Are Medical Care Prices Still Declining? A Systematic Examination of Quality-Adjusted Price Index Alternatives for Medical Care[1]

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)

May 2019

Abstract: More than two decades ago a well-known study provided evidence from heart attack treatments suggesting that prices in medical care were actually declining, when appropriately adjusted for quality. Our paper revisits this subject looking at a large number of conditions and more recent and more comprehensive data sources to compare alternative methods of quality adjustment. A method based on utility theory produces the most robust and accurate results, while the alternative methods used in recent work overstate inflation. Based on claims data for three medical conditions as well as data on medical innovations from over 7,000 cost-effectiveness studies spanning all major condition categories and types of treatment, we find that, when properly adjusted for quality, declining prices from innovation are a prevalent feature of this sector. These findings have important implications for the measurement of medical care output and productivity.

[1] We would like to thank Ana Aizcorbe, Ben Bridgman, David Byrne, David Cutler, Berthold Herrendorf, Charles Jones, Dennis Fixler, Bronwyn Hall, Bob Hall, Peter Klenow, Brett Matsumoto, John Romley, Louise Sheiner, Jon Samuels, Jon Skinner, and the experts on the Bureau of Economic Analysis Academic Advisory panel for comments. We would like to thank conference participants at the NBER CRIW conference, seminar participants at the Society for Economic Measurement in Boston, and participants at the Productivity Growth: Past, Present, and Future conference at Arizona State University and seminar participants at the Federal Trade Commision. We would also like to thank Lasanthi Fernando for research assistance. The views expressed in this paper are those of the authors and not necessarily the views of the Bureau of Economic Analysis, the U.S. Department of the Treasury or the World Bank.

1. Introduction

In a well-known paper in the health economics literature, Cutler et al. (1998) asked the question: "Are medical prices declining?" Measuring the prices of treatments for heart attacks, they find that after accounting for quality improvement, the price of treatment declined over their period of study, even while the unadjusted price of treatment rose. These findings suggest that the mismeasurement of medical care output and productivity could be substantial. It has major implications for individual welfare and economy-wide real output given that medical 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 the gross domestic product (GDP) devoted to medical care rose from 13 percent in 1998 to nearly 18 percent in 2017 (Martin et al. 2018). Health experts believe that much of the growth in this sector is driven by new technologies that improve treatment in the long run (Chernew and Newhouse 2011) and recent work has shown that new innovations have been a key factor behind the rapid growth in expenditures over this period for many conditions, including rheumatoid arthritis, cancer, hepatitis, and HIV (Dunn et al. 2018). Meanwhile, life expectancy at birth in the United States has increased by nearly two years over the same period with medical innovations likely playing a role (Anderson 2001;Kochanek et al. 2017). While innovations are a key contributing factor to the growth in spending for medical care, changes in the quality of medical care are not reflected in U.S. national statistics, leading official statistics to overstate inflation in this sector (Lebow and Rudd 2003; Groshen et al. 2017).

Some evidence of quality change may be gleaned from declining national mortality rates and individuals living more disability-free years (Cutler, Rosen, and Vijan 2006; Cutler et al. 2017). However, given that non-medical factors may influence health outcomes, it can be challenging to

accurately attribute changes in the health of the population to changes in the medical care sector. Price measurement in medical care is further complicated by the rapid pace of technological change, third-party payers, and information asymmetries among other factors (Hall 2016; Sheiner and Malinovskaya 2016). Even for conditions that are more amenable to quality adjustment, there is no consensus on the best method for creating quality-adjusted price indexes. The papers in this literature use different methods of quality adjustment with no discussion of the connections among them (Hall 2016; Sheiner and Malinovskaya 2016).

The goals of this paper are to establish a framework relating the different methods, to illustrate the differences between them empirically, and to demonstrate that price declines are found in a different time period and over a wider set of conditions than previously studied. We show that both the theoretical and empirical differences across methods have substantial implications for measurement in this sector.

The methods we compare are: (1) a utility-based cost-of-living index (COLI) following Cutler et al. (1998); (2) an index measuring the price per unit of health produced from treatment; (3) a hedonic index.; and (4) an index based on the cost of producing the change in quality. The utility-based COLI method of Cutler et al. (1998) is our benchmark method because it is grounded in utility theory and we also show that it is robust to market distortions common in health care markets. Following utility theory, our benchmark method assigns a value to the quality change based on the dollar value of the marginal quality change to a consumer. The question to be addressed is whether the other methods produce similar results or have other distinguishing properties. We show that the second method is only consistent with our benchmark when there is a linear relationship between the health produced and dollars spent on treatment, so that quality changes are valued at the average price per unit of health produced. Stated another way, the second

method values quality changes as if individuals could purchase years of healthy life at a constant average price; when, in fact, quality should be valued at its marginal value, which we argue is significantly higher than the average price. Consequently, the second method tends to undervalue changes in quality. For researchers interested in using a market price to value quality improvements, researchers should target the marginal price of health improvement, and not the average price.

The other two methods also deliver higher price growth than our benchmark utility-based method because they both use changes in spending and costs as proxies for the value of changes in improvements in health. As we will show, however, typically the high valuation put on health and longer life imply that the increases in spending are well below the patient valuation of improved medical technology.

After reviewing the theoretical differences, we apply alternative methods to estimate price indexes to two distinct data sources. We first use claims data to study 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, treatments for these conditions also showed significant improvements in health outcomes as measured by post-hospitalization life expectancy. We find strong evidence that quality adjustment is important, with quality-adjusted indexes growing less than the unadjusted indexes even under the most conservative assumptions, highlighting the importance of quality adjustment.

As expected from our theoretical model, we find that our preferred utility-based COLI price index tends to fall much faster than other quality-adjustment methods because it is the only method that accounts for the full value of improvements in health. Overall the results of the utility-based method show that the average price across the three conditions is declining by 7.4 percent per year relative to an economy-wide deflator and based on the value of a statistical life year of $100,000. We find that the magnitude of the decline is highly dependent on the assumption about the value of extended life, but we estimate that the average price still falls annually by 3.1 percent, even when making the conservative assumption that the value of a statistical life year is worth $50,000.

Next, to determine if these price declines are representative of this sector more generally, we examine evidence from a database of over 7,000 clinical studies from the Tufts Medical Cost Effectiveness Analysis Registry (CEAR) database.[2] The database includes information on thousands of medical innovations including their health benefits and treatment costs, as well as the benefits and costs of prior treatment technologies. We first show that using a measure of the price per unit of health produced, as applied in Hult, Jaffe, and Philipson (2018), can significantly understate the value from new treatments and lead to improper conclusions about the importance of innovations in the sector. To fully capture the value of these innovations, we show that it necessary to apply a utility-based formula to the innovations reported in the CEAR database. Using our preferred index and conservative assumptions, we find price declines from innovations averaging 20 percent or more, relative to the prior standard of care. These declines are similar in magnitude to those observed in high-tech areas of the economy and provide suggestive evidence

[2] Center for the Evaluation of Value and Risk in Health.

that the price declines observed for our three select conditions may be a prevalent feature of the health sector.

The main contribution of this paper is that it shows that the method of constructiong quality adjusted price indexes matters theoretically and empirically. Applying a consistent methodology of utility-based quality adjustment across a wide range of studies (e.g., studies that differ on a variety of dimensions such as how they measure cost and quality and applying widely different data sources) produces surprisingly consistent results of quality adjusted prices declining. These estimates have important implications for the measurement of output and productivity growth. The Bureau of Labor Statistics (BLS) estimates multifactor productivity growth for the hospital and nursing home sector to be negative over the 2001-2014 period, with an annual decline of 0.3 percent. Under the strong assumption that our conservative utility-based measure of quality adjustment for our three conditions studied with the Medicare data is representative of the hospital sector more broadly, we apply the adjustment to the output price index. We find that it implies a multifactor productivity growth rate of 2.8 percent, holding inputs constant.

2. Background on price indexes in health care

Currently, the BLS measures the prices of individual medical services (e.g., price of a doctor's visit) and the Bureau of Economic Analysis (BEA) uses those indexes to deflate expenditures and measure real output for the health care sector.[3] However, there is general agreement among experts that the price in the medical care sector should track the full medical expenditures to treat an episode of a condition, rather than the price of an individual service (National Research Council 2010; World Health Organization 2011). With a treatment-based framework, analysts can better

[3] The BLS has moved towards condition-based measurement for the PPI for general medical and surgical hospitals, but this includes only hospital spending.

measure changes in practice patterns, technologies, outcomes, and associated expenditures on treatments relevant to a condition (National Research Council 2010). The BEA and BLS have already developed experimental treatment-based indexes (Bradley, Hunjan, and Rozental 2015; Dunn, Rittmueller, and Whitmire 2015), although these indexes currently do not control for quality.[4] Our paper focuses on quality adjustment for treatment-based indexes, as does most research on quality-adjusted medical care price indexes (Cutler et al. 1998; Shapiro, Shapiro, and Wilcox 2001; Berndt et al. 2002; Frank et al. 2004).

The measurement of medical care prices lies at the center of an important economic question about the forces driving medical expenditures higher. One explanation for the rise in health spending is a scenario suggested by Baumol (1967), where more expenditures are shifted toward labor-intensive sectors, such as health care, where official measures show low productivity growth. On the other hand, health care has seen significant technological change which has improved health and mortality outcomes over the past 60 years, as discussed in Cutler, Rosen, and Vijan (2006). If official measures of inflation are not capturing this quality improvement, the reverse scenario could be the case, that resources are shifting to health care in response to quality-adjusted prices for health care falling.[5] More accurate measures of price for the health care sector may challenge basic assumptions about price growth in this sector and have implications for understanding across-sector shifts and economy-wide growth.[6]

[4] The experimental indexes of the two agencies are related, although the BLS index contains a downward bias. See Roehrig (2017) for a comparison.

[5] Chandra et al. (2016) show that consumers prefer higher performing hospitals and shift toward higher quality hospitals over time. A related economic puzzle is the current slowdown in measured productivity growth in the U.S. (1.6 percentage point lower growth in labor productivity since 2004), which has received considerable attention (Byrne, Fernald, and Reinsdorf 2016; Syverson, 2017). The full role of the medical care sector contributing to this slowdown is currently unknown given the substantial measurement challenges in this area.

[6] More generally, accurately measuring the price of health care may be important for understanding growth and sectoral shifts in economies, such as the recent decline in manufacturing and growth in the service sector (Ngai and Pissarides (2007); Herrendorf, Rogerson, and Valentinyi (2011); Duernecker, Herrendorf, and Velentinyi 2018).

This work relates to a broader literature on the measurement of quality changes, the value of new goods, and quality-adjusted price indexes, such as the work by Feenstra (1994), Bresnahan and Gordon (1996), Bils and Klenow (2001), Petrin (2002), Broda and Weinstein (2010), Diewert and Feenstra (2018), Redding and Weinstein (2018), and Aghion et al. (2019). Our paper differs from these as it focuses on the medical care sector and applies a method adapted to the unique features of this sector. However, our paper shares the common feature with this literature that our preferred index is grounded in economic theory. Also similar to many of these papers, we find that the magnitude of the quality-adjustment is substantial and has important macroeconomic implications.

Theory

The guiding principles behind price measurement in the health care sector should have theoretical foundations shared by the rest of the economy. For this reason, we view the utility-based true COLI as the ideal foundation for a quality-adjusted index, following Fisher and Shell (1972) and in accordance with the guidelines laid out in "At What Price?" (National Research Council 2002).[7] A utility-based COLI is written as:

$$COLI = \frac{e(p_1,U_0)}{e(p_0,U_0)} = \frac{e(p_0,U_0)-(e(p_0,U_0)-e(p_1,U_0))}{e(p_0,U_0)} \qquad (1)$$

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. The utility-based COLI is the change in expenditures necessary to maintain the same level of utility across periods, given the observed change in prices.

This recent literature is interested in understanding why resources shift to low productivity service sectors, where health care is typically defined as "low productivity" based on our official measures of inflation.

[7] More formally, they recommend a conditional COLI, which is "conditional" in the sense that it ignores factors that are outside of a pre-defined scope, such as public goods or the weather. The COLI approach to quality adjustment is directly applicable to health care, as it is a component of final consumer spending.

The utility-based COLI may also be written as a measure of the change in welfare, as the term $e(p_0, U_0) - e(p_1, U_0)$ is a measure of compensating variation. A full COLI would account for all medical conditions and treatments, as well as all goods and services in the economy. However, following Cutler et al. (1998) we abstract from a more general COLI by estimating a price index specific to one medical condition.

A representative consumer's utility at time t is, $U(H(\alpha_t \cdot m_t), x_t)$ where m_t is the medical care input, $H(\alpha_t \cdot m_t)$ is the medical care technology function that translates medical care into health, and x_t is a numeraire good with a price normalized to 1. The term α_t captures the productivity of medical care in producing health. The consumer has an income Y and is subject to a budget constraint $p_t m_t + x_t \leq Y$.

