital or physician) necessary to produce the change in quality. This resource-cost approach is one of the methods applied in the Producer Price Index (PPI) by the BLS. One disadvantage is that it requires strong assumptions regarding the cost of inputs and their relationship to quality. In this framework, quality cannot improve without costs going up, which poses special problems in health care. For example, Triplett (1982) cites the counterexample of birth control pills where reducing the dose by half lowered costs while delivering the same value to the consumer.

Moreover, in health care, quality can rise even if the price decreases, for example if low-value treatments and diagnostic tests are reduced. No price index has yet been constructed in the research literature for health care by adjusting for resource costs. However, the BLS calculates a cost-based quality adjustment in the hospital PPI and nursing home PPI, which are traditional service-based indexes, not episode-based indexes as studied here.8

As this section has described, multiple approaches have been taken in the literature but without any comparison of them, either empirical or theoretical. Over the next several sections, we will lay out the results of a simple theoretical model connecting the methods and showing how their results might differ and why Cutler et al.’s COLI-based method is to be preferred to the other methods.

8 It is our understanding that the BLS quality adjustments for health care currently has a very small impact on inflation rates for this sector, but we have not been provided with exact magnitudes. These quality adjustments are based on quality indicators for hospitals collected by the Department for Health and Human Services. However, the index has some limitations. First, the BLS calculates a composite quality score index for each condition where the weights for each quality score are based on how many people the score applies to. As they note themselves, it would be preferable to weight the scores by medical importance but data for that is not easily available. The percent change in the composite score is then multiplied by an estimate of costs formed from the median charges for the Diagnosis-Related Group (DRG) and the cost-to-charge ratio for operating and capital costs. As they also note, it would be preferable to use more direct proxies for costs of each quality score but they were unable to identify ones for each score. There is also a quality adjustment for nursing homes that is based on the nursing staff ratios. Their approach illustrates the potential difficulties of constructing a quality-adjusted price index in health care based on changes in resource costs.
3. A Comparison of Methods

In this section, we lay out the consumer theory behind the COLI and describe the results of a simple model showing the relationship between the methods for quality adjustment in health care that have been used previously. (The details of the model are given in Appendix 1.)

From the consumer’s perspective, the theoretical ideal for a price index is a COLI conditional on the environment and population. A COLI is written as:

$$ e(p_1, U_0) \over e(p_0, U_0) $$

(1)

where \( e(\cdot) \) is the expenditure function that expresses the minimum expenditure to achieve a certain level of utility given a certain set of prices \( p \). The COLI is the ratio of expenditures needed at new prices to maintain base-period utility in period 1 to the expenditure in the base period.

The COLI is often rewritten as \( \frac{Y-CV}{Y} \) where \( Y \) represents income, which is held constant over time so \( Y = e(p_0, U_0) = e(p_1, U_1) \). \( CV \) is the compensating variation, the monetary amount required to compensate consumers in period 1 to restore them to their utility level of the base period, given period 1 prices. Cutler et al. (1998) create quality-adjusted price indexes for heart attacks by approximating the compensating variation from its utility-based definition:

$$ U(H_1(m_1), Y - p_1 m_1 - CV) = U(H_0(m_0), Y - p_0 m_0) $$

(2)

In this equation, \( m_1 \) is the quantity of medical care, \( p_1 \) is the price of medical care, and \( H_t(\cdot) \) translates medical care into health \( H \). Taking a first-order Taylor approximation at period 0 yields:
\[ CV = \frac{U_H H_m}{U_x} (m_1 - m_0) - (p_1 m_1 - p_0 m_0) \]  

(3)

where \( U_H \) is the marginal utility of health, \( H_m \) is the marginal effect on health of medical care, and \( U_x \) is the marginal utility of non-health consumption \((x = Y - p_0 m_0)\). Cutler et al. (1998) note that the first term in this equation is the increased benefit in monetary terms of medical care for the condition being treated \((\Delta B = B_1 - B_0)\) and the second term is the change in spending on the condition \((\Delta S = S_1 - S_0 = p_1 m_1 - p_0 m_0)\) where \( S_t = p_t m_t \).

The COLI for a patient with the condition is therefore:

\[ COLI_{pat} = \frac{Y - (\Delta B - \Delta S)}{Y} \]  

(4)

In general, however, we are interested in the COLI across the population, not just for individuals with the condition. Sheiner and Malinovskaya (2016) therefore adjust this index by the prevalence of the condition \( r \):

\[ COLI_{pop} = \frac{Y - r(\Delta B - \Delta S)}{Y} \]  

(5)

From this expression, they then derive the quality-adjusted price index for the individual condition. Let \( \alpha \) be the per capita income share of spending on the condition in the initial period: \( \alpha = \frac{r S_0}{Y} \). Assuming homothetic preferences, the change in quality-adjusted prices for the condition multiplied by its income share will be equal to the change in the COLI, holding all other prices constant:

\[ \alpha \left( \frac{S_1 - S_0}{S_0} \right) = \frac{-r(\Delta B - \Delta S)}{Y} \]
Solving for $\hat{S}_i$, we obtain $\hat{S}_i = S_1 - \Delta B$. So the quality-adjusted price index for this individual condition is equal to:

$$\frac{S_1 - \Delta B}{S_0}$$

(6)

This formula in fact corresponds to the formula for a direct quality adjustment in the Consumer Price Index (CPI) by the BLS (Bureau of Labor Statistics 2015). Dividing health expenditures from period 1 by this index provides a measure of expenditures necessary in period 0 to derive the same units of utility as in the base period. Empirically, if we can obtain an accurate measure of the additional dollar value benefit of improved medical treatment, it can serve as the direct measure of quality.

Appendix 1 lays out the details of a model relating the different index methods to each other with the above notation. The indexes under consideration are:

1. A quality-adjusted index using the method of Cutler et al. 1998. Cutler et al. based their index on life expectancy outcomes so we call this the LE index.

2. A quality-constant index measuring the price of achieving a treatment endpoint based on Berndt et al. 2002, which we call the TE index. We show this index is equivalent to an index that prices a quality-adjusted life-year (QALY) such as in Lucarelli and Nicholson 2009, Howard et al. 2015, and Hult et al. 2017.

3. A constant-quality index similar to that in Frank et al. 2004, which holds technology constant by holding the market shares of different treatment baskets constant and then measuring the price change. We call this the “basket-price” (BP) index.
4. A quality-adjusted index with the quality adjustment based on the production cost change of the quality change. We call this the “resource cost” (RC) index.

We consider the performance of the indexes in a stylized model with two treatments of a conditions, one of which is both more expensive and more effective, and whose market share is increasing over time as it diffuses. The model gives the following results:

1. The LE index gives the correct result in all circumstances, including circumstances where the assumptions of the above setup are violated: there is no change in market shares of the treatments, the treatments are equally effective, the more expensive treatment is actually less effective, or the treatments differ in effectiveness but have the same price.

2. The validity of the BP and RC indexes, in particular, rests on the assumption that quality changes are reflected in changes in spending, so they are invalid if quality rises (falls) but spending falls (rises). However, as mentioned above, these scenarios could come to pass in health care, for example, if spending is lowered and quality increased simultaneously by reducing low-value and wasteful services.

3. In general, the three other indexes are equal to the LE index only when spending is equal to the monetized medical benefit, either in their levels or in their changes over time.

The approach of the TE index of pricing the achievement of a treatment endpoint or, equivalently, a QALY delivered by medical care has proven to be the most popular method in recent years (Berndt et al. 2002, Romley et al. 2015, Howard et al. 2015, Hult et al. 2017). As our model shows, however, the TE index and LE index are only equal when the price of achieving the endpoint in the second period is equal to the monetized value of achieving that endpoint. Sheiner and Malinovskaya (2016) explain the difference by arguing that because
consumers of medical care are at a corner solution prior to the introduction of a new technology, the marginal utility of buying another QALY can be greater than the marginal cost of the QALYs they have purchased so far. That is, they purchase all the QALYs that are possible under existing technologies and their budget constraint, but it is impossible to purchase more even though they would like to, given the price they have paid already. When new and effective technologies enter, the constraint is lifted but the new QALYs are more expensive than the previous one, although their price is still less than the marginal utility of buying one. An index that prices QALYs would therefore show an increase even though welfare has actually improved. The LE index shows correctly that the price has dropped (since it was effectively infinite before) because the quality adjustment based on the added QALYs is positive and therefore that patient welfare has improved.