To form our benchmark index, we start by defining the compensating variation (CV) in the following relationship:

$$U(H(\alpha_1 \cdot m_1), Y - p_1 m_1 + CV) = U(H(\alpha_0 \cdot m_0), Y - p_0 m_0) \qquad (2)$$

The CV is the additional dollars necessary to make consumers indifferent between the first and second period treatments. Taking a first-order Taylor-series approximation at period 0 yields:

$$CV = \frac{U_H H_m (\alpha_1 \cdot m_1 - \alpha_0 \cdot m_0)}{U_x} - (p_1 m_1 - p_0 m_0) \qquad (3)$$

where U_H is the marginal utility of health, H_m is the marginal effect of medical care on health, and U_x is the marginal utility of non-health consumption ($x_t = Y - p_t m_t$). The term α_t makes clear that technological change can lead to a higher CV, without any change in medical care inputs, m_t. Cutler et al. (1998) note that the first term in equation (3) is the monetary benefit from improvements in medical care treatment. The change in benefit is measured as the improvement

in health due to medical care, $H_m(\alpha_1 \cdot m_1 - \alpha_0 \cdot m_0)$, times the monetary benefit of improvements in health, $\frac{U_H}{U_x}$. The second term is the change in spending to treat the condition ($\Delta S = S_1 - S_0 = p_1 m_1 - p_0 m_0$) where $S_t = p_t m_t$.

We are interested in a price index specific to treating a medical care condition assuming all other prices and income do not change. We obtain this index by subtracting base period numeraire expenditures from the numerator and the denominator, so that the disease-specific index captures the change in medical expenditures necessary to maintain the same level of utility across periods:[8]

$$\frac{e(p_0,U_0)-x_0-(e(p_0,U_0)-e(p_1,U_0))}{e(p_0,U_0)-x_0} = \frac{S_0-CV}{S_0}.$$

Using the Taylor-series approximation for CV:

$$\frac{S_0-CV}{S_0} = \frac{S_0-\left(\frac{U_H H_m(\alpha_1 \cdot m_1-\alpha_0 \cdot m_0)}{U_x} - (p_1 m_1-p_0 m_0)\right)}{S_0} = \frac{S_1-\left(\frac{U_H H_m(\alpha_1 \cdot m_1-\alpha_0 \cdot m_0)}{U_x}\right)}{S_0}. \quad (4)$$

Equation (4) gives the formula for our target COLI index. Intuitively it can be understood as adjusting the numerator of the unadjusted price index $\frac{S_1}{S_0}$ with the term $\frac{U_H H_m(\alpha_1 \cdot m_1-\alpha_0 \cdot m_0)}{U_x}$ which is the marginal monetary valuation of health $\frac{U_H}{U_x}$ times the change in health, $H_m(\alpha_1 \cdot m_1 - \alpha_0 \cdot m_0)$, giving the total benefit to the patient of improvements in health in monetary terms. An advantage of this index is that the benefit to the consumer are derived from health changes

[8] See Sheiner and Malinovskaya (2016) for a more complete discussion of a full COLI and how it relates to a disease-specific COLI. They importantly note some limitations of the COLI specified in Cutler et al. (1998) and propose an alternative. The disease-specific index in our paper builds on the insight from Sheiner and Malinovskaya (2016). To see how this relates to the aggregate index, suppose the share of expenditures on health is $\frac{e(p_0,U_0)-x_0}{e(p_0,U_0)}$ and the share on the numeraire good is $\frac{x_0}{e(p_0,U_0)}$ with a price index of 1, then individual indexes relate to the aggregate by multiplying by the spending share for each category: $\frac{e(p_0,U_0)-x_0-CV}{e(p_0,U_0)-x_0}\left(\frac{e(p_0,U_0)-x_0}{e(p_0,U_0)}\right) + \frac{x_0}{e(p_0,U_0)} = \frac{e(p_0,U_0)-CV}{e(p_0,U_0)}$.

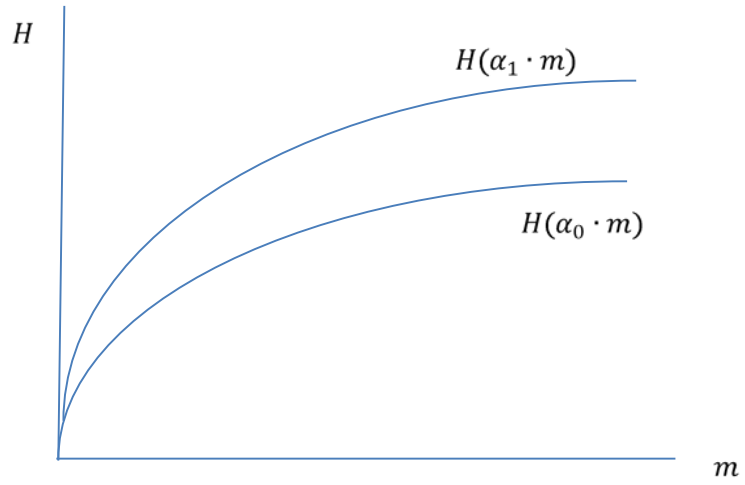
(observed in practice or in clinical trials) and estimates of the marginal value of health (e.g., based on the value of a statistical life year), so that the index is not reliant on equilibrium market conditions and is robust to potential market distortions.

When we apply this index to the data, we will refer to this target COLI as the life-expectancy (LE) index because the change in health, $H_m(\alpha_1 \cdot m_1 - \alpha_0 \cdot m_0)$, is often measured by changes in life expectancy or changes in quality-adjusted life years.[9]

Health Producing Technology. The change in quality depends on the health production function, $H(\alpha_t \cdot m_t)$. While the health production technology is not necessary to form the LE index, it is helpful for understanding key differences in the quality-adjusted price indexes used in the literature. The health production function is unknown, but following Skinner and Staiger (2015), we assume that medical care technology is limited, so that additional medical care inputs have diminishing returns. For example, one can think of physicians applying the lowest price, highest impact medical treatments first (e.g., an aspirin after a heart attack is low cost and highly beneficial), but the last treatment applied may have a smaller impact on health per dollar spent (e.g., bypass surgery). In other words, we expect $H(m)$ to be concave as in Figure 1. Additional health may be delivered for the same level of medical care if technology improves, as reflected in the figure below by an increase in α_t.

[9] Ideally, life expectancy would account for not just the quantity of life, but also the quality of life through considering morbidity factors.

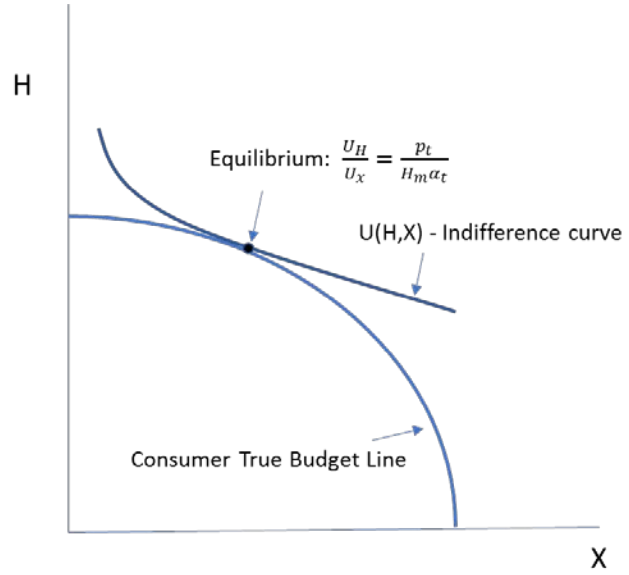
Figure 1. Health Production Technology



A concave health production function implies that the price per marginal unit of health $\frac{p_t}{H_m \alpha_t}$ will be increasing in health. Therefore, if we express the consumer's budget constraint as a trade-off between health, H, and x, the numeraire good, the budget constraint will be curved, as shown in Figure 2. We follow Hall and Jones (2007) who argue that the marginal utility of health stays relatively constant with increasing health, unlike the marginal utility of other goods that decline with additional units of consumption.[10] We therefore model the utility from health with a constant marginal valuation of health for all but the smallest values of x.

[10] They specifically start with a utility function of the form $U(c) = B + \frac{c^{1-\gamma}}{1-\gamma}$, where B is a constant and the indifference curve for consumption quickly becomes flat. Consumers receive this level of utility for each year of life, so the full utility function is roughly, $U(c, H) = H(B + \frac{c^{1-\gamma}}{1-\gamma})$, leading to relatively flat indifference curves. Other functional forms for health and consumption give similar properties of declining marginal utility from consumption relative to health, which results in an indifference curve that is relatively constant at the equilibrium point.

Figure 2. Consumer's Utility Maximization Problem



In a hypothetical scenario where the consumer does not have health insurance, the consumer consumes up to the point where the marginal value of an additional unit of health, $\frac{U_H}{U_x}$, is equal to the marginal cost of an additional unit of health $\frac{p_t}{H_m \alpha_t}$:

$$\frac{U_H}{U_x} = \frac{p_t}{H_m \alpha_t}. \qquad (5)$$

The equilibrium condition from equation (5) is depicted in Figure 2 and implies that the consumer increases health consumption up to the point where the marginal value of health (the left-hand side) is equal to the marginal price of another unit of health (the right-hand side). Although equilibrium conditions are not necessary for our preferred index, alternative quality-adjusted methods use market prices as part of the quality-adjustment. Using the above hypothetical model can help us evaluate these indexes in a "best case" scenario where the prices reveal information about the value of treatment.

Alternative Price Indexes

Life-Expectancy index. The utility-based LE index $\frac{S_1-\left(\frac{U_H H_m(\alpha_1 \cdot m_1-\alpha_0 \cdot m_0)}{U_x}\right)}{S_0}$ values gains in health at the marginal monetary value of the gains in health, $\frac{U_H}{U_x}$. The valuation is consistent with the consumer's utility maximization problem (5), but also values health gains correctly, even when the first-order conditions (5) do not hold.[11] This is because the estimated value of health to the consumer, $\frac{U_H}{U_x}$, is computed directly from external information regarding the consumer's value of additional health and does not depend on the price of medical care. In this way, it is robust to market distortions thought to affect the health care sector. We will contrast the LE index with three alternatives and show why, in most situations, the other indexes understate welfare gains from improvements in medical technology.

Treatment-Endpoint (TE) index. This index measures the price per unit of health produced from a treatment. If we let σ_t represent some measure of the health produced at time *t*, so that $\sigma_t \approx H_t$, then the index may be written as:

$$TE = \frac{S_1/\sigma_1}{S_0/\sigma_0} \quad (6)$$

Often the value σ_t is measured as the rate of obtaining a successful treatment endpoint (e.g., survival after 30 days or remission from a health condition), so we refer to the index as the treatment-endpoint (TE) index. Berndt et al. (2002) took this approach in measuring the

[11] It also does not rely on knowing the shape of the health care production function.

incremental price of achieving remission of major depression with medical care relative to no treatment. More recently, Romley, Goldman, and Sood (2015) took a similar approach when they measured the output of hospitals by measuring the number of successful treatments, where they defined a successful treatment as survival through a certain time-period without an unplanned readmission.

One of the more popular implementations of this approach in recent years has been to measure the price per quality-adjusted life-years (QALYs) added by medical treatment (Lucarelli and Nicholson 2009; Howard et al. 2015; Hult, Jaffe, and Philipson 2018). The QALY is just another measure of the health produced from treatment, H_t, where a one value unit of a QALY represents one year of life in perfect health, typically accounting for morbidity and mortality factors. This type of index is effectively identical to the index formed based on a successful treatment endpoint. For example, if σ_t is the rate of achieving a successful treatment endpoint that adds M QALYs, an index that prices QALYs= $M \cdot \sigma_t$ would be written:

$$\frac{S_1/_{M\sigma_1}}{S_0/_{M\sigma_0}} \quad (7)$$

From this it can be seen that M cancels out and we are left with the TE index.

The TE index, although intuitive and popular, has the potential to substantially overstate quality-adjusted inflation in health care and understate welfare improvements from new medical technologies. If we assume that the rate of achieving a treatment endpoint, σ_t, is a reasonable proxy for health (as it is intended), $\sigma \approx H$, then the index may be written as: $TE = \frac{S_1/_{H(1)}}{S_0/_{H(0)}}$ where $H(\alpha_t \cdot m_t)$ is written as $H(t)$ to simplify notation. Following Sheiner and Malinovskaya (2016),

we rewrite the TE index as: $\frac{S_1 - \frac{S_1}{H(1)}(H(1)-H(0))}{S_0}$. The functional form is nearly identical to the LE index with the unadjusted price of treatment in the numerator subtracted by an adjustment term that accounts for the observed change in health, divided by the unadjusted price of treatment in the base period. The primary difference in the indexes is the value placed on improvements in health from treatment. The value of improvements in health in the TE index is proportional to the average price per unit of health produced, which may be seen by rewriting the adjustment term, $\frac{S_1}{H(1)} = \frac{p_t}{H(1)/m_1}$.