With regard to the resource cost index, interestingly, recent empirical results by Doyle et al. (2015) find results for the cost of a quality difference that fall in the range of what we might expect. They find that the cost of producing an additional year of life to be approximately $80,000, which falls in the range of estimates of the value of a statistical life-year from the consumer’s perspective. However, we would discourage researchers from tackling this measurement problem from the producer’s perspective because, as discussed above, marginal utility and marginal cost have the potential to greatly diverge in health care. In addition, empirically, there are huge challenges in backing out this cost estimate from the producer’s perspective, as researchers must hold both technology and patient characteristics constant, while varying the level of resources used for treatment. Doyle et al. (2015) are able to do this through arguably pseudo-random assignment of ambulatory companies. More generally the assumption of random assignment is flawed as results from Chandra et al. (2016) and Romley and Goldman
(2011) show that patients shift toward facilities with higher quality. Finally, there are the problems with directly relating costs to quality in health care mentioned above, such as quality sometimes being improved by providers delivering less low-value care and therefore lowering costs.

There are other advantages of the LE index. As discussed by Hall (2016), the LE index is better at incorporating highly innovative medical treatments because it uses a universal metric (the QALYs added by medical treatment) for its quality adjustment. (This point is laid out formally in Appendix 1.) If there is an innovation such that the goal of treatment shifts entirely, there is no clear way to construct a comparable TE index over time. (An example is the introduction of antiretrovirals for HIV/AIDS treatment, at which point the goal of treatment shifted from treating AIDS-related infections to keeping the viral load low.) Similarly, the BP index measures the prices of the same treatments over time and it is unclear how to incorporate new treatments into it.

In summary, the LE index is the most flexible method in theory, correctly approximating the COLI in the widest set of circumstances, and we regard it as the benchmark to which all other indexes should be compared. In the next section, we will construct the LE, TE, and BP indexes for two acute illnesses and the LE and TE indexes for a third illness where outcomes have improved in recent years and show how the indexes diverge in value from each other.

4. Data and methods

Hall (2016) notes that many economics papers calculate quality adjustments based on observed health outcomes. When measuring aggregate outcomes, the challenge is to separate the effects of
medical care (which should be included in the quality adjustment) from the effects of environmental factors (which ought to be held constant) such as behavior, risk factors, and demographics. Due to this issue, many economics papers in this literature choose to measure quality based on observed short-term mortality outcomes of acute illnesses because mortality outcomes are more likely to be observed in data, because measuring them is relatively straightforward without medical expertise, and because measuring around an acute event allows for isolating the effects of medical care. We follow the same approach in this paper as we calculate quality-adjusted price indexes for three acute high-mortality inpatient illnesses among Medicare beneficiaries based on short-term mortality outcomes during or after the hospitalization.

Constructing the indexes described above with quality based on observed mortality outcomes and treatments requires data connecting treatments received, spending, and death dates for individual patients. We therefore use Medicare fee-for-service (FFS) claims as this is the only dataset where spending and details of treatments can be reliably connected to death dates of patients. Our sample consists of elderly FFS Medicare beneficiaries who had an inpatient admission between 2001 and 2014 for one of the following conditions: acute myocardial infarction (AMI), congestive heart failure (CHF), or pneumonia. The full details of how the sample was put together and how risk adjustment was performed are in the appendix. Beneficiaries were included if they had a full year of FFS enrollment prior to the index admission (for risk adjustment based on diagnoses) and a full year after the admission or death.

---

9 Medicare beneficiaries may choose to remain in fee-for-service or “traditional” Medicare which is operated by the Center for Medicare and Medicaid Services (CMS), or they may enroll in a Medicare Advantage plan operated by a private insurer contracting with Medicare. In the former case, their medical claims are held by CMS.
within the year after the admission, to measure outcomes. Enrollment and death dates are taken from the enrollment file.

Our study has several limitations due to well-known data constraints. First, as aforementioned, we needed to be able to tie diagnoses and treatments to mortality outcomes up to a year after the index event and the Medicare claims files are the only data source with this capability. However, as a result, our study is limited to creating price indexes for these conditions for elderly FFS Medicare beneficiaries.\textsuperscript{10} The price indexes are not representative of the U.S. population because we have no information on the commercially insured, Medicare Advantage enrollees, Medicaid enrollees, or the uninsured. Moreover, parallel to other papers in this literature, we only measure health outcomes with mortality and do not address quality of life. Finally, we lack spending and treatment data on outpatient pharmaceuticals for all beneficiaries in our sample.

5. Descriptive Statistics

The selected health conditions in the paper tend to afflict the oldest Medicare beneficiaries. Over 70 percent of the events in our sample are for individuals over the age of 75, even though half of the population in Medicare is between 65 and 75. As shown in Appendix Table A1, these beneficiaries have a high rate of comorbidities for which we make an adjustment described below. The last line of Table 1 gives the number of patients with each condition across years in the sample; there are from 8,000 to 30,000 observations in each year.

\textsuperscript{10} We have removed the disabled and end-stage renal disease (ESRD) population to create a more homogenous population to evaluate the impact of quality change.
As discussed above, the goal is a conditional COLI with the environment held constant. In this application that means adjusting measures of spending and outcomes for patient demographics and comorbidities to accurately capture the changes in health care technology and quality conditional on those factors. We therefore adjust for severity by applying standard regression techniques that control for the demographic and other health conditions of individuals and details of these methods are outlined in Appendix 2. We include those health factors listed in Appendix Table A1.

Table 1 summarizes the statistics for each condition. The table shows the percentage of males, age groups, and races for acute myocardial infarction, congestive heart failure, and pneumonia. The number of observations for each condition is also provided.

Table 2 shows 30-, 60-, and 90-day average adjusted expenditures (in 2014 dollars) in 2001, in 2007, and at the end of our sample in 2014. As it shows, the bulk of expenditures are incurred in the first 30 days following the event. For CHF and pneumonia, average adjusted spending per patient in the year following the event rose from 2001 to 2014 across all three windows.
Spending for AMI patients rose from 2001 to 2007 and has since declined to a level below its initial level in 2001, again across all three windows. The decline in growth in expenditures later in the period corresponds to a reduction in the growth rates of Medicare fees after 2010.

Table 2

<table>
<thead>
<tr>
<th>Days after hospitalization</th>
<th>30</th>
<th>60</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute myocardial infarction</td>
<td>2001</td>
<td>$24,693</td>
<td>$28,593</td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>$25,901</td>
<td>$30,159</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>$24,430</td>
<td>$28,322</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>2001</td>
<td>$14,613</td>
<td>$18,736</td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>$16,829</td>
<td>$21,561</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>$17,521</td>
<td>$22,685</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>2001</td>
<td>$14,351</td>
<td>$17,725</td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>$14,807</td>
<td>$18,570</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>$15,966</td>
<td>$19,986</td>
</tr>
</tbody>
</table>

Notes: Figures are deflated with the GDP deflator to 2014 levels.

Table 3 shows risk-adjusted survival rates to 30, 60, and 90 days and life expectancy following the event measured out to 365 days. For each condition, survival across all three windows and life expectancy improved from 2001 to 2014. Most of the improvements, however, took place between 2001 and 2007; there is relatively little improvement in the second half of the period. The increases in life expectancy and survival are larger for AMI and pneumonia than for CHF, which had longer life expectancy and survival rates to begin with.
Figures 1A and 1B show a clear and consistent pattern across the three conditions. The prices of treatments rise or stay nearly level over the period for the three conditions (Figure 1A), but the risk-adjusted 30-day mortality rates are falling (Figure 1B). Figure 1 highlights the goal of this paper: combining the price and quality information (i.e., the mortality estimates) into quality-adjusted price indexes. The next section discusses more specifically how we construct the theoretical price indexes described in the previous section.

Table 3
Risk-adjusted survival rates and life expectancy following hospitalizations

<table>
<thead>
<tr>
<th>Year</th>
<th>Survival rates</th>
<th>Life expectancy (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acute myocardial infarction</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>83.8% 79.5% 76.8%</td>
<td>70.7</td>
</tr>
<tr>
<td>2007</td>
<td>88.7% 85.5% 83.2%</td>
<td>80.6</td>
</tr>
<tr>
<td>2014</td>
<td>90.1% 87.0% 85.2%</td>
<td>83.5</td>
</tr>
<tr>
<td></td>
<td>Congestive heart failure</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>90.0% 84.6% 80.7%</td>
<td>111.8</td>
</tr>
<tr>
<td>2007</td>
<td>91.8% 87.1% 83.5%</td>
<td>119.7</td>
</tr>
<tr>
<td>2014</td>
<td>92.4% 87.1% 83.3%</td>
<td>121.5</td>
</tr>
<tr>
<td></td>
<td>Pneumonia</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>86.2% 80.8% 77.4%</td>
<td>93.2</td>
</tr>
<tr>
<td>2007</td>
<td>91.1% 86.6% 83.5%</td>
<td>107.7</td>
</tr>
<tr>
<td>2014</td>
<td>91.7% 87.0% 83.9%</td>
<td>109.6</td>
</tr>
</tbody>
</table>

Notes: Life expectancy is measured up to 365 days. Survival rates and life expectancy are risk-adjusted as described in Appendix 2.
Figure 1A. 30-Day Risk Adjusted Spending

Figure 1B. Risk-Adjusted Mortality Rate, 30-Day
6. **Empirical approach and results**

In this section we implement alternative methods for producing quality-adjusted price indexes and show how alternative assumptions influence the levels of the price indexes.

**LE index:** As discussed previously, the LE index approximates a COLI by adjusting the numerator of the price index by the change in value ($\Delta B$) received from medical care. Following our previous analysis, we construct the COLI for the entire population as:

$$LE \equiv \frac{S_1 - \Delta B}{S_0}.$$

The key challenge of the LE index is evaluating the monetary benefit of the quality change, $\Delta B$. The benefit of the change is from the increased life expectancy induced by improvements in treatment. However, if we measure this increase simply with the observed change in life expectancy, we run into two issues. The first issue is how to isolate the benefits for this condition, when the mortality rate for other conditions may be affecting our outcome variable. Cutler et al. (1998; 2001) addressed this by comparing the mortality rate of the treated population with that of the general population. One problem with this approach for the purposes of national accounting, however, is that it may be difficult to apply when looking at a broader set of conditions, as it would not be clear then how to define the general population.\(^{11}\)

A second problem is that price indexes need to be produced in a timely fashion so it is not possible to wait for the resolution of long-term outcomes. To both isolate the effect of treatment for the specific conditions on health and to create an index based on data available in the short-term, we take an approach different from that of Cutler et al. (1998).

\(^{11}\) In addition, there are likely to be improvements for other health conditions, leading to a reduction in relative benefits when looking at a control and comparison group.
We begin with the assumption that there is a point γ after the acute event before which survival of the event can be attributed to medical care for the event and after which it is determined by other factors such as lifestyle and medical care for other conditions. However, life expectancy (LE) will still overall be shorter after the event than it would be for similar patients who did not have the event. Life expectancy for patients who have the event is mechanically a weighted average of the LE of those who die before γ and that of those who die after. If we let \( m_t \) = the share of patients who die before γ, \( LE_{γ,t} \) = the LE of patients who die before γ in period t, and \( LE_{p,t| γ} \) = LE of survivors who die post-γ, then:

\[
LE_{MC,t} = m_t LE_{γ,t} + (1 - m_t) LE_{p,t| γ}
\]

The change in LE of these patients over time is given by:

\[
\Delta LE_1 = [m_1 LE_{γ,1} + (1 - m_1) LE_{p,1| γ}] - [m_0 LE_{γ,0} + (1 - m_0) LE_{p,0| γ}]
\]

\[
\Delta LE_2 = m_1 LE_{γ,1} - m_0 LE_{γ,0} + (1 - m_1) LE_{p,1| γ} - (1 - m_0) LE_{p,0| γ}
\]

\( m_t \) and \( LE_{γ,t} \) can be measured from the data in the short term. The question then is how to approximate \( LE_{p,t} \). The disadvantage of measuring this term directly in the data is that, as described above, it is affected by improvements in treatments of other conditions and measuring it fully requires waiting for the resolution of long-term outcomes. To solve both those problems, we hold \( LE_{p,t| γ} \) constant at its 2001 level. Then:

\[
\Delta LE_1 = m_1 LE_{γ,1} - m_0 LE_{γ,0} + (m_0 - m_1) LE_{p,0| γ}
\]

Because it is unclear at what point medical care for the event ceases to have an effect on post-event life expectancy, we create indexes with γ set at either 30, 60 or 90 days, which is the window in which we allow the benefits to change. After the 30-, 60- or 90-day window, we make the assumption that the health of the population that experienced the event is identical to the health of the population that survived the event in the initial period of the data. For example, if
the window is selected to be 30 days, and an individual in 2006 survives a heart attack for more
than 30 days, we assume that the number of years that the person survives after the 30-day
window is the same as someone that survived the 30-day window in 2001, where we observe
survival over a 13-year period. In other words, conditional on surviving through the initial
window (i.e., 30, 60 or 90 days), we hold life expectancy to be the same for the following 13
year period. This approach only allows for benefits to be realized if they occur in the window
around the event, so that changes in the treatments for other conditions are less likely to play a
role in the changes in outcomes. For example, if the window is 30 days and we see no change in
30-day life expectancy, then we would measure no change in quality. The shorter the window,
the lower the likelihood that other conditions will impact the outcome measure. However, a
shorter window may also miss some of the benefits if treatments influence long-term outcomes
after the window. For example, a new treatment may not affect 30-day life expectancy, but could
improve survival between 30 and 60 days.

Table 4 shows the results of these calculations for all three conditions and the 60-day window.
The last column, the synthetic life expectancy, is a weighted average of life expectancy before 60
days in each year and life expectancy conditional on surviving past 60 days in 2001, with the
weights being the 30-day mortality rate and its inverse. With the window set at 60 days, this
synthetic life expectancy following a hospitalization for an AMI increased nearly 144 days
between 2001 and 2014. The improvements in this synthetic life expectancy are almost entirely
driven by the improvement in the 60-day survival rate from 80 percent to 87 percent with a small
contribution from the 2-day increase in life expectancy of those who die in the first 60 days.

---

12 Specifically, for those that survived the acute event in 2001, we look at their mortality rate in the following 10 years. For all
subsequent years, we assume that individuals that survive the event have the same post-event survival rate as those that survived
in 2001. Ideally, we would compare mortality of the population that has an AMI with a comparable population that does not, but
this is not possible for very recent mortality estimates. Our analysis provides a practical alternative.
When the window is set at 30 days, life expectancy post-AMI increases less, around 115 days, and when it is set at 90 days, life expectancy increases more, around 170 days. These differences are driven by the fact that 90-day mortality improved more than 60-day mortality and 60-day mortality improved more than 30-day mortality. Short-term survival and synthetic life expectancy rose for all three conditions from 2001 to 2014; however, Table 4 shows that the bulk of the increases were between 2001 and 2007, with little improvements between 2007 and 2014.

<table>
<thead>
<tr>
<th>Acute myocardial infarction</th>
<th>Life expectancy (days) before 60 days</th>
<th>60-day survival rate</th>
<th>Life expectancy (days) conditional on surviving 60 days in 2001</th>
<th>Synthetic life expectancy (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>13.3</td>
<td>79.5%</td>
<td>1941.9</td>
<td>1547.1</td>
</tr>
<tr>
<td>2007</td>
<td>14.5</td>
<td>85.5%</td>
<td>1941.9</td>
<td>1662.0</td>
</tr>
<tr>
<td>2014</td>
<td>15.3</td>
<td>87.0%</td>
<td>1941.9</td>
<td>1691.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Congestive heart failure</th>
<th>Life expectancy (days) before 60 days</th>
<th>60-day survival rate</th>
<th>Life expectancy (days) conditional on surviving 60 days in 2001</th>
<th>Synthetic life expectancy (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>22.2</td>
<td>84.6%</td>
<td>1254.2</td>
<td>1064.6</td>
</tr>
<tr>
<td>2007</td>
<td>23.2</td>
<td>87.1%</td>
<td>1254.2</td>
<td>1094.9</td>
</tr>
<tr>
<td>2014</td>
<td>24.8</td>
<td>87.1%</td>
<td>1254.2</td>
<td>1095.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pneumonia</th>
<th>Life expectancy (days) before 60 days</th>
<th>60-day survival rate</th>
<th>Life expectancy (days) conditional on surviving 60 days in 2001</th>
<th>Synthetic life expectancy (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>19.3</td>
<td>80.8%</td>
<td>1418.7</td>
<td>1150.3</td>
</tr>
<tr>
<td>2007</td>
<td>22.0</td>
<td>86.6%</td>
<td>1418.7</td>
<td>1230.8</td>
</tr>
<tr>
<td>2014</td>
<td>23.1</td>
<td>87.0%</td>
<td>1418.7</td>
<td>1237.5</td>
</tr>
</tbody>
</table>

Notes: Life expectancies and survival rate are risk-adjusted as described in Appendix 2. Long-term life expectancy is measured through ten years.

Once we have measured the increased value of medical care in terms of life expectancy, we must also settle on a monetary value for a year of life to put $\Delta B$ in monetary terms. We follow Pandya et al. (2015) in using estimates based on three values for a year of life: $50,000$, $100,000$ and $150,000$ (in 2014 dollars). As they note, the $150,000$ amount has been justified as an upper
threshold by the World Health Organization (WHO) because it is approximately three times that of the GDP per capita (Neumann et al., 2014).\textsuperscript{13}

In calculating our LE indexes, we use a range of values for both the length of the mortality window over which we measure the short-term benefits of treatment and for the monetary value of a life-year. We allow the window to be 30, 60 or 90 days and we allow the value of a life year to be $50,000, $100,000, or $150,000. The estimates of unadjusted indexes and indexes adjusted for the value of change life expectancy are reported in Table 5. The unadjusted indexes report the growth of the average annual costs for each condition deflated with the GDP deflator.