This adjustment term could be justified if consumers pay a constant dollar price for each additional unit of health. However, in the previous section we argued that the production of health has diminishing returns and therefore each additional unit of health is costlier than the last (Figure 1). The previous section showed that, under normal assumptions, the benefits of medical care should be measured with its marginal benefit to consumers. In the hypothetical scenario of equation (5), the marginal price per unit of health rises as a patient is treated until the last unit of medical care where the marginal price per unit of health $\frac{p_t}{H_m \alpha_t}$ is equal to the marginal benefit $\frac{U_H}{U_x}$. Therefore, the marginal price of the last unit of health purchased would be larger than the average price per unit of health from treatment, $\frac{p_t}{H_m \alpha_t} > \frac{p_t}{H(1)/m_1}$. Consequently, an adjustment based on the average price per unit of health will understate the value of the quality change.[12, 13]

[12] Even from the perspective of a producer maximizing revenue, quality should be valued at the marginal revenue received for producing a marginal improvement in health, not the average revenue per unit of health.

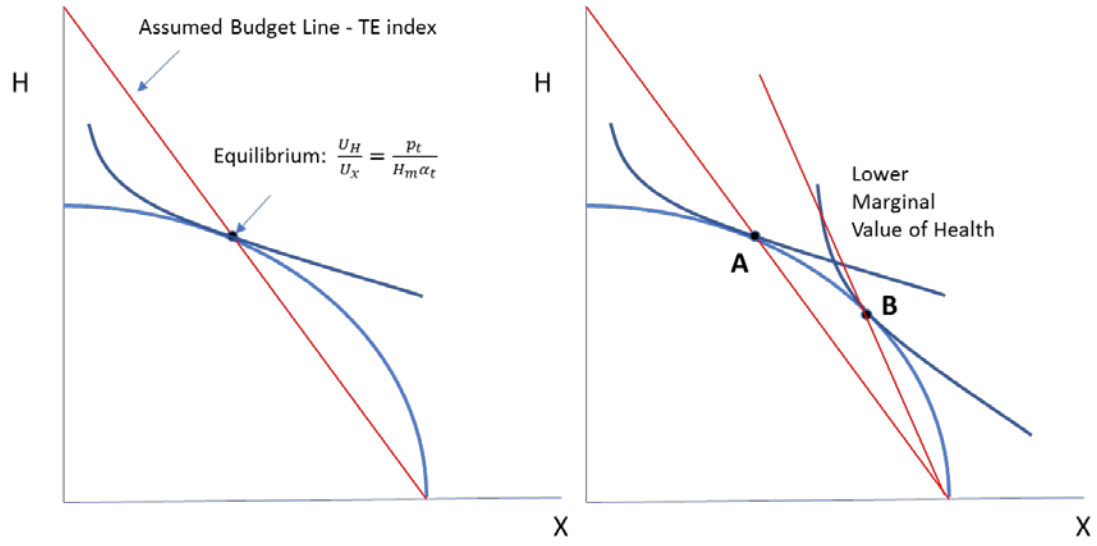
[13] Sheiner and Malinovskaya (2016) find a similar result but their model assumes linear costs and then uses a technological constraint resulting in a corner solution to explain why a difference would arise between the LE and TE indexes. Here we show it is not necessary to have a corner solution for there to be a difference between the LE

This result may be seen graphically in the left panel of Figure 3 which depicts the budget constraint implied by the TE index overlaying the curved budget constraint of Figure 2. The TE index assumes a linear cost for producing additional health, which intersects at the same point as the curved budget constraint on the X-axis and at the equilibrium point. The TE index is adjusted with the average price of H, which is the slope of the straight-line TE budget constraint. The true marginal rate of substitution for health at the equilibrium point, however, will always be higher than the value assigned by the TE index. This leads the TE index to undervalue changes in quality. This argument is shown more formally in the appendix.

Furthermore, as the right panel of Figure 3 shows, the discrepancy between the TE index and the LE index will be greater if the marginal valuation of health is higher. At point A, the utility curve is drawn so that the marginal valuation of H is relatively higher and the difference between the slope of the tangency and the slope of the straight-line budget constraint is greater, while at point B, the marginal valuation is lower and the difference between the slopes is smaller.

and TE indexes, but this difference should be expected more generally when there are diminishing returns to health inputs.

Figure 3. Consumer's Utility Maximization Problem and the TE Quality Adjustment Assumption



Overall, while the TE index is intuitively appealing, we would caution against its use if we believe that the marginal valuation of health is relatively high. For researchers interested in using a "market price" to value quality, the above discussion shows that the theoretically more relevant price is the marginal price, $\frac{p_t}{H_m \alpha_t}$, and not the average price.[14] While it is sometimes argued that an attractive feature of the TE index is that one does not have to place a value on a statistical life year, the formula $TE = \frac{S_1 - \frac{S_1}{H(1)}(H(1) - H(0))}{S_0}$ shows that this method unavoidably places a value on health that may have no economic foundation.

Hedonic index. The next method for constructing quality-adjusted medical price indexes is a hedonic index that uses a hedonic regression to control for the characteristics of treatment or new

[14] The market price for a marginal increase in health may be challenging to estimate because of selection issues. Moreover, if the goal of quality adjustment is a price index relevant for final consumption, this method will produce inaccurate estimates unless the FOC holds. The closest measure we are aware of in the literature to $\frac{p_t}{H_m \alpha_t}$ is Doyle et al. (2015). Exploiting exogenous variation in ambulance assignment, they find the implied cost of producing an additional year of life to be at least $80,000.

innovations following the work of Frank et al. (2004) who applied this method to study schizophrenia treatment.

In general, Pakes (2003) shows hedonic indexes provide an upper bound to a utility-based COLI index using arguments similar to Konüs (1939). The argument is applicable to the health care setting under the strong assumption of utility maximization and no market distortions. Let the hedonic function for period t be $g_t(m_t)$, which is an estimate of the price of purchasing medical technologies m_t in period t, so that $g_t(m_t) = S_t$. The function $g_t()$ captures relevant technologies and characteristics of medical care inputs m_t, in period t. The hedonic adjustment in period 1 is $g_0(m_0) - g_1(m_0)$, which is the dollar value in spending in period 0 minus the cost of purchasing the period 0 treatment in period 1. This difference is a lower bound for the CV because in period 1 individuals prefer treatment m_1, even though treatment m_0 is still available, so any change in the cost of purchasing m_0 is less than the full compensating variation adjustment: $\big(g_0(m_0) - g_1(m_0)\big) < CV$. The hedonic price index is then, $\frac{S_0 - (g_0(m_0) - g_1(m_0))}{S_0} = \frac{g_1(m_0)}{S_0}$, which is an upper bound for the price change implied by a full CV adjustment (i.e., $\frac{S_0 - (g_0(m_0) - g_1(m_0))}{S_0} > \frac{S_0 - \mathrm{CV})}{S_0}$).

There are three important considerations relevant for the application to health care. First, the hedonic index provides an upper bound, but Pakes (2003) warns that it may be far from the least upper bound that is desirable as it will not account for the full utility change, especially for innovative markets.[15] Second, determining what treatment characteristics to include is both

[15] Specifically, when new goods enter a market, consumers purchase up until the marginal utility of the marginal consumer equals the marginal price. However, there may be large inframarginal gains from technological improvements, from consumers that are considerably better off because of the introduction of the new goods.

important and challenging, as it may require significant understanding of treatment technologies.[16] Third, the assumptions needed for a hedonic adjustment may be violated because it is possible that inefficient technologies that do not add to societal welfare may be adopted. As a simple example, if an individual has an indemnity insurance plan that covers 90 percent of expenditures, she would have an incentive to seek treatment costing $1,000, even if the health benefit is worth only $500, because the out-of-pocket cost (i.e., $100) is less than the benefit. Empirically, Brot-Goldberg et al. (2017) show that consumers are not necessarily optimizing, as they 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. These examples undermine the rationale for the hedonic adjustment, as society may be worse off with treatments selected in period 1, implying the hedonic index may overstate the gains in welfare.

In summary, it is challenging to control for the right product characteristics in this framework. If consumers and doctors are not making optimal decisions for society, then the hedonic adjustment may be far from the correct adjustment and could either overstate or understate changes in welfare. Even under ideal conditions, when the right characteristics are controlled for and consumers and doctors are making optimal decisions for society, the basket price index provides an upper bound to an index that accounts for the full CV adjustment.

Resource-cost index. Quality-adjusted price indexes may be formed from the perspective of a producer using inputs to produce treatments. This producer problem is the theoretical basis of a

[16] For example, in health care it is common to receive nearly identical treatments at lower costs (i.e., highly elastic treatment alternatives). Consumers switching toward lower-cost close substitutes such as from branded to generic drugs (Griliches and Cockburn 1994; Feenstra 1997) or shifts from inpatient to outpatient treatments (Aizcorbe and Nestoriak 2011) should theoretically be counted as a reduction in price. However, controlling for the characteristics of treatment, such as "generic" or "inpatient" erases these price changes that are theoretically appropriate and economically important for obtaining a tighter bound on CV.

resource-cost index (Fisher and Shell 1972). The arguments for the producer are parallel to those presented from the consumer's perspective. Suppose the revenue function of a representative producer in the economy is $R(H(\alpha_t \cdot z_t^m), x_t)$ where z_t^m is an intermediate input devoted to medical care and x_t is the numeraire intermediate input.[17] As before, the function, $H(\alpha_t \cdot z_t^m)$ is the health production function and α_t captures health technology changes, but in this case z_t^m are the inputs of the producer. The resource constraint of the economy is $w_t \cdot z_t^m + x_t \leq M$ where w_t is the price of the medical care input and the price of the numeraire input has been normalized to 1.[18] The producer pays for the inputs at a price equal to its costs. We can then form parallel arguments to those presented for the consumer.[19] The ideal producer price index based on this framework is:

$$\frac{S_1 - \frac{R_H H_z \alpha_0}{R_x}(\alpha_1 z_1 - z_0 \alpha_0)}{S_0}. \quad (8)$$

[17] To keep the model simple, the revenue function produces has two inputs, medical care producing health and the numeraire good. For the producer, one could potentially extend the model to allow them to gain revenue based on the number of patients treated (i.e., $(\alpha_t \cdot z_t^m) \cdot \#of\ treatments$), rather than just the health produced per treatment. However, an additional treatment at a fixed quality would be quite similar to a change in a typical good in the economy. That is, we would assume a constant-returns to scale to the number of treatments, which poses no fundamental measurement challenges. We focus only on the health produced, which poses the measurement challenge.

[18] Both inputs are produced one for one with labor $z_t^m = l_t^m$ and $x_t = l_t^x$ where there is a fixed amount of labor.

[19] The dollar value in producer inputs, W, that holds revenues constant over the two periods is: $R(H(\alpha_1 \cdot z_1^m), M - w_1 \cdot z_1^m - W) = R(H(\alpha_0 \cdot z_0^m), M - w_0 \cdot z_0^m)$. Taking a first order Taylor series approximation at time 0, the value of the producer inputs necessary to hold revenues constant is: $W = \frac{R_H H_z \alpha_0}{R_x}(\alpha_1 z_1 - z_0 \alpha_0) - (w_1 z_1 - w_0 z_0)$. The first term measures the dollar value of a change in quality from a change in the input $(\alpha_1 z_1 - z_0 \alpha_0)$, where the dollar value is measured as the opportunity cost of output, $\frac{R_H H_z \alpha_0}{R_x}$, from producing additional health relative to the output that could be generated by the numeraire input, x_t. Assuming a competitive equilibrium, this will be equal to the marginal value of the quality change for the consumer. If we also assume that the output is competitively produced, so that the producer receives the marginal product of its output, then an alternative interpretation is a representative consumer utility model as in Aghion et al. (2019).

In this framework, the quality adjustment term is based on the opportunity cost (measured in marginal revenue) of devoting additional resources to improving health, rather than producing additional units of the numeraire good.

The ideal producer price index in (8) is distinct from how the resource-cost index is applied in practice. First, the productivity improvement reflected in the technology change $(\alpha_1 - \alpha_0)$, may be of great importance, as large improvements in quality may involve a shift in technology, such as in Figure 1, where more output may be produced for the same level of inputs. In practice, this productivity term is ignored.[20] Second, it may be challenging to derive the opportunity cost of inputs, $\frac{R_H H_Z \alpha_0}{R_X}$. However, if one assumes that the first order conditions of the producer holds, then $\frac{R_H H_Z \alpha_0}{R_X} = w_0$ and the index becomes:

$$\frac{S_1 - w_0(z_1 - z_0)}{S_0} \quad (9)$$

The quality-adjustment term is then, $w_0(z_1 - z_0)$, which is the cost of producing the change in quality, which is the resource-cost index that is typically applied in practice. A complication, for equation (9), is that capturing the actual resource-cost of the innovation may be tremendously complex for medical care (e.g., purchasing a new MRI machine improves diagnosis across many conditions). Moreover, this correction requires strong assumptions regarding the cost of inputs and their relationship to quality as quality cannot improve without costs going up. A simple

[20] For example, in regards to quality adjustment for new vehicles, the BLS writes: "Occasionally, new technology makes it possible to achieve recognizably better quality at no increase in cost—or possibly even at lower cost. While the values associated with these changes provide BLS with reference information, they are not reflected in BLS quality adjustment amounts."

counterexample is taking an aspirin after a heart attack event, which may have a large effect on the outcome, but costs almost nothing.

In general, the resource-cost approach using equation (9) ignores productivity changes, assumes that firms are producing efficiently in a competitive environment, and that they are receiving their marginal product for the quality that they produce. All of these assumptions may be problematic in health care. Similar to the other methods, this index will also be close to the LE index in cases where changes in benefits are similar to the change in cost but would tend to diverge in other cases. Due to the practical challenges of applying this method, we do not apply this approach in our empirical analysis.

Returning briefly to the ideal producer price index (8), one may be interested in estimating the ideal producer index directly, as this provides an economically meaningful quality-adjustment. Empirical work by Grieco and McDevitt (2016) provides insight to this topic. Specifically, they measure the production function of dialysis centers that consider two dimensions of output, quality and quantity. They find a quality-quantity trade-off and measure the opportunity cost of the production of an additional unit of quality, roughly providing a measure of $\frac{R_H H_Z}{R_X}$.[21] They find that the opportunity cost of reducing one infection is $75,000 (i.e., the opportunity cost in revenue lost from the reduction in quantity to produce more quality). In contrast, the societal benefit of reducing one infection (using a conservative value of a statistical life of $50,000 and 1.8 life-years saved) is $90,000, plus the additional hospitalization costs averted of $25,000, for a total of $115,000. In the case of dialysis treatment, the value of quality from an ideal resource-cost perspective ($75,000) is below the value of quality from a utility-based perspective ($115,000). In this

[21] They analyze productivity for Medicare enrollees where prices are fixed.

example, applying a quality adjustment using a value of $75,000 per infection prevented would be a valid adjustment from the producer's perspective, but since we are concerned with measuring output for final consumption, this value understates our preferred measure.[22]

Summary of Methods

The benchmark LE index gives the correct result across many scenarios, including circumstances where standard optimization assumptions are violated.[23] For the TE index, quality is adjusted based on the average price of producing health, which tends to understate the full marginal benefit (and marginal price) of quality improvements. For researchers interested in using a "market price" to value quality improvements, the marginal price of producing an additional unit of health is the theoretically more appropriate price measure. The validity of the hedonic and resource cost indexes rest 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, 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.[24]

4. Data and methods

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

[22] For researchers interested in creating a price index from the producer's perspective, the methods of Grieco and McDevitt (2016) provide an alternative, which has not yet been applied in the literature. Similar to the LE index, the methods used in Grieco and McDevitt (2016) provide economic foundations for assigning value to quality, where value is assigned based on the measured opportunity cost of producing the quality change.

[23] When a new higher cost technology is introduced, the correct adjustment is derived across several scenarios: when there is no observed change in technologies; when treatments are equally effective; or when the more expensive treatment is actually less effective. The correct adjustment is also derived when the treatments differ in effectiveness but have the same price.

[24] A stylized model in the appendix presents a comparable analysis for the simple case of two discrete treatments of a condition, which highlights many of the same points made here.

hospitalization. Following others in the literature, we use Medicare fee-for-service (FFS) claims where spending and details of treatments can be reliably connected to death dates of patients.[25] 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 three selected conditions account for a large number of inpatient hospital stays, ranking among the 10 most frequent conditions for inpatient admission for those over the age of 65 according to estimates from the Healthcare Cost and Utilization project (HCUP), Nationwide Inpatient Sample (NIS) 2010.[26] According to this data, over 65 percent of the stays for these conditions are for individuals over the age 65 captured in our Medicare data. This share understates the economic importance of these conditions for this population as the the severity of the illnesses typically increases with age as reflected in longer lengths of inpatient stays and higher mortality rates for those over the age of 65 (HCUP 2016).

Beneficiaries were included if they had a full year of FFS enrollment prior to the index admission (to use comorbidities prior to the event to use in risk adjustment) and a full year after the admission or death within the year after the admission, to measure outcomes. Enrollment and death dates are taken from the enrollment file. The full details of how the sample was put together and how risk adjustment was performed are in the Appendix.

When measuring medical care quality, 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. Our analysis

[25] 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.
[26] The estimates are based on statistics available from the HCUP webstie: https://www.hcup-us.ahrq.gov/.

of the claims data follows many economics papers in this literature that choose to measure quality based on observed short-term mortality outcomes of acute illnesses because mortality outcomes are important health measures observed in the data, measuring them is relatively straightforward without medical expertise, and measuring around an acute event allows for isolating the effects of medical care (Hall 2016). To study conditions more generally, including non-acute conditions, clinical-trial data that randomizes patients across treatments may be necessary. Later in this paper we use a database of studies from the medical literature as the basis for the empirical analysis, which covers a broader range of medical treatments and conditions.

The analysis of these claims data has several limitations due to well-known data constraints. First, our study is limited to creating price indexes for these conditions for elderly FFS Medicare beneficiaries.[27] While the Medicare FFS population likely accounts for a majority of the population afflicted with the conditions studied in this paper, the price indexes may not be representative of the U.S. population because we have no information on non-Medicare FFS beneficiaries. 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 the beneficiaries in our sample.

5. Descriptive Statistics

Table 1 provides some descriptive statistics for individuals with one of the three select conditions. These conditions 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 the ages of 65 and 75. Table 1 also shows that these beneficiaries have a high rate of

[27] 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.

comorbidities, with around 80 percent of patients having hypertension and over 36 percent with diabetes. The last line of Table 1 gives the number of patients observed with over 150,000 observations for each condition.[28] Additional demographic and condition information is provided in A1 and A2 of the appendix.

Table 1
Summary statistics

	Acute myocardial infarction	Congestive heart failure	Pneumonia
Male	44.2%	37.6%	39.0%
Age group:			
Age: 65-69	12.0%	8.3%	9.1%
Age: 70-74	17.2%	13.0%	14.2%
Age: 75-79	19.7%	17.7%	18.3%
Age: 80-84	20.5%	21.9%	21.4%
Age: 85-89	17.4%	21.1%	19.7%
Age: >=90	13.2%	17.9%	17.3%
Cardiovascular conditions:			
History of PCI	6.4%	7.9%	4.6%
History of CABG	8.8%	13.2%	7.0%
History of AMI	13.3%	10.8%	5.4%
History of heart failure	29.6%	66.9%	36.9%
Unstable angina	23.9%	16.6%	8.3%
Chronic atherosclerosis	23.9%	24.8%	15.1%
Cardiopulmonary-respiratory failure and shock	25.6%	30.6%	32.2%
Valvular heart disease	31.2%	44.7%	23.8%
Other comorbidities:			
Hypertension	80.9%	84.0%	78.2%
Stroke	11.8%	12.1%	13.4%
Renal failure	30.4%	41.0%	27.2%
COPD	30.4%	43.4%	52.3%
Pneumonia	5.2%	6.8%	10.3%
Diabetes	41.7%	47.7%	36.7%
Number of observations for each condition	173,277	314,560	340,675

As discussed above, the goal is a conditional utility-based 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

[28] We observe around 8,000 to 30,000 observations per year for each condition.

quality conditional on those factors. We therefore adjust for severity by applying standard regression techniques that control for the demographic and health conditions of individuals. Details of these methods are outlined in Appendix 2. We include those health factors listed in Table 1 and additional factors listed in Appendix Table A1. The estimates of quality and spending measures are only as good as the risk adjustment applied to the data. Recent work by Doyle, Graves, and Gruber (2014) tests the validity of standard risk adjustment techniques by exploiting quasi-random assignment of patients to hospitals using ambulatory patterns and find that the standard methods perform quite well. Similar risk-adjustment methods have been applied in other recent work, such as Skinner and Staiger (2015) and Chandra et al. (2016).

The top panel of Figure 4 shows risk-adjusted trends in the 30-day price of treatment measured as the spending per patient in 2014 dollars using an economy-wide GDP deflator. For CHF and pneumonia, the risk-adjusted spending per patient in the year following the event rose from 2001 to 2014. Spending for AMI patients rose from 2001 to 2007 and has since declined to a level below its initial level in 2001. The decline in growth in expenditures later in the period corresponds to a reduction in the growth rates of Medicare fees after 2010. The higher price growth in the private sector would suggest slightly faster price growth for the full population, which we estimate to grow about 0.6 percent faster per year than the Medicare sample.[29] The bottom panel of Figure 4 shows the risk-adjusted 30-day mortality rates. For all three conditions, survival 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 survival are larger for AMI

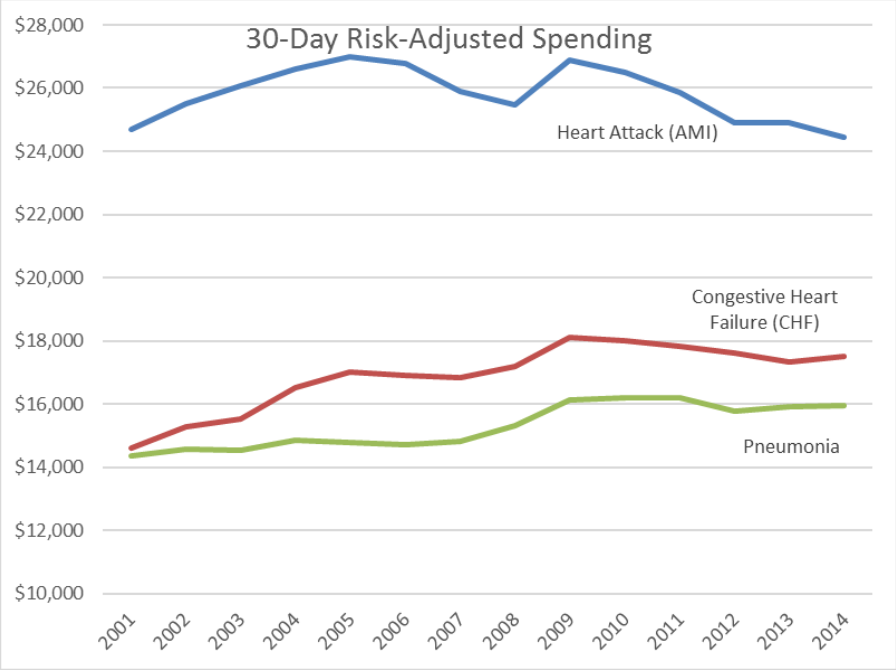
[29] Estimates from the BLS PPI show the growth in both private and Medicare hospital prices. Assuming utilization changes are comparable across populations, this price difference may be used to estimate for the full population treatment price growth. Specifically, we find the growth rate for the private hospital market is 1.6 percent faster per year relative to Medicare. Assuming 65 percent of the relevant population is in Medicare, this would result in treatment price growth that is 0.6 percent faster per year than the Medicare estimates.