<table>
<thead>
<tr>
<th>Window length</th>
<th>Annual value of life</th>
<th>30 days</th>
<th>60 days</th>
<th>90 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual value</td>
<td>$50,000</td>
<td>$100,000</td>
<td>$150,000</td>
<td>$50,000</td>
</tr>
<tr>
<td>Unadjusted index</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
</tr>
<tr>
<td>COLI</td>
<td>-4.8%</td>
<td>-9.8%</td>
<td>-15.1%</td>
<td>-5.1%</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>Unadjusted index</td>
<td>1.4%</td>
<td>1.5%</td>
<td>1.5%</td>
</tr>
<tr>
<td>COLI</td>
<td>-0.4%</td>
<td>-2.3%</td>
<td>-4.4%</td>
<td>-0.2%</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>Unadjusted index</td>
<td>0.8%</td>
<td>0.9%</td>
<td>1.0%</td>
</tr>
<tr>
<td>COLI</td>
<td>-4.4%</td>
<td>-9.9%</td>
<td>-15.8%</td>
<td>-4.3%</td>
</tr>
</tbody>
</table>

Notes: Estimates are computed as compound annual growth rates. The COLI estimates are computed by rebasing the amounts in each year. The price indexes are calculated with dollars deflated to 2014 values with the GDP deflator.

We make a few observations about the results in Table 5. First, quality adjustment turns out to be important across all of the assumptions. For each scenario we observe the quality adjustment having a significant impact, relative to the unadjusted index. The unadjusted indexes show

\textsuperscript{13} Government agencies often assign a value of a statistical life to conduct cost-benefit analysis. The Department of Transportation issues guidance on the value of a statistical life of $9.6 million in 2016 and the Environmental Protection Agency uses the value of $7.4 million in 2006 dollar values. However, these values would need to be transformed into a value of a statistical life year to be applicable in this study. Estimates of a value of a statistical life year reported in Aldy and Viscusi (2008) suggest that our values are relatively conservative as their value of a life year typically falls above $150k per year. However, no research we are aware of produces the value of a statistical life for the Medicare population age 65+.
annual price increases slightly above general inflation across conditions, while the growth rates of the quality-adjusted indexes are lower and often negative.

Table 5 shows that for pneumonia and heart attacks, which saw greater drops in mortality rates, quality adjustment has a larger impact than for CHF. This result highlights the necessity of disease-specific adjustment. Furthermore, for those conditions for which quality adjustment matters more, the estimates are much more sensitive to the variations in the value assigned to a life than to variations in the time period over which we are measuring benefits. Fixing the value of a life at $100,000, the table shows that the time period over which benefits are measured has a moderate impact on inflation for these conditions, with a difference of 1 to 2 percentage points. However, assigning the value of a year of life to be $50,000 compared to a value of $150,000 can change the inflation rate by a larger amount. For instance, holding days of measured benefit to be 60 days, we find the annual inflation rate is lower by between 2 and 6 percentage points more if the value of a life is placed at $50,000 compared to $150,000. For congestive heart failure, where the survival rate increases by the smallest amount, we see that the specific assumptions have a smaller impact on the quality-adjusted estimates.

Although our methods and assumptions are not identical, our results are a little stronger than those of Cutler et al. (1998). They find an annual quality-adjusted inflation rate for the treatment of AMI of around 1 to 2 percentage points below general inflation. They used relatively conservative estimates of the value of a year of life of $25,000 in 1991 dollars which is equivalent to $39,000 in 2014 dollars. Using our conservative value of a life year of $50,000 and allowing benefits to change up to a 30-day window we find that the average inflation rate across conditions, weighting by expenditure share across conditions, is 3.5 percentage points below
general inflation. Using a more central estimate with the $100,000 per life year estimate and 60-day window we arrive at an average inflation rate of 8.1 percentage points below general inflation. For AMI specifically, we find an annual inflation rate of -4.8 percent for our most conservative estimate, relative to general inflation.

**TE index:** We construct the treatment endpoint (TE) index in the same way as Berndt et al. (2002) construct their index but with the endpoints for the conditions as defined by Romley et al. (2015) who study the same acute inpatient conditions that we consider here. Berndt et al. (2002) construct their index by measuring the price of the expected achievement of an endpoint (remission of major depression). Their prospective measurement is based on the estimated medical value of the treatments being given. Romley et al. (2015) meanwhile measure hospital output retrospectively, defining successful output as meeting the endpoint of survival to 30 days without an unplanned readmission.

For each condition, we define the price in each period as the average annual incremental per patient cost of successfully achieving the treatment endpoint:

\[
\frac{S_1}{\sigma_1} / \frac{S_0}{\sigma_0}
\]

where \(S_t\) is average spending as defined above and \(\sigma_t\) is the percent of treatments that are successful relative to no treatment.\(^{15}\) We measure this by measuring the fraction of outcomes that are successful, relative to our assumption regarding untreated cases, and do not distinguish

---

\(^{14}\) To construct the weights, we multiply the number of observations for each condition by the 30-day spending estimate for each condition. The weights are 31 percent for heart attacks, 33 percent for heart failure, and 36 percent for pneumonia.

\(^{15}\) In the stylized model of Appendix 1, \(\sigma_t = q_t \pi_1 + (1 - q_t) \pi_2 - \pi_3\), where \(\pi_3\) is the success of the untreated cases and \(q_t \pi_1 + (1 - q_t) \pi_2\) is the success of the treated cases.
among treatments. Similar to Romley et al. (2015), we define “successful” treatment as surviving up to 30, 60 or 90 days without an unplanned readmission within 30, 60 or 90 days of discharge, with unplanned readmissions identified with the algorithm used by the Centers for Medicare & Medicaid Services (CMS). Survival rates with "successful" treatment are risk-adjusted as indicated above.

A challenge of constructing a TE index is that, because it measures the change in the incremental price relative to no treatment, it is necessary to know the rate of reaching the endpoint without any medical treatment. Berndt et al. (2002) were able to estimate the rate of remission of major depression without any treatment based on expert opinion because it was not uncommon for major depression to go untreated. However, for the conditions we are considering, every patient we observe receives treatment so it is difficult to know the success rate for untreated patients. At one extreme, the illnesses studied here are sufficiently severe that one may view non-treatment as a complete failure, so that the rate of success for untreated cases is arguably zero. However, there seems to be the potential for survival without treatment for all of these conditions. For example, prior to the 1960s when modern treatments were unavailable, the in-hospital mortality rate for AMI was 30 percent (Braunwald, 2012). Similarly, according to one cardiologist, in-hospital mortality from heart attacks in the 1970s for older patients was about 40 percent (National Heart, Lung, and Blood Institute, 2012). We view these estimates as an approximate baseline for “non-treatment.”

Table 6 shows some of the detail of the TE index calculations with the treatment/spending window held at 60 days and assuming a 20 percent survival rate for untreated cases. The quality-constant price of AMI treatment, for example, is $72,022 in 2001 and drops to $56,565 in 2014 as survival greatly improved but per-case spending declined slightly. Table 7 shows alternative
indexes based on different window lengths for measuring outcomes and spending and with differing assumptions for untreated cases. Again, adjusting for quality has a substantial impact on measured inflation and it has a larger impact on the indexes for AMI and pneumonia than for CHF. In addition, as we increase the assumed success rate of untreated cases, the incremental change in outcomes has a larger impact on the quality adjustment. If one assumes that the untreated success rate is large, so that medical care plays a smaller role in determining mortality, then the incremental changes in quality observed will show more progress in the advancement of medical treatments.16

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean total 60-day spending per patient</th>
<th>Rate of successful treatment (survival to 60 days without an unplanned readmission)</th>
<th>Assumed success rate of untreated cases</th>
<th>Price per incremental successful treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>$28,593</td>
<td>59.7%</td>
<td>20.0%</td>
<td>$72,022</td>
</tr>
<tr>
<td>2007</td>
<td>$30,159</td>
<td>66.6%</td>
<td>20.0%</td>
<td>$64,760</td>
</tr>
<tr>
<td>2014</td>
<td>$28,322</td>
<td>70.1%</td>
<td>20.0%</td>
<td>$56,565</td>
</tr>
</tbody>
</table>

Table 6

Prices per incremental successful outcome 2001-2014

<table>
<thead>
<tr>
<th>Year</th>
<th>Acute myocardial infarction</th>
<th>Congestive heart failure</th>
<th>Pneumonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>$18,736</td>
<td>61.4%</td>
<td>63.0%</td>
</tr>
<tr>
<td>2007</td>
<td>$21,561</td>
<td>64.7%</td>
<td>68.7%</td>
</tr>
<tr>
<td>2014</td>
<td>$22,685</td>
<td>66.2%</td>
<td>70.1%</td>
</tr>
</tbody>
</table>

Table 6

Prices per incremental successful outcome 2001-2014

<table>
<thead>
<tr>
<th>Year</th>
<th>Acute myocardial infarction</th>
<th>Congestive heart failure</th>
<th>Pneumonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>$18,736</td>
<td>61.4%</td>
<td>63.0%</td>
</tr>
<tr>
<td>2007</td>
<td>$21,561</td>
<td>64.7%</td>
<td>68.7%</td>
</tr>
<tr>
<td>2014</td>
<td>$22,685</td>
<td>66.2%</td>
<td>70.1%</td>
</tr>
</tbody>
</table>

Notes: Spending is deflated to 2014 dollars with the GDP deflator. Spending and survival rates are risk-adjusted as described in appendix 2.

16 For example, if the untreated successful rate of treatment is 40 percent, then a increase in the treated success rate going from 60 percent to 65 percent would be (65-40)/(60-40)=25/20=1.25 or a 25 percent increase in quality. However, if the untreated success rate is 0 percent, then the quality change is (65-0)/(60-0)=1.083 or just 8.3 percent. In our study, moving from a 0 percent success rate for untreated cases to a 40 percent success rate, leads to an annual inflation rate that is lower by about 1-3 percentage points.
BP index: The next method constructs a technology-constant index following Frank et al. (2004). This method holds the quality of care constant by holding the share of patients receiving specific treatments constant over time while allowing the prices of those treatments to change. Specifically, we run the following generalized linear model (GLM) regression, separately for each condition and year:

$$Y_i = \alpha + X_i \beta + Z_i \gamma + \varepsilon_i.$$  

where $Y_i$ is the annual health care spending related to the index admission of patient $i$, $X_i$ is a vector of patient-level covariates as indicated above, and $Z_i$ is a vector of evidence-based treatment types or therapies received within 30 days of the index admission. We then construct a Laspeyres-type index where the average price for each year is the average predicted spending with the prediction run on the data from 2001 using the $\hat{\beta}$ and $\hat{\gamma}$ from that year’s regression. We
construct a Paasche-type index using the same method on the 2014 data. The final index is a Fisher index, the geometric average of the two.

For both AMI and CHF, we are able to identify relevant technologies to include in \( Z_i \). Pneumonia treatment, however, mostly relies on antibiotics. Given the difficulty in using ICD-9 codes in the Medicare claims data to identify the many different antibiotic recommendations for treating pneumonia, we did not create BP indexes for the pneumonia cohort.

When we apply the BP method to AMI and CHF, we find that there is very little difference between the BP indexes and the unadjusted indexes. Given the limited change in these estimates relative to the unadjusted figures, we do not report these estimates separately but show them in the next section when we compare across methods (Figures 2-4).

The BP indexes also diverge substantially from the outcomes-based LE and TE indexes, both of which fall much faster than the unadjusted and BP indexes. This divergence suggests that the shift in the shares of the treatment baskets that we have defined are not actually related to the changes in observed outcomes captured in the two outcomes-based indexes. The improvements in mortality outcomes of AMI and CHF that we observe may have been caused by shifts among other treatments that we did not count separately, or by improvements in quality not captured in the claims data, such as improved coordination among hospital staff.

17 For the AMI cohorts, the treatments in \( Z_i \) are cardiac catheterization (CATH) only, percutaneous coronary intervention (PCI) only, coronary artery bypass grafting (CABG) only, and various combinations of CATH, PCI and CABG. The reference group is medical management which indicates the receipt of none of the heart attack procedure regimens. The medical management regimen is the least intensive, while CABG is the most intensive. The therapies for the CHF cohorts are the following: implantable cardioverter defibrillator (ICD) only, cardiac resynchronization therapy defibrillators (CRT-D) only, cardiac resynchronization therapy pacemaker (CRT-P) only and various treatment combinations of ICD, CRT-P, and CRT-D. We also include two infrequently used therapy options, which are present in the data: implantation of left ventricular assist device (LVAD), and heart transplantation. The reference group again is medical management, again indicating the receipt of none of the heart failure procedures identified above.
**RC Index:** The resource-cost approach is often preferred by the BLS when adjusting the PPI for quality and is used by the BLS to adjust the current hospital and nursing home PPIs. A challenge in the health care field is that it may be difficult to assign a cost to a quality change. In the hospital setting, many costs are also fixed and would need to be dispersed across a wide range of conditions (e.g., purchases of capital equipment like MRI machines). As mentioned previously, there may also be quality changes where there is no or little associated cost (e.g., receiving an aspirin within 24 hours of admission for an AMI) or even negative costs (e.g., reducing medical care that delivers no value). Since we have no data on the costs of producing treatments and their relationship to improvements in quality, we do not produce an RC index.

Sheiner and Malinovskaya (2016) show, however, that the RC index (which they call the “cost of quality improvement approach”) is equivalent to the TE index (which they call the “redefine the good approach”) under the assumption that costs are a purely linear function of successful treatments. We derive their result here in a slightly more extended form.

The average cost of attaining a successful health outcome in period $t$ is written as $\frac{S_t}{\sigma_t}$. We also write the change in the share of health outcomes that are successful between period 0 and period 1 as: $(\sigma_1 - \sigma_0)$. If cost is linear in successful treatments, then the estimated cost of the quality change can then be written as the average cost per outcome times the change in the outcome:

$$\frac{S_t(\sigma_1 - \sigma_0)}{\sigma_t}$$. The price index can then be written as:

$$RC = \frac{S_t}{S_0} \cdot \frac{S_t(\sigma_1 - \sigma_0)}{\sigma_t}$$.

Substituting $\frac{S_t}{\sigma_t}$ with period 1’s price per quality change, $\frac{S_1}{\sigma_1}$, the result then becomes:
Therefore, under the linear cost assumption, applying these assumptions, we see that the resource-cost method is identical to the treatment endpoint method.

As Sheiner and Malinovskaya (2016) note, however, this result only holds if there is zero profit. If we relax that assumption, to derive an actual cost of treatment, we would need to remove the profit margin $m$ of the provider and perhaps remove other costs that are unrelated to the treatment quality. In other words, it may be that only a percentage of the total expenditures, $\alpha = 1 - m$, are associated with the treatment of the condition. In this case, the cost per quality change may be scaled by some factor that is related to the cost specific to the treatment of the condition, which would tend to reduce the effect of the quality change. In this case, the resource-cost index becomes.

$$RC = \frac{S_1 - S_1(\sigma_1 - \sigma_0)}{\sigma_1} = \frac{S_1}{S_0}$$

If the profit margin is 10 percent, then $\alpha=0.90$ and the quality change applies to only 90 percent of the expenditures.

The assumptions that we would make to calculate the RC index lead to an index that is extremely close to the TE index, so we do not report calculations for this index separately. However, as the above equation shows, quality adjustment would tend to have a dampening effect on the RC index relative to the TE index.
Violation of the assumption that quality is related to cost has different implications for this approximation of the RC index from the theoretical RC index. Recall that one criticism of the RC index is that if quality changes without a corresponding change in the cost, there would be no change in the index. Because we are now assuming that changes in costs are directly related to the quality change and can therefore be inferred from a change in observed outcomes, this reformulation of the RC index assumes that if the quality changes by an amount \((\sigma_1 - \sigma_0)\), then there is necessarily a quality adjustment using the RC formula.

7. **Across-Method Comparison**

Next we graphically compare results from three of the methods choosing a single index from each approach. For the LE index we choose the estimate using $100,000 value of a year of life, which is the middle value of our range of assumptions. For the TE index, we assume a 20 percent success rate without treatment, also in the middle of our assumptions. The growth rates of the LE and TE indexes can be sensitive to the assumptions underlying the indexes but the differences across indexes tend to be so large that variations in these assumptions have a limited impact on their relative trends.

Figures 2 to 4 illustrate the differences between the indexes. All calculations were made with amounts measured in 2014 dollars so the growth rates pictured are relative to economy-wide inflation as measured by the GDP deflator. Across the three conditions we find similar patterns. We find that the unadjusted index and the BP index are nearly identical within conditions. We find that inflation is considerably lower when measured by outcomes in both the TE index and LE index, relative to the unadjusted index, but the amount of adjustment is much larger for the LE index.
Figure 2. Acute myocardial infarction

Figure 3. Congestive heart failure
As discussed previously, differences can arise between the LE index and other indexes when the increase in the benefits of treatment diverges greatly from the increase in spending. That seems to be the case for the conditions under study here. For AMI, for example, 60-day risk-adjusted spending per patient actually fell between 2001 and 2014 (Table 2) while monetized benefits (with short-term benefits measured through a 60-day window and with a year of life valued at $100,000) increased by $39,410 (Table 4). For CHF, the change in the benefits ($8,559) is only just over twice as high as the change in spending ($3,949), so the indexes look more similar to each other. For pneumonia, spending per pneumonia patient increased by $2,261 while the monetized value of the increase in life expectancy is $23,889, more than ten times as much.