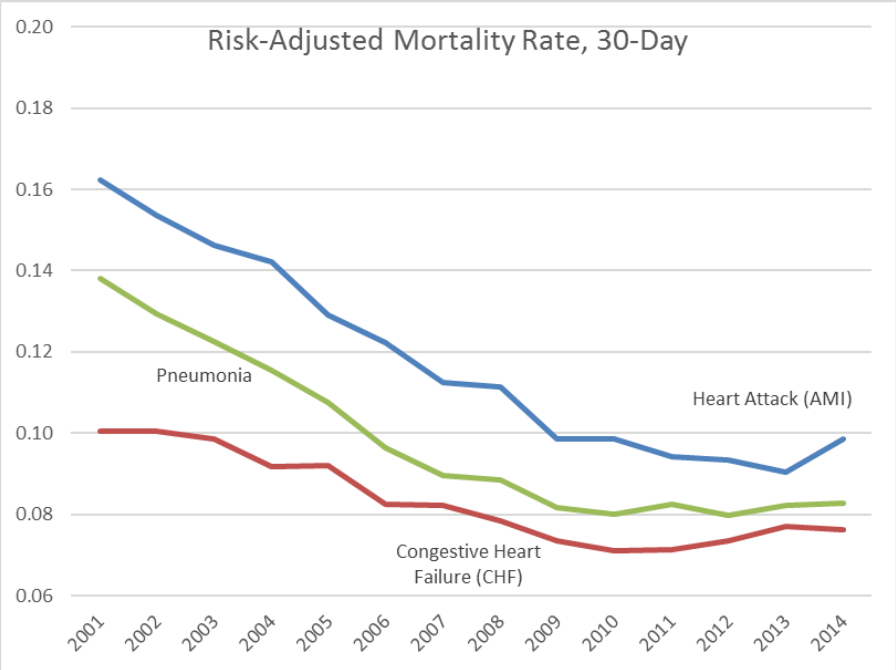
and pneumonia than for CHF, which had longer life expectancy and survival rates to begin with.[30] These improvements in survival rates occurred over a period where there are documented improvements in treatment quality. In particular, the Hospital Compare database tracks "process of care" measures of quality for each of these three conditions with these quality measures first being reported in 2004. For each of these conditions the data shows marked improvement in the share of patients given appropriate treatment, with much of the improvement occurring in the first couple of years (See Table A10). For the case of pneumonia, the percent of patients given the most appropriate initial antibiotic rose 18 percent from 2004 to 2009 (from 77 percent to 91 percent), with two-thirds of the improvement occurring in the first two years. For heart attacks, the improvements in the speed of treatment and coordination among hospital staff is believed to have greatly improved outcomes.[31] It is also interesting to note that many of the process of care measures of quality are not necessarily costly (e.g., given an aspirin), highlighting that increases in treatment quality are not necessarily accompanied by higher costs. Based on Figure 4, factoring in quality change is clearly important, but we will show that the impact on quality-adjusted price indexes greatly depends on the specific index and assumptions applied.

[30] Similar patterns for the price of treatment and mortality may be observed when considering additional days after the initial event, such as a window of 60 or 90 days. These estimates are shown in the appendix in Tables A3 and A4.

[31] "A Sea Change in Treating Heart Attacks", June 19, 2015. New York Times. https://www.nytimes.com/2015/06/21/health/saving-heart-attack-victims-stat.html

Figure 4. 30-Day Risk-Adjusted Spending and Mortality Rates





6. Empirical approach and results

In this section empirically compare alternative quality-adjusted price indexes.

LE index: As discussed above, we construct the LE index as:

$$LE = \frac{S_1 - \frac{U_H}{U_x} H_m(\alpha_1 \cdot m_1 - \alpha_0 \cdot m_0)}{S_0} \quad (4)$$

The key challenge of the LE index is evaluating the monetary benefit of the quality change, $\frac{U_H}{U_x} \cdot H_m(\alpha_1 \cdot m_1 - \alpha_0 \cdot m_0)$. This term has two parts, the marginal valuation of health $\frac{U_H}{U_x}$ and the change in health delivered by medical care $H_m(\alpha_1 \cdot m_1 - \alpha_0 \cdot m_0)$ which we will consider separately.

For the marginal valuation of health $\frac{U_H}{U_x}$, we follow Cutler et al. (1998) who used external research on the value of a statistical life (Viscusi 1993), which attempts to infer the value of life from individual's decisions (e.g., analyzing workers marginal willingness to take a riskier job for different wages). For selecting a range of estimates for the value of a statistical life year, 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).[32] These values are based on a variety of empirical sources such as surveys on willingness to pay and revealed preference studies.[33]

[32] As Pandya et al. 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, Cohen, and Weinstein 2014). While there may be heterogeneity in the value of health in the population, it is often assumed in this literature that $\frac{U_H}{U_x}$ is a constant value representing the dollar value for an additional healthy year of life.

[33] 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+.

Similar to Cutler et al. (1998), we measure $H_m(\alpha_1 \cdot m_1 - \alpha_0 \cdot m_0)$ with the increased life expectancy induced by improvements in treatment. However, simply measuring the observed life expectancy does not isolate the benefits of improved treatment for the condition because changes in treatments 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, which has a few challenges for our application. First, we cannot guarantee that those that survive a heart attack, pneumonia or heart failure are comparable to the rest of the population (e.g., Table 1 shows the comorbidities afflicting each group are distinct). Second, it may be difficult to apply when looking at a broad set of conditions, as it would not be clear how to define the general population (e.g., should we choose those without the particular condition or those without any condition).[34] Finally, one must wait for the resolution of long-term outcomes for the full population, resulting in a significant delay in the estimates.

For these reasons, we take a different approach. We focus on short time horizons around the events, as improvements in survival just after the event are likely attributable to the treatment. Specifically, we only allow the mortality changes to take place over a relatively short window (e.g., 60 days). However, over a longer horizon, trends in the treatment of other conditions and technologies may play an important role. To remove these other factors that affect outcomes over the longer horizon, we assume that the survival rate after the window (e.g., post-60 days) is fixed at the level observed for individuals surviving the event at the beginning of the sample. Additional details are provided in the appendix.

[34] 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.

In calculating our LE indexes, we use a range of values for both the length of the mortality window over which we measure health outcomes 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.[35] The estimates of unadjusted indexes and indexes adjusted for the changes in quality are reported in Table 2.

Table 2
Annual growth rates of LE indexes across different assumptions

Window length	30 days			60 days			90 days		
Annual value of life	$50,000	$100,000	$150,000	$50,000	$100,000	$150,000	$50,000	$100,000	$150,000
Heart Attack (AMI)									
Unadjusted index	-0.1%			-0.1%			-0.1%		
COLI	-4.8%	-9.8%	-15.1%	-5.1%	-10.5%	-16.2%	-5.5%	-11.3%	-17.5%
Congestive heart failure (CHF)									
Unadjusted index	1.4%			1.5%			1.5%		
COLI	-0.4%	-2.3%	-4.4%	-0.2%	-2.1%	-4.2%	-0.1%	-2.0%	-4.1%
Pneumonia									
Unadjusted index	0.8%			0.9%			1.0%		
COLI	-4.4%	-9.9%	-15.8%	-4.3%	-10.0%	-16.1%	-4.1%	-9.6%	-15.6%

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 2. First, quality adjustment turns out to be important across all assumptions. For each scenario, we observe the quality adjustment having a significant impact relative to the unadjusted index. The unadjusted indexes show annual price increases slightly above general inflation across conditions, while the growth rates of the quality-adjusted indexes are all negative.

[35] The estimates are rebased by the amount each year and the index growth rate is chained across years. One advantage of this approach is that it avoides potentially negative values in the index that may occur from drastic changes in index values. For example, if the change in welfare is particularly large, then the numerator of the index could become negative. This issue is avoided by rebasing and chaining the index keeping the innovations incremental. This issue will be discussed further in a later section of this paper.

Table 2 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 window over which we measure health benefits. Fixing the value of a statistical life year (VSLY) 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 VSLY to be $50,000 compared to a value of $150,000 can change the inflation rate by a larger amount. Averaging across conditions based on expenditure share[36] and holding the days of measured benefit to be 60, the average annual price decline is 3.1 percent for VSLY of $50,000, 12.0 percent for VSLY of $150,000, and our central estimate is a decline of 7.4 percent for VSLY of $100,000.

Our results are similar to Cutler et al. (1998, 2001) in showing rapid price declines, although the declines we find over our period of study are smaller. Estimates in Cutler et al. (2001) show rapid price declines of around 14.4 percent a year based on the relatively conservative estimates that the value of a statistical life year is worth $25,000 in 1991 dollars ($39,000 in 2014).[37] In the period we study, for AMI specifically, we find an annual price decline of 4.8 percent for our most

[36] The expenditure share is calculated based on 60-day spending on treatment in the base year 2001 multiplied by the number of cases. Expenditure share for heart attacks is 29 percent, expenditure share for congestive heart failure is 35 percent, and expenditure share for pneumonia is 36 percent.

[37] Sheiner and Malinovskaya (2016) use economy-wide COLI estimates reported in Cutler et al. (2001) to form a disease-specific index more comparable to our estimate, but consistent with the utility theory in Cutler et al. (2001). Cutler et al. (1998; 2001) find an annual inflation rate of around 1 to 2 percentage points below general inflation. However, there are some important differences in how they derive this estimate. As Sheiner and Malinovskaya (2016) note, the index formed by Cutler et al. (1998; 2001) uses income in the denominator, which provides more of an indicator of the change for the aggregate deflator, rather than forming a disease-specific index. Sheiner and Malinovskaya (2016) show how a disease-specific utility-based price index may be formed from data reported by Cutler et al. (2001).

conservative estimate assuming $50,000 per statistical life year, relative to general inflation.[38] The faster decline found by Cutler et al. (2001) is in line with expectations, as Cutler et al. (2001) study price trends of heart attack treatments during a period of rapid technological improvement for treating this condition, including the expanded use of effective treatments such as bypass surgery, beta blockers, aspirin, ace inhibitors, and angioplasty.

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, Goldman, Sood (2015) who study the same acute inpatient conditions that we consider here. For each condition, we define the price in each period as the average annual incremental per patient cost of successfully achieving the treatment endpoint shown in equation (6), $\frac{S_1/\sigma_1}{S_0/\sigma_0}$, where S_t is average risk-adjusted spending as defined above and σ_t is the risk-adjusted percent of treatments that are successful relative to no treatment.[39] Similar to Romley, Goldman, Sood (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).

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 or assume the rate of reaching the endpoint without any medical treatment. Berndt et al. (2002) estimated the rate of remission of major

[38] 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 3.1 percentage points below general inflation. To construct the weights, we multiply the number of observations for each condition by the 60-day spending estimate for each condition. The weights are 29 percent for heart attacks, 35 percent for heart failure, and 36 percent for pneumonia.

[39] In the stylized model of Appendix 1, $\sigma_t = q_t\pi_1 + (1-q_t)\pi_2 - \pi_3$, where π_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.

depression without any treatment based on expert opinion because it was not uncommon for major depression to go untreated. 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, as assumed in Romley, Goldman, Sood (2015). However, prior to the development of modern treatments, there was the potential for survival for all three conditions, so we estimate the quality-adjusted indexes based on different assumptions regarding the success of untreated cases.[40]

Table 3 shows alternative indexes based on differing assumptions for untreated cases and different window lengths for measuring outcomes and spending.[41] 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. As we increase the assumed success rate of untreated cases, the incremental change in health has a larger impact on inflation. As expected based on our theoretical discussion, the inflation rates observed here are higher than the inflation rates observed based on the LE index.

[40] 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."

[41] The Appendix Table A6 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. Using the TE index, the quality-adjusted 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 3
Annual growth rates of treatment endpoint index under alternative assumptions

	Window length								
	30 days			60 days			90 days		
	Assumed success rate of untreated cases								
	0%	20%	40%	0%	20%	40%	0%	20%	40%
Acute myocardial infarction									
Unadjusted index	-0.1%			-0.1%			-0.1%		
Quality-constant index	-1.0%	-1.4%	-2.1%	-1.3%	-1.8%	-3.3%	-1.5%	-2.2%	-4.4%
Congestive heart failure									
Unadjusted index	1.4%			1.5%			1.5%		
Quality-constant index	1.0%	0.8%	0.5%	0.9%	0.6%	-0.1%	0.8%	0.5%	-0.8%
Pneumonia									
Unadjusted index	0.8%			0.9%			1.0%		
Quality-constant index	0.2%	0.0%	-0.5%	0.1%	-0.3%	-1.2%	0.1%	-0.4%	-1.8%

Notes: Estimates are computed as compound annual growth rates. Price index is based on dollar figures deflated to 2014 dollars with the GDP deflator.

Hedonic index: The next method follows Frank et al. (2004), by using hedonics to control for the characteristics of treatment over time. Specifically, we run the following generalized linear model (GLM) regression, separately for each condition and year:

$$Y_i = \alpha_t + X_i\beta_t + Z_i\gamma_t + \varepsilon_i.$$

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.[42] We then construct a Laspeyres-type index where the average price for year t is the average predicted treatment price with the prediction run on the population and treatments from 2001 using the $\widehat{\alpha_t}$, $\widehat{\beta_t}$ and $\widehat{\gamma_t}$from year t, essentially using the approach suggested by Pakes (2003). Next, we construct a Paasche-type index using the

[42] We apply a GLM model using a log-link and gamma distribution due to the skewness of the expenditure data.

same method on the population and treatments in 2014. The final index is a Fisher index, that is the geometric average of the two, following the method of Frank et al. (2004).[43]

For both AMI and CHF, we are able to identify relevant technologies to include in Z_i.[44] 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 hedonic indexes for the pneumonia cohort.

When we apply the hedonic method to AMI and CHF, we find that there is very little difference between the hedonic 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 3-5).

The hedonic indexes diverge from the LE and TE indexes that explicitly incorporate health outcomes, and which decline substantially. 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 of AMI and CHF that we observe may have been caused by shifts among other treatments that are not contained in the claims data. As mentioned previously, for heart attacks, the improved speed of

[43] Rather than using only a base-period technology for our hedonic function, we use a Fisher index. Our results are robust to alternative methods. For instance, we ran a simple hedonic model with year dummies and hedonic controls and found similar results.

[44] 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.

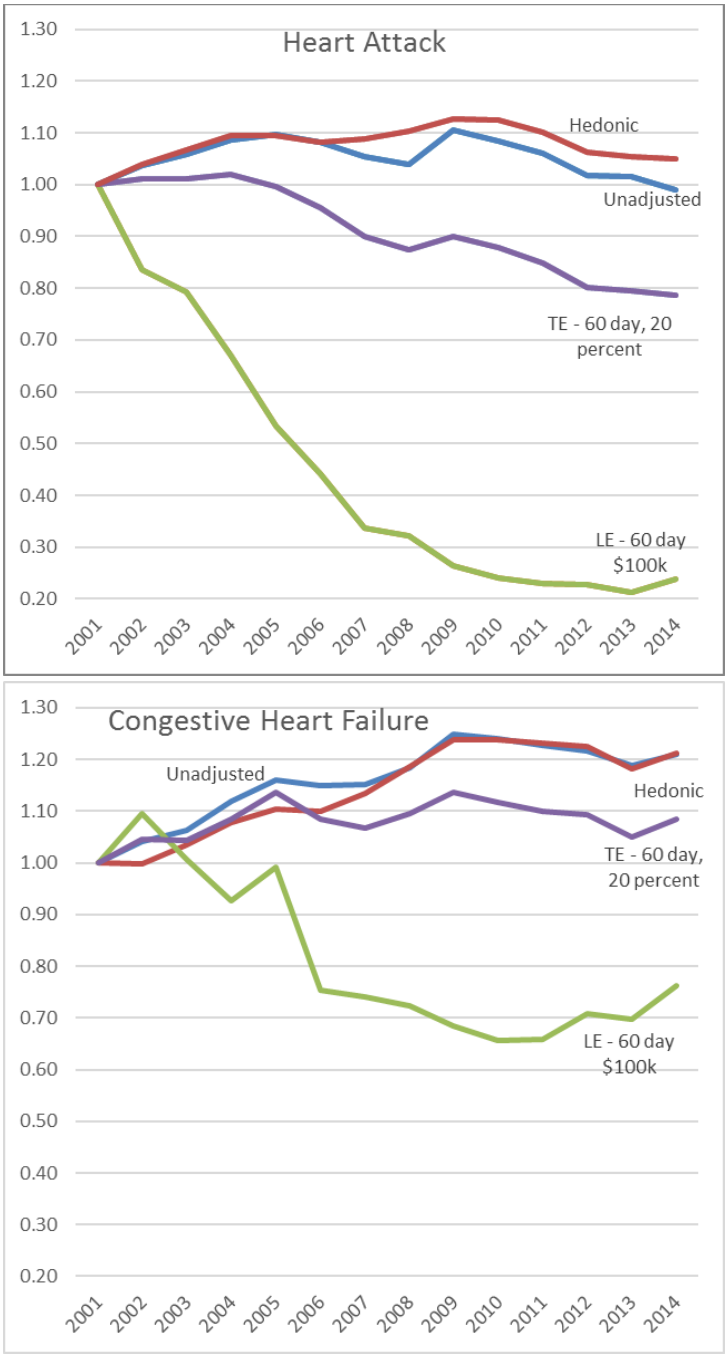
of treatment and coordination of care greatly improved outcomes, which are factors not captured by claims data sources. In addition, many of the "process of care" measures of quality reflected in the Hospital Compare database discussed previously showed large improvement, even though many of these quality measures did not involve costly treatment.

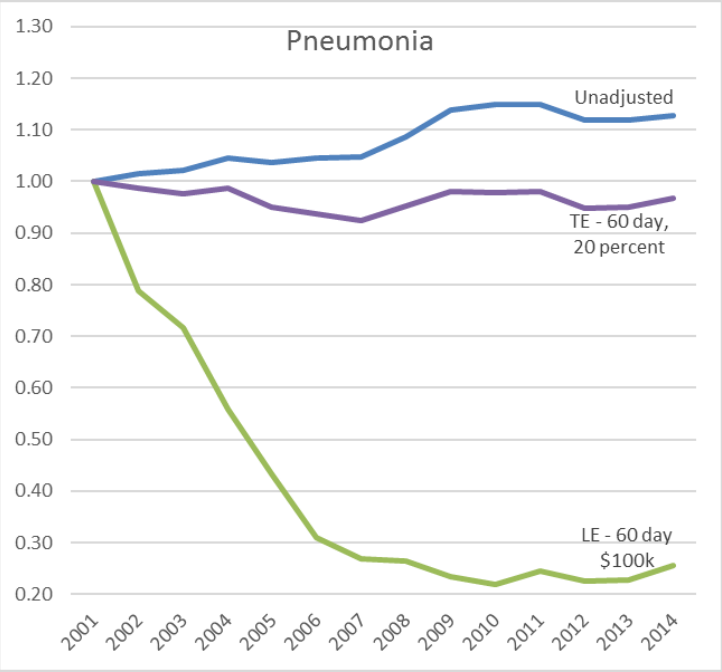
Across-Method Comparison

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, which is also in the middle of our assumptions. For both indexes, we select a 60-day window.

Figures 5 illustrate the differences between the indexes. Across the three conditions we find similar patterns. We find that the unadjusted index and the hedonic index are nearly identical within conditions. We find that inflation is considerably lower when measured by health outcomes in both the TE index and LE index, relative to the unadjusted index, but the amount of the adjustment is much larger for the LE index. While we are presenting the indexes based on a single set of assumptions, the difference in the growth rates in Tables 2 and 3 suggests that this difference between the TE and LE indexes is robust to alternative assumptions. Consistent with our theoretical discussion, the estimates for these conditions suggest that the TE and hedonic indexes tend to overstate the rate of inflation, relative to our preferred LE index.

Figure 5. Comparison of Indexes for Heart Attacks, Congestive Heart Failure and Pneumonia





7. Study of New Innovations from the Tufts Cost-Effectiveness Database

We show that quality adjustment is important for the three selected conditions, but it is not clear if those conditions are representative of the impact of innovation on health care price more generally. To address this concern, we reconsider the findings of Hult, Jaffe, and Philipson (2018). Their study uses a dataset of cost-effectiveness studies from the Tufts Medical Center Cost-Effectiveness Analysis Registry (CEAR) database. The registry database, intended to be a comprehensive database of a wide variety of treatments and diseases, summarizes and reviews published original cost-effectiveness studies, where each article is screened and reviewed before inclusion in the registry. To satisfy the criteria for inclusion in the database the research must be published in English, be an original cost-effectiveness analysis, and measure health benefits as QALYs. Review articles, editorials, and articles missing key features (e.g., quality measures) are excluded. Each article is reviewed by two readers that have been trained in cost-effectiveness and decision analysis. These readers follow a standardized set of forms and instructions and extract

over 40 variables for each article, as well as provide specific ratings regarding the quality of the study. The studies vary on numerous dimensions that are recorded in the data: type of intervention (e.g., pharmaceutical), condition treated (e.g., cardiovascular), funding source (e.g., government), as well as numerous other variables. The types of studies vary in the methods that are applied, which are described in the abstract of each paper that is one of the included data elements. In contrast to claims-based approach applied in the previous section, which relies on risk adjustment to remove potential biases in the quality and cost estimates, the studies here present a diverse array of methods applied in the medical literature. Based on a word search of the title and abstract, we find that about 37 percent of the articles have the word "random" or "trial".[45] However, many of the studies may be meta-studies or disease-model simulations that are often based on randomized trials. The quality of each study is rated by the readers of the study based on a variety of criteria (e.g., health economic methodology, consideration of uncertainty, and transparency). The methods forming both the cost and QALY estimates vary depending on the study, but they are unified in their goal of estimating the key elements that are necessary to evaluate the cost effectiveness of treatment, which are the same elements needed to form a price index.

The latest version of this database applied in our study contains 7,287 cost-effectiveness studies with about 90 percent of the studies coming from the 2004 to 2017 period. Many of the studies in the database contain the critical four elements for understanding the price impact of new innovations: (1) the price of treatment for the new innovation (i.e., insurer plus patient costs); (2) the price of treatment for the previous standard of care (SOC); (3) the QALYs produced by the innovation; and (4) the QALYs produced by the previous standard of care. The standard of care treatment typically represents the incumbent treatment prior to the arrival of the new innovation.

[45] This is based off of a simple search of the title and abstract for the word "random" or "trial".

About 50 percent of the articles in the database includes all four of these elements, so not every study may be used to form a quality-adjusted price index. However, a single article may contain multiple comparisons of treatments, increasing the number of innovations that may be analyzed. Our version of the data contains three additional years relative to Hult, Jaffe, and Philipson (2018), and we have a total of 10,000 observations for which we observe the necessary elements to form quality-adjusted price indexes.

Before reporting our results from the Tufts database, we start by analyzing the results reported in Hult, Jaffe, and Philipson (2018). Hult, Jaffe, and Philipson (2018) use the database to calculate quality-adjusted prices for a wide set of medical treatments using a TE index formula based on the average price of a QALY, as described previously: $TE = \frac{s_1/QALY_1}{s_0/QALY_0}$ where the innovation corresponds to period 1 treatment and the standard of care corresponds to period 0 Based on this formula, they find the median quality-adjusted price change for a new innovation to be an increase of 4 percent relative to the prior standard of care. As we have shown, measuring the price per successful treatment or QALYs using a TE index may understate the gains in welfare relative to our preferred utility-based LE index. To relate the LE formula to estimates reported in Hult, Jaffe, and Philipson , we first re-write equation (4) where the innovation corresponds to period 1 treatment and the prior standard of care corresponds to the base period 0, as:

$$LE = \frac{S_0 - \mathrm{CV}}{S_0} = \frac{S_0 - \left((U_H/U_x)\Delta H \; - \; (S_1 - S_0)\right)}{S_0} = \frac{S_0 - \left(\frac{U_H}{U_x} \; - \; \frac{(S_1 - S_0)}{\Delta H}\right)\Delta H}{S_0}$$

where $\Delta \mathrm{H} = QALY_1 - QALY_0$ is the change in health (assuming $\Delta \mathrm{H} \neq 0$), measured by QALYs, added by the new innovation relative to the prior standard of care treatment (or $H_m(\alpha_1 \cdot m_1 - \alpha_0 \cdot m_0)$). The second term in the numerator is a measure of compensating

variation from the new innovation. In the above formula, the compensating variation is rewritten as the net value gained (or lost) per unit increase in health from the new innovation, $\left(\frac{U_H}{U_x} - \frac{(S_1 - S_0)}{\Delta H}\right)$, times the observed change in health, ΔH. The marginal cost per increase in health, $\frac{(S_1 - S_0)}{\Delta H}$, is often referred to as the incremental cost-effectiveness ratio (ICER). Based on estimates reported in Hult, Jaffe, and Philipson (2018) they find the median value of the ICER in their data is \$17,415. If we conservatively assume the value of a QALY ($\frac{U_H}{U_x}$) is \$50,000, then the term $\left(\frac{U_H}{U_x} - \frac{(p_1 m_1 - p_0 m_0)}{\Delta H}\right) = \$50{,}000 - \$17{,}415 = \$32{,}485$, which indicates the value gained per QALY for the median innovation. Since this value is positive, the LE index is less than one indicating falling quality-adjusted prices for the median innovation.[46] In other words, based on the estimates reported in Hult, Jaffe, and Philipson (2018) over half of the new innovations in the database lead to falling prices using the LE index formula.

Next, to obtain a more complete picture of the price decline we turn to the micro-data from CEAR to estimates the quality-adjusted price change for all innovations in the database. To clean the data, we first take the same steps outlined in the work by Hult, Jaffe, and Philipson to remove some of the outlier studies and estimates.[47] In the top of Table 4 we report the same descriptive statistics that are provided in Hult, Jaffe, and Philipson, but using our larger sample. This includes information for the "innovator" and the prior standard of care "SOC". These elements include the

[46] This assumes that $\Delta H > 0$.

[47] The selection rules outlined in Hult, Jaffe, and Philipson (2018): "We omit observations with quality values greater than 100, since it does not make sense for a treatment to add more than 100 years to someone's life. We also omit studies with negative quality values. We omit observations with negative cost for either the innovation or the SOC. We also omit observations where the ICER, price, or price per QALY for the innovation or the SOC is over \$10." In order to normalize expenditures in the studies across years to the year 2014, we use a medical care deflator to ensure that the same quantity of medical care may be purchased in 2014 as in the year of the study. We convert medical expenditures into 2014 dollars using the PCE deflator for medical care, rather than the medical CPI, which is only relevant for out-of-pocket costs (Dunn, Grosse, and Zuvekas2018). However, the main findings are not changed by the use of either index. We convert to U.S. dollars using yearly exchange rates.

innovator QALY, SOC QALY, Innovator Price, SOC Price, Innovator Price per QALY, SOC Price per QALY and the ICER. Overall the descriptive statistics are very similar to those reported in Hult, Jaffe, and Philipson.