8. Implications for Productivity

To demonstrate the potential economic importance of the quality adjustment, we examine the impact of applying our quality-adjusted index amounts to multifactor productivity estimates computed by the Bureau of Labor Statistics. The estimate of multifactor productivity from BLS
that is most relevant for our study is that for Hospitals and Nursing and Residential Care Facilities (NAICS 622, 623). The official estimates show a multifactor productivity growth rate that declines by 0.3 percent per year (see Table A3 of Appendix 2).

To adjust the multifactor productivity estimate for the quality change, we first determine the magnitude of the quality adjustment bias, which we use to adjust the output price index. We calculate the bias by taking the difference between a weighted average of the 60-day $50,000 per life-year LE indexes, where the weights are the total 60-day expenditure shares of each condition in 2001. The average of the unadjusted indexes grows at 0.8 percent per year while the average of the LE indexes falls at 3.2 percent per year so the bias is 4.0 percentage points. We restrict our adjustment only to the hospital sector by applying the quality adjustment to 80 percent of output since hospitals account for 80 percent of expenditures for NAICS industries 622 and 623. Therefore, the adjustment amount becomes 3.2 percent per year. We incorporate the quality adjustment by deflating the output price index by 3.2 percent per year over the period of study and then recompute a new quality-adjusted output and new productivity index. With this alternative estimate, we find that the new annual productivity growth rate becomes 2.8 percent per year.

To arrive at this adjusted estimate, many strong assumptions were applied. Most importantly, we assume that the magnitudes of quality adjustment that we estimate for our select conditions can apply to a wider set of medical conditions than those we consider here. In fact, we do not know the actual magnitude of the true quality adjustment, which could either be higher or lower than the estimates applied. The quality changes may also be different for privately insured or Medicaid patients, which are not included in our sample.
While the true growth rate of productivity is unknown, this hypothetical estimate highlights the importance of measuring price indexes and output correctly. This point may be important not only for the measurement of health care, but for measurement of the economy more broadly.

9. Conclusion

This paper reviews different methods for quality adjustment of medical price indexes. We contrast these methods theoretically and empirically for three acute conditions. In the model, we find that a COLI-type index whose quality adjustment is based on the monetized value of the increase in the health benefits of treatment such as that constructed by Cutler et al. (1998; 2001) gives the most theoretically accurate results and that the other indexes are closer to the COLI when the increase in medical spending is closer to the value of the increase in health benefits. If these values diverge, however, the other indexes are not equivalent. In particular, the somewhat popular approach of pricing outcomes or QALYs can give wrong results in the presence of technological constraints preventing the purchase of more QALYs (Sheiner and Malinovskaya 2016).

We conducted an empirical application by constructing these indexes for three acute conditions in the elderly Medicare population in the period from 2001 to 2014: acute myocardial infarction (AMI), congestive heart failure (CHF), and pneumonia. All of these conditions showed improvements in post-hospitalization life expectancy during this period, AMI and pneumonia more so than CHF. In this empirical application, we find that quality adjustment has a significant effect on the levels of the price indexes. Given that we also found that the monetized increases in
life expectancy exceeded the increases in spending by considerable margins, the quality-constant indexes diverged greatly from the benchmark COLIs for each condition.

Our work has the important limitation that we are only considering three conditions and the results should therefore not necessarily be generalized to all of healthcare. Howard et al. (2015) calculate a quality-adjusted price index for new cancer drugs from 1995 to 2013 and find that annual inflation in that period was 10 percent. Much of the other previous research on condition-level medical price indexes was generally conducted on conditions for which there was considerable technological advancement in the period being studied, which would bias towards finding lower inflation. Shapiro et al. (2001), for example, found that inflation measured at the medical condition level was considerably lower than inflation measured at the service level even without any quality adjustment when they measured it for one specific condition, cataracts. The health-care satellite account introduced by the BEA, however, showed that this result did not hold across conditions in a later time period because it found that inflation measured by condition was actually slightly higher than inflation measured by current methods (Dunn et al., 2015).

Our results show that accurate measurement of prices and output in the health-care sector could have substantial implications for the measurement of productivity. If health outcomes are improving due to medical care, inflation in health care is actually lower and the value of real output higher, implying higher rates of productivity growth in the sector if inputs are held constant. Pinning down this divergence for the entire health sector, as well as for the broader economy, is essential to fully understand productivity changes in the national economy.
Extending these methods to other conditions will be challenging, however. The approach used by Cutler et al. (1998) and by us only works well for acute conditions with high mortality. Tracking medical spending and outcomes immediately following an acute event allows for isolation of the effects of medical care from other factors such as behavior. Measuring the QALYs added by medical care for a chronic condition where behavioral risk factors play a strong role and where outcomes are measured by quality of life instead of mortality requires significant medical expertise. Going forward, however, there is considerable promise for further development of quality-adjusted price indexes for medical conditions as measurements of quality of life are improved, more detailed data on treatments and outcomes become available, and valuations of the quality and amount of life become more certain. The need for assessing the true value of changes in medical spending is only going to become more pressing in upcoming years.
References


Appendix 1

A simple model for comparing across methods: To compare these methods for creating quality-adjusted or quality-constant price indexes for medical care, and to deepen our understanding of how they relate to one another, consider the following simple model for a condition that has two treatments ($T_1$ and $T_2$) and has an endpoint that delivers $M$ QALYs. $M$ multiplied by the monetary value of one QALY is equal to $B$:

- $T_i$ has cost $C_{it}$ in period $t$ and patients receiving $T_i$ reach the endpoint with a probability of $\pi_i$.
- The proportion of patients in period $t$ receiving $T_i$ is $q_t$ so $1 - q_t$ receive $T_2$.
- If the condition receives no medical care, patients reach the endpoint with a probability of $\pi_3$. While $\pi_3$ in this model represents the case without medical care, in practice everyone receives medical care.
- $C_{it} > C_{2t}$ in each period $t$ and $\pi_1 > \pi_2 > \pi_3$. $T_i$ is both more expensive and more effective than $T_2$, and $T_2$ is more expensive and more effective than no medical treatment at all.
- $T_i$ is reimbursed to the provider at $R_{it} = C_{it} \cdot m_t$ where $m_t$ is the markup in period $t$.
- There are two periods, 0 and 1.

From this set-up it can be extrapolated that:

- The average spending on the condition in period $t$ is given as $S_t = q_t R_{1t} + (1 - q_t) R_{2t}$.
- The percent reaching the endpoint of treatment in period $t$ is equal to $q_t \pi_1 + (1 - q_t) \pi_2$.
- The incremental percent of total cases for which medical care is responsible for reaching the endpoint is equal to $q_t \pi_1 + (1 - q_t) \pi_2 - \pi_3$, i.e., the percent receiving the endpoint if no one received medical care subtracted from the percent reaching the endpoint in actuality.
- The change in the percent of patients reaching the treatment endpoint between period 0 and period 1 is written $\Delta q * (\pi_1 - \pi_2)$, where $\Delta q = q_1 - q_0$. 

50
The unadjusted index (UI) is written \( UI = \frac{s_1}{s_0} \).

We can then write down the associated formulas each of the four indexes, assuming data for all of the variables above are available.

**Life expectancy (LE) index:** An index adjusted for quality by making a direct quality adjustment based on the changes in the benefits of medical care is written \( \frac{s_1 - \Delta q^* \pi_1 - \pi_2 * B}{s_0} \). Because \( B \) is measured with life expectancy in Cutler et al. (1998), we will call this the LE index.

**Treatment endpoint (TE) index:** A constant-quality index that measures the relative change in price of meeting the treatment endpoint, such as that created by Berndt et al. (2002), will be written \( \frac{s_1}{s_0} \cdot \frac{q_1 \pi_1 + (1-q_1) \pi_2 - \pi_3}{q_0 \pi_1 + (1-q_0) \pi_2 - \pi_3} \). A variation on this approach is to price outcomes directly as in Lucarelli and Nicholson (2009) and Hult et al. (2017), both of which construct quality-adjusted prices by measuring the price of QALYs added by medical care. An index constructed with this price would be written \( \frac{s_1}{s_0} \cdot \frac{q_1 \pi_1 + (1-q_1) \pi_2 - \pi_3}{q_0 \pi_1 + (1-q_0) \pi_2 - \pi_3} \). However, it can be seen easily that \( M \) cancels out and that this index is therefore equivalent to the index that prices reaching the treatment endpoint.