The bottom of the Table reports the distribution of quality-adjusted prices using both the TE index and the LE index.[48] The TE index shows a median index of 1.04, indicating a 4 percent increase, which matches the result in Hult, Jaffe, and Philipson. The mean price increase based on the TE index is 34 percent. In other words, based on the TE index the average innovation represents a price increase, again matching the finding in Hult, Jaffe, and Philipson. These estimates contrast with the estimates obtained from the LE index that shows clear quality-adjusted price declines at both the mean and median of the distribution across all VSLY estimates. In fact, the mean LE index level is negative for VSLY of $100k or $150k. The negative level is caused by the welfare improvement exceeding the treatment price, which is a problem that may occur for drastic improvements in technology (Trajtenberg 1990). This implies that for the individuals to be indifferent between receiving the standard of care and the innovation, they would need to receive the standard of care product for free and additional cash to make up for the total loss in quality from giving up the newer technology. While we can interpret these negative index levels, they cannot be used as deflators to calculate real output.

To avoid negative index values resulting from large technical change when examining pneumonia, heart attack, and heart failure for the Medicare population, we chained index growth rates, but this is not possible for examining the innovations where there is only one price change. Instead, we address this issue by following the advice of Trajtenberg (1990) and construct an alternative utility-

[48] For all indexes the top and bottom 1 percent of the indexes are not considered in the reported distribution. This is to avoid outliers influencing the mean estimate. Qualitatively the results are robust to the inclusion or exclusion of these extreme values in the distributions.

based index based on the reservation price of the new technology. In this index, the denominator represents the reservation price of the new technology that makes individuals indifferent between the innovation and the previous technology (i.e., : LE reservation price=$\frac{e(p_1,U_1)-x_1}{e(p_0,U_1)-x_1} = \frac{e(p_1,U_1)-x_1}{e(p_1,U_1)+(e(p_0,U_1)-e(p_1,U_1))-x_1} = \frac{S_1}{S_1+CV} = \frac{S_1}{S_0+\left(\frac{U_H H_m(\alpha_1 \cdot m_1-\alpha_0 \cdot m_0)}{U_x}\right)}$). By construction, this LE reservation price index is positive for all values of improved quality. Both indexes are equally valid based on utility theory, and Trajtenberg (1990) suggests taking an average of the two, but using this average would still show a negative average price based on the VSLY of $100k or $150k. The LE reservation price index is shown at the bottom of Table 4 and shows clear declines in price from innovation at both the mean and median across all VSLYs.

We conclude that a conservative measure of the average price decline from innovation would be around 20 percent, since the mean price drop falls near 20 percent or more for five of the six average LE indexes and for four of the six median LE indexes. If medical care markets are responsive to price so that technologies with lower quality-adjusted prices are more likely to be adopted and diffused, then both the median and mean quality-adjusted prices could actually understate the impact of new technologies.

Table 4
Innovations and Quality Adjusted Price Estimates

	Mean	Median	p5	p95	sd	obs
Innovator QALY	9.76	8.05	0.19	25.56	9.90	10,066
SOC QALY	9.36	7.50	0.10	25.56	9.81	10,048
Innovator Price	$108,682	$22,799	$265	$372,886	$459,697	10,537
SOC Price	$92,513	$17,723	$77	$318,882	$414,086	10,525
Innovator Price per QALY	$22,630	$4,563	$28	$91,267	$142,215	9,905
SOC Price per QALY	$19,851	$3,796	$16	$84,969	$239,409	9,740
ICER	$69,437	$16,407	-$133,495	$405,937	$612,663	$17,459
		Quality-Adjusted Price Indexes				
TE Index	1.35	1.04	0.65	2.86	1.14	9,455
LE Index						
($50,000 VSLY)	0.21	0.94	-4.04	2.54	4.69	9,455
($100,000 VSLY)	-1.17	0.79	-10.26	2.51	9.55	9,455
($150,000 VSLY)	-2.55	0.63	-16.72	2.57	14.65	9,455
LE Reservation Price Index						
($50,000 VSLY)	0.80	0.92	0.07	1.64	6.69	9,453
($100,000 VSLY)	0.77	0.79	0.04	1.58	0.56	9,453
($150,000 VSLY)	0.75	0.69	0.02	1.60	3.22	9,453

Notes: Estimates with outlier values in QALYs and costs specified in the text have been removed prior to the construction of this table. For the indexes, the bottom and top 1 percent of the distribution have been removed for the construction of this table so that outliers have a limited effect on the mean. Results are robust to the outlier removal procedure. For instance, removing observations that are outliers for any one of the indexes produces nearly identical results.

As the TE index methodology values QALYs based on the price per QALY, it is clear that this approach will tend to undervalue technological change, as the median price per QALY is around $4,000 (far below any estimate of the value of a statistical life). This finding is consistent with the results of Figures 3, which suggests that the average price per unit of health will be much lower than the marginal value per unit of health, leading to the empirical difference we observe across the indexes. Showing the estimates from a well-known example helps to highlight this point. Consider the case of Sovaldi, a well-publicized hepatitis treatment, which was viewed as a costly, but effective new innovation. For a patient with cirrhosis the innovation using Sovaldi had a price

of treatment of $99,908 with a QALY of 9.40, while the standard of care had a price of $76,915, with a QALY of 8.28. In this case, the TE index is 1.14, while the LE index is 0.57 (VSLY $50,000). The LE index shows the Sovaldi treatment to be lowering quality-adjusted prices, while the TE index implies that it is driving quality adjusted prices higher. This is caused by the TE index implicitly valuing the additional 1.12 years of life at just $10,000.

Tables A8 and Table A9 in the appendix show additional details based on disease categories of the innovation (e.g., cardiovascular or musculoskeletal),type of intervention (e.g., pharmaceutical or device), the funding sponsor (e.g., government or pharmaceutical maker), and type of study based on a simple word searches of the title and abstract (e.g., randomized or simulation).[49] Table A8 shows estimates for the LE reservation price index (VSLY $100k) and Table A9 shows estimates for the TE index, respectively. These tables also show an additional breakout of high-quality studies based on the evaluations of the readers scoring the quality of the research studies along various dimensions. While there are some differences in the mean and median across disease categories, type of intervention, funding sponsor, and type of study, what stands out most is the persistent difference between the LE reservation price indexes and TE indexes within all categories. The LE index shows consistent price declines, while the TE index shows price increases. Overall, the LE indexes reported in Table A8 strongly suggests that price declines from innovation are a prevalent feature of the health care sector, showing declines at both the mean and median across all categories.

[49] If the title or abstract contain the word random or trial and does not contain the word "meta", then we categorize the study as randomized. If the title does not contain the word random or meta, but includes the word simulation or markov, then we categorize the model as a simulation. If the word title contains the word meta, then we categorize it as a meta-study.

Simple correlations in these data suggest that not properly accounting for quality improvements will lead to systematic biases. A regression of the log price of the new innovation on the log price of the standard of care treatment, the log QALY of the standard of care treatment, and the log incremental gain in quality from the innovation is shown in Table 5. The regression shows three things. First, the cost of new treatments tend to be correlated with the costs of previous treatment, as we might expect. Second, holding the cost of the standard of care treatment constant, the magnitudes of improvements in treatment outcomes, relative to the standard of care, are correlated with higher prices of innovative treatments. Only the incremental improvement in the QALY is related to price, while the standard of care QALY alone shows little correlation. Consequently, not placing any value on quality improvement will lead to a systematic upward bias in the price of new innovations. As explained throughout this article, choosing the correct value to place on the quality improvements is critical for obtaining economically meaningful estimates.

Table 5
Regression of Log(Innovator Price) on log(QALY) difference, log(SOC Price), and log(SOC QALY)

	Full Sample	Year>=2013	Year<2013	Manufacturer Funded	Not Manufaturer Funded
log(QALY Innovation)-log(QALY SOC)	0.612***	0.754***	0.500***	0.632***	0.605***
	(0.0621)	(0.0427)	(0.0751)	(0.125)	(0.0525)
log(QALY SOC)	0.0255	0.0256	0.0289*	0.0409***	0.0193
	(0.0168)	(0.0248)	(0.0130)	(0.0106)	(0.0202)
log(SOC Price)	0.923***	0.939***	0.908***	0.924***	0.924***
	(0.00849)	(0.00563)	(0.0129)	(0.0156)	(0.0113)
Number of Observations	9571	5072	4499	2998	6573
Adjusted R2	0.926	0.930	0.923	0.937	0.921

Notes. Standard errors in parentheses with * p<0.10, ** p<0.05, *** p<0.01. Standard errors are clustered by disease category for all estimates.

Hult, Jaffe, and Philipson (2018) note that their findings imply that health care is somehow different from other high-technology industries that are typically characterized by large quality-adjusted price drops. However, here we show that when a more theoretically grounded method is applied, the price changes we observe from new innovations actually seem to be quite similar to those in other high-technology industries.

8. Implications for Productivity

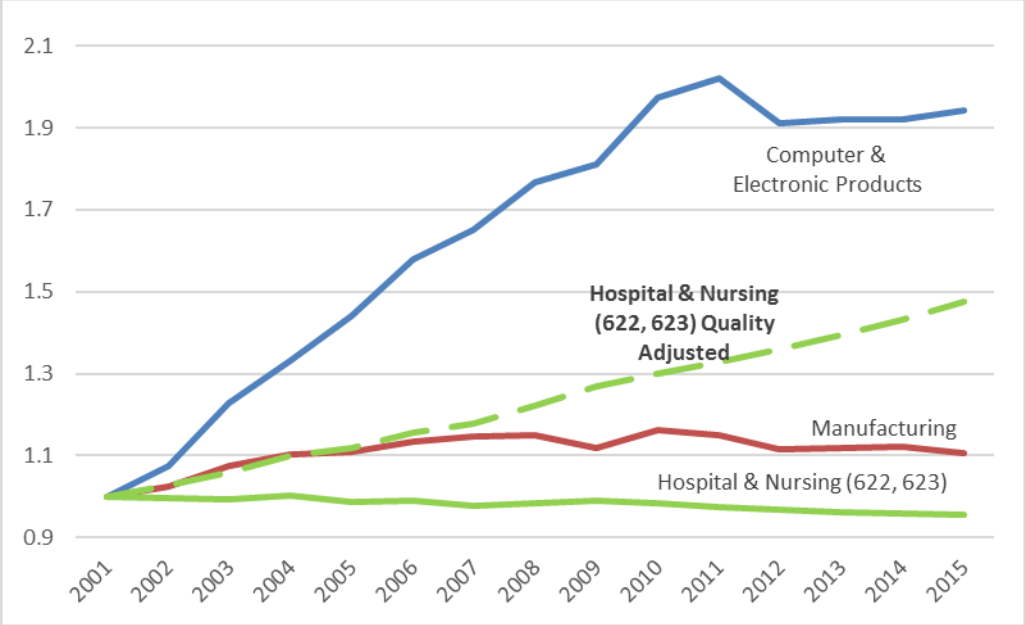
If official health care price indexes do not account for changes in quality, this has implications for official measures of output and multifactor productivity growth that rely on these indexes. The official estimate of multifactor productivity growth most related to our study is from BLS and covers Hospitals and Nursing and Residential Care Facilities (North American Industry

Classification System (NAICS) industries 622 and 623). The official estimate shows a multifactor productivity growth rate that declines by 0.3 percent per year from 2000-2014.

The potential effect of quality adjustment on multifactor productivity growth depends on the magnitude of the quality adjustment bias. Evidence from the Tufts registry implies that the quality-adjustment bias from new innovations is prevalent and potentially quite large, but it is difficult to determine the specific annual quality-adjusted price change that would be broadly representative based on these data. Instead, we turn to the price indexes based on the three conditions we studied. To keep our estimates conservative, we use the value of a statistical life year of $50,000, which implies a bias adjustment amount of 3.1 percent per year.[50] We incorporate the quality adjustment by deflating the output price index by 3.1 percent per year over the period of study and then re-computing a new quality-adjusted output and new productivity index (see Table A7 of the Appendix). With this alternative estimate, we find that the quality-adjusted productivity growth rate becomes 2.8 percent per year. Figure 6 shows the multifactor productivity estimates from BLS for three categories for comparison: computer & electronic products, manufacturing, and hospital and nursing (NAICS 622, 623). For hospital and nursing we also show the quality-adjusted estimate as the dashed line. After the quality-adjustment, the hospital productivity estimate exceeds that of the manufacturing sector and is more comparable to a high productivity growth sector such as computer and electronic products.