**Basket price (BP) index:** A constant-technology index that measures the changes in the prices of treatment baskets and aggregates these prices holding the shares receiving the treatment or
technology constant using a Fisher index formula, such as that created by Frank et al. (2004), will be written
\[
\sqrt{\frac{q_0R_{11} + (1-q_0)R_{21}}{q_0R_{10} + (1-q_0)R_{20}}} \cdot \frac{q_1R_{11} + (1-q_1)R_{21}}{q_1R_{10} + (1-q_1)R_{20}} = \sqrt{UI} \cdot \frac{q_0R_{11} + (1-q_0)R_{21}}{q_1R_{10} + (1-q_1)R_{20}}.
\]

**Resource-cost (RC) index:** An index based on the change in costs originating from quality improvements will be constructed by applying that change to the unadjusted index. The total change in spending can be written:

\[
S_1 - S_0 = \Delta q \cdot (C_{11}m_1 - C_{21}m_1) + q_0 \cdot (C_{11}m_1 - C_{10}m_0) + (1 - q_0) \cdot (C_{21}m_1 - C_{20}m_0)
\]

The first term represents the change in spending coming from the change in quality and is therefore the quality adjustment to be put into the cost-based index, which we will call the RC index:

\[
\frac{S_1 - \Delta q \cdot (C_{11}m_1 - C_{21}m_1)}{S_0} = UI - \frac{\Delta q \cdot (C_{11}m_1 - C_{21}m_1)}{S_0}.
\]

When constructing this type of index based on production costs, the BLS includes the markup to costs in the adjustment so this index can then be written: \( UI - \frac{\Delta q \cdot (R_{11} - R_{21})}{S_0} \) (BLS, 2014). The last two terms measure the changes in the reimbursements of the same treatments over time and therefore capture pure inflation.

Next, we examine how the different indexes may deviate from each other and from a COLI estimate of a quality change. By construction, the LE index is a COLI which gives an accurate measure of price changes from the perspective of the consumer under all conditions and we regard it as the benchmark index. We then explore how the other indexes perform relative to the LE index under alternative scenarios where the assumptions of the model given above are violated.
1. If \( q_1 = q_0 \), there are no changes in treatment patterns and therefore no need for quality adjustment. In that case, all four indexes are appropriately equal to the unadjusted index.

2. If \( B = 0 \), that is, if achieving the treatment endpoint does not deliver any benefit at all in reality, the LE index will be appropriately equal to the unadjusted index but the other three indexes will not. The TE index, for example, will still measure the changes in the price of achieving the treatment endpoint whether or not achieving that endpoint has any meaning. It is essential therefore when constructing this type of index to choose a treatment endpoint that is medically meaningful.

3. If \( \pi_1 = \pi_2 \), that is, if both treatments are equally effective and there is therefore no actual change in quality, the LE index and the TE index are both appropriately equal to the unadjusted index. The BP and RC indexes, however, will differ from the unadjusted index. This reflects a weakness of these indexes, that whether or not they are meaningful depends on whether the shifts in \( q \) reflect actual improvements in treatment. However, it is questionable whether shifts to newer, more expensive treatments or increases in intensity of treatment always reflect actual differences in efficacy in health care.

4. If both treatments cost the same in both periods but \( q_1 \neq q_0 \), so there is quality change but no change in spending other than general inflation, the BP and RC indexes are inappropriately equal to the unadjusted index. These indexes assume quality changes are only reflected in changes in spending. However, as noted above, quality in health care can improve (decline) without increases (decreases) in spending.

In general, the other indexes approximate the LE index most closely when the value of the changes in quality lines up with the changes in spending, consistent with the result described in Triplett (1982).
If we set the LE and TE indexes equal, for example, and solve the value of the change in quality \( \Delta q \ast (\pi_1 - \pi_2) \ast B \), they are equal when \( B = \frac{S_1}{q_1\pi_1 + (1-q_1)\pi_2 - \pi_3} \), or in other words, when the monetized medical value of achieving the treatment endpoint is equal to the price of achieving that endpoint in period 1.

Similarly, if we set the LE and BP indexes equal, we find they are equal when \( \Delta q \ast (\pi_1 - \pi_2) \ast B = S_0 \ast (UI - BP) = S_0 \ast (\%\Delta spending - \%\Delta quality\text{-}constant\ spending) \). They are therefore equal when the monetized value of the change in outcomes is equal to the rise in spending that is due to quality change.

Finally, the LE and RC indexes are equal when \( (\pi_1 - \pi_2) \ast B = R_{11} - R_{21} \) or when the monetized value of the differences in outcomes between the two treatments is exactly equal to the difference in their prices in period 1.

**Incorporating innovative new treatments:** Let us hypothesize a medical innovation with a new treatment endpoint that delivers \( B_2 > B \) in monetized quality-adjusted life years (QALYs), that costs \( R_{31} \), and that 100% of patients receive in period 1, the first period it is available. The LE index can be calculated as \( \frac{S_1 - B_2}{S_0} \) because monetized QALYs are a universal metric that can be used to compare the values of all treatments. However, constructing the other three indexes require treatments to be comparable across periods. The TE and BP indexes cannot be calculated without identical endpoints or treatment baskets across periods. For the TE index, it is important to choose an outcome metric that is common across all technologies. The RC index is challenging to calculate as well because \( S_1 - S_0 = R_{31} - (q_0R_{10} + (1-q_0)R_{20}) \), so it may be difficult to split up spending into those components deriving from general inflation and those deriving from the quality change. The advantage of the BP and RC indexes, however, is that
they can be constructed without knowing B or observing outcomes. That information is often unknown to the economist constructing the index. They do, however, require creating treatment baskets which cannot be computed without medical expertise.
Appendix 2

Data sources

This study uses 2000-2015 Medicare claims data from the inpatient, outpatient, and carrier (physician) files. However, we perform the analysis only for the period 2001-2014. The 2000 data sets were used to identify a 365-day history of certain conditions for index admissions occurring in 2001 and the 2015 data sets were used to get the full 365-day spending and survival measures for index admissions occurring in 2014. We obtain patient demographic, enrollment and mortality information from the enrollment files.

Patient disease cohorts

In constructing the sample, we generally followed the method of Chandra et al. (2013). The analytical sample includes Medicare beneficiaries aged at least 65 years with an inpatient hospitalization and a primary discharge diagnosis for acute myocardial infarction (AMI), congestive heart failure (CHF), or pneumonia between 2001 and 2014. The index event was restricted to an inpatient setting in order to consider only acute cases of the condition. The International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes were used to identify the conditions. The heart attack cohort was identified using the diagnosis code 410.xx, excluding the fifth digit of 2 (that is, subsequent episode of care). The cohort of CHF patients was identified using the following diagnosis codes: 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.x, and 428.xx. For the pneumonia cohort, the following diagnosis codes were used: 481, 482.x, 482.xx, 483.x, 485, 486, and 487.x. The choice of these codes for each cohort was based on prior studies (Krumholz et al. 2006a, 2006b; Bratzler et al. 2011).

We restrict the samples to fee-for-service beneficiaries who were continuously enrolled for at least 365 days before the index admission and for at least 365 days (or until death) after the index admission. The requirement for enrollment for at least 365 days prior to the index admission is to ensure that we have a full 1-year history of certain conditions that we use as risk adjusters and the requirement for enrollment for at least 365 days after the index admission is to ensure that we are able to capture the full 1-year spending and survival measures after the index admission. We
require at least a 365-day window after an index admission of a particular patient before that patient can have another index admission. However, a patient can appear in a different disease cohort during the 365-day window of one cohort. A single beneficiary can therefore appear multiple times within a particular disease cohort or appear in different disease cohorts during the sample period.

**Outcome variables**

The outcome measures used are life expectancy (number of days survived after the index admission), survival rates up to a certain period and spending up to a certain period. As discussed in the paper, the periods over which health outcomes and spending are measured range from 30 days to 365 days. The spending variable encapsulates all medical care expenses incurred in an inpatient, outpatient or physician office settings during and after the index admission and is inflation-adjusted to 2014 dollars using the U.S. gross domestic product implicit price deflator.

**Risk adjusters**

To obtain risk-adjusted average survival days, survival rates and spending for each disease cohort, we estimated a generalized linear model (GLM) with a logit link function and assuming a negative binomial, binomial and gamma distributions for observed survival days, survival rates and spending, respectively. We adjusted for a number of patient-level covariates. In particular, we control for age groups (i.e., 5-year intervals with those aged at least 90 years as one group), sex and racial/ethnic groups (i.e., non-Hispanic Whites, non-Hispanic Blacks, non-Hispanic Asians, and Hispanics – the reference group is “Others”) in each cohort regression. Additionally, we control for certain hierarchical condition categories (HCC) that prior studies have found to be important risk-adjusters (Krumholz et al., 2006a, 2006b; Bratzler et al., 2011). The particular HCC variables were obtained using all diagnosis and procedure fields in the inpatient, outpatient, and physician claims data for the 365 days prior to the index admission and the secondary diagnosis and procedure fields in the index hospitalization. Specifically in each cohort regression, we control for the history (excluding the index hospitalization) of the following conditions: Percutaneous coronary intervention (PCI), Coronary artery bypass graft (CABG),

---

18 Hierarchical Condition Categories (HCC) is a grouping of the over 15, 000 ICD-9-CM codes into 189 clinically coherent groups
AMI, and Heart failure and for the following HCC groupings: Unstable angina, Chronic atherosclerosis, Cardiopulmonary-respiratory failure and shock, Valvular heart disease, Hypertension, Stroke, Renal failure, COPD, Pneumonia, Diabetes, Protein-calorie malnutrition, Dementia, Hemiplegia-paraplegia-paralysis-functional disability, Peripheral vascular disease, Metastatic cancer, Trauma in last year, Major psychiatric disorders, and Chronic liver disease. Additional cohort-specific covariates include two dummy variables for the AMI locations\(^\text{19}\) in the AMI cohort, Cerebrovascular diseases in the CHF and pneumonia cohorts and Severe hematological disorders, Iron deficiency and other/unspecified anemias and blood disease, Depression, Parkinson's and Huntington's diseases, Seizure disorders and convulsions, Fibrosis of lung and other chronic lung disorders, Asthma, and Vertebral fractures in the pneumonia cohorts.

\(^{19}\) The two dummy variables are for codes 410.1x and codes 410.2x, 410.3x, 410.4x, 410.5x, and 410.6x, respectively. The reference group is all others.
<table>
<thead>
<tr>
<th>Cardiovascular conditions:</th>
<th>Acute myocardial infarction</th>
<th>Congestive heart failure</th>
<th>Pneumonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>History of PCI</td>
<td>6.4%</td>
<td>7.9%</td>
<td>4.6%</td>
</tr>
<tr>
<td>History of CABG</td>
<td>8.8%</td>
<td>13.2%</td>
<td>7.0%</td>
</tr>
<tr>
<td>History of AMI</td>
<td>13.3%</td>
<td>10.8%</td>
<td>5.4%</td>
</tr>
<tr>
<td>History of heart failure</td>
<td>29.6%</td>
<td>66.9%</td>
<td>36.9%</td>
</tr>
<tr>
<td>Unstable angina</td>
<td>23.9%</td>
<td>16.6%</td>
<td>8.3%</td>
</tr>
<tr>
<td>Chronic atherosclerosis</td>
<td>23.9%</td>
<td>24.8%</td>
<td>15.1%</td>
</tr>
<tr>
<td>Cardiopulmonary-respiratory failure and shock</td>
<td>25.6%</td>
<td>30.6%</td>
<td>32.2%</td>
</tr>
<tr>
<td>Valvular heart disease</td>
<td>31.2%</td>
<td>44.7%</td>
<td>23.8%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other comorbidities:</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension</td>
<td>80.9%</td>
<td>84.0%</td>
<td>78.2%</td>
</tr>
<tr>
<td>Stroke</td>
<td>11.8%</td>
<td>12.1%</td>
<td>13.4%</td>
</tr>
<tr>
<td>Renal failure</td>
<td>30.4%</td>
<td>41.0%</td>
<td>27.2%</td>
</tr>
<tr>
<td>COPD</td>
<td>30.4%</td>
<td>43.4%</td>
<td>52.3%</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>5.2%</td>
<td>6.8%</td>
<td>10.3%</td>
</tr>
<tr>
<td>Diabetes</td>
<td>41.7%</td>
<td>47.7%</td>
<td>36.7%</td>
</tr>
<tr>
<td>Protein-calorie malnutrition</td>
<td>4.8%</td>
<td>7.3%</td>
<td>10.8%</td>
</tr>
<tr>
<td>Dementia</td>
<td>14.3%</td>
<td>17.1%</td>
<td>25.3%</td>
</tr>
<tr>
<td>Hemiplegia, paraplegia, paralysis, functional disability</td>
<td>6.2%</td>
<td>7.1%</td>
<td>8.5%</td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
<td>30.6%</td>
<td>36.2%</td>
<td>31.2%</td>
</tr>
<tr>
<td>Metastatic cancer</td>
<td>3.8%</td>
<td>4.3%</td>
<td>8.4%</td>
</tr>
<tr>
<td>Trauma in last year</td>
<td>4.9%</td>
<td>6.6%</td>
<td>7.6%</td>
</tr>
<tr>
<td>Major psychiatric disorders</td>
<td>4.4%</td>
<td>5.8%</td>
<td>8.1%</td>
</tr>
<tr>
<td>Chronic liver disease</td>
<td>1.1%</td>
<td>2.0%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td>6.2%</td>
<td>7.1%</td>
<td>8.5%</td>
</tr>
<tr>
<td>Severe hematological disorders</td>
<td>3.8%</td>
<td>4.3%</td>
<td>8.4%</td>
</tr>
<tr>
<td>Iron deficiency and other blood disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parkinson's and Huntington's diseases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seizure disorders and convulsions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fibrosis of lung and other chronic lung disorders</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertebral fractures</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Number of observations for each condition           | 173,277                     | 314,560                   | 340,675   |
Table A2: The number of events for each condition in each year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Acute myocardial infarction</th>
<th>Congestive heart failure</th>
<th>Pneumonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>15,839</td>
<td>24,596</td>
<td>27,184</td>
</tr>
<tr>
<td>2002</td>
<td>16,224</td>
<td>25,030</td>
<td>29,097</td>
</tr>
<tr>
<td>2003</td>
<td>15,942</td>
<td>26,683</td>
<td>30,393</td>
</tr>
<tr>
<td>2004</td>
<td>14,953</td>
<td>26,653</td>
<td>27,955</td>
</tr>
<tr>
<td>2005</td>
<td>13,703</td>
<td>25,744</td>
<td>30,230</td>
</tr>
<tr>
<td>2006</td>
<td>12,753</td>
<td>24,945</td>
<td>26,557</td>
</tr>
<tr>
<td>2007</td>
<td>12,066</td>
<td>23,023</td>
<td>24,299</td>
</tr>
<tr>
<td>2008</td>
<td>11,719</td>
<td>21,956</td>
<td>24,276</td>
</tr>
<tr>
<td>2009</td>
<td>10,699</td>
<td>21,569</td>
<td>21,766</td>
</tr>
<tr>
<td>2010</td>
<td>10,830</td>
<td>21,012</td>
<td>21,312</td>
</tr>
<tr>
<td>2011</td>
<td>10,099</td>
<td>19,799</td>
<td>21,462</td>
</tr>
<tr>
<td>2012</td>
<td>10,164</td>
<td>18,862</td>
<td>20,297</td>
</tr>
<tr>
<td>2013</td>
<td>9,539</td>
<td>18,113</td>
<td>19,643</td>
</tr>
<tr>
<td>2014</td>
<td>8,747</td>
<td>16,575</td>
<td>16,204</td>
</tr>
<tr>
<td>Total</td>
<td>173,277</td>
<td>314,560</td>
<td>340,675</td>
</tr>
</tbody>
</table>
Table A3. Hypothetical Adjustment to BLS Multifactor Productivity Estimate for Hospitals and Nursing and Residential Care Facilities (NAICS 622, 623)

<table>
<thead>
<tr>
<th>Year</th>
<th>BLS (current)</th>
<th>Alternative Adjusted Productivity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Real output</td>
<td>Price indexes</td>
</tr>
<tr>
<td>2001</td>
<td>74.66</td>
<td>79.87</td>
</tr>
<tr>
<td>2002</td>
<td>79.44</td>
<td>81.90</td>
</tr>
<tr>
<td>2003</td>
<td>81.96</td>
<td>84.32</td>
</tr>
<tr>
<td>2004</td>
<td>84.00</td>
<td>87.16</td>
</tr>
<tr>
<td>2005</td>
<td>88.77</td>
<td>90.01</td>
</tr>
<tr>
<td>2006</td>
<td>91.27</td>
<td>92.93</td>
</tr>
<tr>
<td>2007</td>
<td>93.82</td>
<td>95.66</td>
</tr>
<tr>
<td>2008</td>
<td>95.87</td>
<td>98.77</td>
</tr>
<tr>
<td>2009</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>2010</td>
<td>103.41</td>
<td>101.94</td>
</tr>
<tr>
<td>2011</td>
<td>106.24</td>
<td>105.00</td>
</tr>
<tr>
<td>2012</td>
<td>111.13</td>
<td>106.07</td>
</tr>
<tr>
<td>2013</td>
<td>114.05</td>
<td>108.06</td>
</tr>
<tr>
<td>2014</td>
<td>117.17</td>
<td>110.06</td>
</tr>
<tr>
<td>2015</td>
<td>124.61</td>
<td>110.96</td>
</tr>
</tbody>
</table>

Notes: The BLS estimates of multifactor productivity taken from the table of productivity for the nonmanufacturing industries (https://www.bls.gov/mfp/mprdload.htm). The adjustment to the BLS estimates is based on the difference in the weighted unadjusted price index, which grows at 1.1 percent per year, and the LE quality-adjusted index that grows at -1.9 percent per year, assuming a 60 day window and a value of a life of $50,000 per year.