[50] 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 3.1 percent per year so the bias is 3.9 percentage points. We restrict our adjustment only to the hospital sector by applying the quality adjustment to 80 percent of output because hospitals account for 80 percent of expenditures for NAICS industries 622 and 623. Therefore, the adjustment amount becomes 3.1 percent per year.

Figure 6. Multifactor Productivity Growth Comparison (Based to 1 in 2001)



This hypothetical estimate makes the strong assumption that the magnitude of the quality-adjustment bias that we estimate for our select conditions can apply to a wider set of medical conditions than those we consider here. While this estimate should be viewed as a bit crude, the broad evidence from both the three conditions and the CEAR database, suggest that price declines from innovation are broadly occurring and with a substantial magnitude. Given the prevalence and magnitude of the declines from innovation, we view our measure of the quality-adjustment bias as a reasonable lower bound.

Conclusion

This paper provides comprehensive evidence that innovations commonly lead to quality-adjusted price declines in the medical care sector. We find that applying the appropriate quality-adjustment methodology is critical for obtaining a meaningful quality-adjusted index. The utility-based COLI price 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 and robust results. Important differences can arise between the utility-based method and other indexes when the marginal valuation of life differs from the average price per unit of health produced. These differences are found to be of great empirical importance for the thousands of cost-effectiveness studies in the Tuft's CEAR database and for the three actue conditions studied using Medicare claims.

Applying the utility-based method of quality-adjustment to the three conditions from our claims database as well as the more comprehensive CEAR database suggest substantial quality-adjusted price declines from new innovations. The robustness of these findings across data sources, disease categories and types of interventions suggest that quality-adjusted prices declining from new innovations is a prevalent feature of the sector. Although quality-adjustment from innovations is shown to be substantial in this study, these quality changes are not currently reflected in official estimates. This work suggests that quality-adjustment may be of great practical importance for understanding price trends, output and productivity in the health care.

An observed decline in quality-adjusted prices in itself does not imply that the health care system is functioning optimally following the price fall since it alone says nothing about whether or not full efficiency has been achieved. If the decline results from better employment of existing technology (either a reduction in non-cost-effective technology or an increase in cost-effective technology), the decline will correspond to an improvement in health-care efficiency but further improvements (and price declines) may yet still be possible.

While we are able to show that there may be substantial quality-adjusted price declines from new innovation, more work is needed to incorporate this information into annual disease-based price indexes. It will be important for academic researchers and statistical agencies to continue research

to build a consensus around quality adjustment methods that may be applied to the health care sector more broadly (Schreyer 2010). Until a consensus is formed, it may be important to report a range of estimates for the quality-adjusted prices, rather than applying a single method or set of assumptions. 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 become available, and valuations of health become more certain.

References

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li, "Missing Growth from Creative Destruction," American Economic Review, forthcoming, 2019.

Aizcorbe, Ana, and Nicole Nestoriak, "Changing mix of medical care services: stylized facts and implications for price indexes," Journal of Health Economics, 30(3) (2011), 568-574.

Aldy, Joseph E., and W. Kip Viscusi, "Adjusting the value of a statistical life for age and cohort effects," The Review of Economics and Statistics, 90(3) (2008), 573-581.

Anderson, Robert N, "United States life tables," National vital statistics reports, 48(18) (1998), 1-40.

Baumol, William J, "Macroeconomics of unbalanced growth: the anatomy of urban crisis," The American economic review 57(3) (1967), 415-426.

Berndt. Ernst R., A. Bir, S. H. Busch, R. G. Frank, and S. T. Normand "The Medical Treatment of Depression, 1991-1996: Productivity Inefficiency, Expected Outcome Variations, and Price Indexes," Journal of Health Economics, 21(3) (2002), 373-96.

Bils, Mark, and Peter Klenow, "Quantifying Quality Growth," American Economic Review, 91(4) (2001), 1006-1030.

Bradley, Ralph, Jaspreet Hunjan, and Lyubov Rozental, "Experimental Disease-Based Price Indexes," BLS working paper, (2015).

Bratzler, Dale W., Sharon-Lise T. Normand, Yun Wang, Walter J. O'Donnell, Mark Metersky, Lein F. Han, Michael T. Rapp, and Harlan M. Krumholz,"An Administrative Claims Model for Profiling Hospital 30-Day Mortality Rates for Pneumonia Patients," PLoS ONE, 6(4) (2011), e17401.

Braunwald, Eugene "The treatment of acute myocardial infarction: the Past, the Present, and the Future," European Heart Journal: Acute Cardiovascular Care, 1(1)(2012), 9-12.

Bresnahan, Timothy F., and Robert J. Gordon, “Introduction to “The Economics of New Goods”,” in The Economics of New Goods ed (The Universirty of Chicago Press, 1996).

Broda, Christian, and David E. Weinstein, “Product Creation and Destruction: Evidence and Price Implications,” American Economic Review, 100(3) (2010), 691–723.

Brot-Goldberg, Zarek C., Chandra, A., Handel, B.R. and Kolstad, J.T., “What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics,” The Quarterly Journal of Economics, 132(3) (2017), 1261-1318.

Byrne, David M., John G. Fernald, and Marshall B. Reinsdorf, “Does the United States have a productivity slowdown or a measurement problem?” Brookings Papers on Economic Activity, 2016(1) (2016), 109-182.

Center for the Evaluation of Value and Risk in Health. The Cost-Effectiveness Analysis Registry [Internet]. (Boston), Institute for Clinical Research and Health Policy Studies, Tufts Medical Center. Available from: www.cearegistry.org Accessed on February 2019.

Chandra, Amitahb, Amy Finkelstein, Adam Sacarny and Chad Syverson, “Healthcare Exceptionalism? Performance and Allocation in the U.S. Healthcare Sector,” American Economic Review, 106 (8) (2016), 2110-2144.

Chandra, Amitabh, Maurice A. Dalton, and Jonathan Holmes, “Large increases in spending on postacute care in Medicare point to the potential for cost savings in these settings,” Health Affairs, 32(5) (2013), 864-872.

Chernew, Michael, and Joseph P. Newhouse, “Health Care Spending Growth,” Handbook of Health Economics, 2 (2011), 1-43.

Cutler, David M., Michael Chernew, Kaushik Ghosh, and Mary Beth Landrum, “Understanding the Improvement in Disability Free Life Expectancy in the U.S. Elderly Population,” NBER Working Paper No. 22306, 2017.

Cutler, David M., M. McClellan, J.P. Newhouse, and D. Remler, “Are medical prices declining? Evidence for heart attack treatments,” Quarterly Journal of Economics, 113(4) (1998), 991-1024.

Cutler, David M., McClellan, M., Newhouse, J. P., and Remler, “Pricing Heart Attack Treatments,” In Medical Care Output and Productivity. David Cutler and Ernst Berndt, ed (The University of Chicago Press, 2001), 305-62.

Cutler, David M., Allison B. Rosen, and Sandeep Vijan, “The Value of Medical Spending in the United States, 1960-2000,” New England Journal of Medicine, 355(9) (2006), 920-927.

Diewert, Erwin, and Robert Feenstra, "Estimating the Benefits and Costs of New and Disappearing Products," NBER Working Papers, (2018).

Doyle, Joseph J., John A. Graves, and Jonathan Gruber. "Evaluating Measures of Hospital Quality." The Review of Economics and Statistics. Forthcoming.

Doyle Jr, Joseph J., John Graves, Jonathan Gruber, and Samuel Kleiner, "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns," Journal of Political Economy, 123(1) (2015), 170-214.

Duernecker, Georg, Berthold Herrendorf, and Akos Valentinyi, "Quantity Measurement, Balanced Growth, and Welfare in Multi–Sector Growth Models," (2018).

Dunn, Abe, Scott D. Grosse, and Samuel H. Zuvekas, "Adjusting Health Expenditures for Inflation: A Review of Measures for Health Services Research in the United States," Health Services Research, 53(1) (2018), 175-196.

Dunn, Abe, Lindsey Rittmueller, and Bryn Whitmire, "Introducing the New BEA Health Care Satellite Account," Survey of Current Business, 95(1) (2015), 1-21.

Dunn, Abe, Bryn Whitmire, Andrea Batch, Lasanthi Fernando, and Lindsey Rittmueller, "High Spending Growth Rates For Key Diseases In 2000–14 Were Driven By Technology And Demographic Factors," Health Affairs, 37(6) (2018), 915-924.

Feenstra, Robert C, "New Product Varieties and the Measurement of International Prices," American Economic Review, 84(1) (1994), 157-177.

Feenstra, Robert C., "Generics and new goods in pharamaceutical price indexes: Comment," The American Economic Review, 87(4) (1997), 760-767.

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

Frank Richard G., E.R. Berndt, A.B. Busch, and A.F. Lehman, "Quality-constant "prices" for the ongoing treatment of schizophrenia: An exploratory study," Quarterly Review of Economics and Finance, 44(3) (2004), 390-409.

Groshen, Erica L., Brian Moyer, Ana Aizcorbe, Ralph Bradley, and David Friedman, "How Government Statistics Adjust for Potential Biases from Quality Change and New Goods in an Age of Digital Technologies: A View from the Trenches," Journal of Economic Perspectives, 31(2) (2017), 187-210.

Grieco, Paul LE, and Ryan C. Mcdevitt, "Productivity and quality in health care: Evidence from the dialysis industry," Review of Economic Studies, 84(3) (2016), 1071-1105.

Griliches, Zvi, and Iain Cockburn, "Generics and New Goods in Pharamceutical Price Indexes," American Economic Review, 84(5) (1994), 1213-1232.

Hall, Anne E, "Adjusting the Measurement of the Output of the Medical Sector for Quality: A Review of the Literature," Medical Care Research and Review, 74(6) (2016), 639-667.

Hall, Robert E., and Charles I. Jones, "The value of life and the rise in health spending," The Quarterly Journal of Economics, 122(1) (2007), 39-72.

Hospital Cost and Utilization Project (HCUP) "HCUP Projections: Acute Myocardial Infarction (AMI) and Acute Stroke 2005 to 2016," (2016) Report # 2016-1.

Howard, David H., Bach, P. B., Berndt, E. R., and Conti, R. M. "Pricing in the market for anticancer drugs," The Journal of Economic Perspectives, 29(1) (2015), 139-162.

Hult, Kristopher J., Sonia Jaffe, and Tomas J. Philipson, "How Does Technological Change Affect Quality-Adjusted Prices in Health Care? Systematic Evidence from Thousands of Innovations," American Journal of Health Economics, 4(4) (2018), 433-453.

Kochanek, Kenneth D., S.L. Murphy, J. Xu, and E. Arias, "Mortality in the United States, 2016," The Centers for Disease Control and Prevention, (2017).

Konüs, Alexander A., "The problem of the true index of the cost of living," Econometrica: Journal of the Econometric Society, (1939), 10-29.

Krumholz, Harlan M., Yun Wang, Jennifer A. Mattera, Yongfei Wang, Lein Fang Han, Melvin J. Ingber, Sheila Roman and Sharon-Lise T. Normand, "An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-Day Mortality Rates Among Patients With an Acute Myocardial Infarction," Circulation 113(13) (2006), 1683-1692.

Krumholz, Harlan M., Yun Wang, Jennifer A. Mattera, Yongfei Wang, Lein Fang Han, Melvin J. Ingber, Sheila Roman and Sharon-Lise T. Normand, "An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-Day Mortality Rates Among Patients With Heart Failure," Circulation, 113(13) (2006), 1693-1701.

Lebow, David E., and Jeremy B. Rudd, "Measurement error in the consumer price index: where do we stand?" Journal of Economic Literature, 41(1) (2003), 159-201.

Lucarelli, Claudio, and Sean Nicholson, "A Quality-Adjusted Price Index for Colorectal Cancer Drugs," NBER Working Paper No. w15174, (2009).

Martin, Anne B., Hartman, B. Washington, A. Catlin, and The National Health Expenditures Account Team, "National health care spending in 2017: Growth slows to post-Great Recession rates; share of GDP stabilizes," Health Affairs, 38(1) (2018), 96-106.

National Research Council, "*At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes,"* ed. (National Academies Press, 2002).

National Research Council, "Accounting for Health and Health Care: Approaches to Measuring the Sources and Costs of Their Improvement," ed. (The National Academies Press, 2010).

National Heart, Lung, and Blood Institute, "Morbidity and Mortality: 2012 Chart Book on Cardiovascular, Lung, and Blood Diseases," National Institutes of Health, (2013).

Neumann, Peter J., Joshua T. Cohen, and Milton C. Weinstein, "Updating Cost-Effectiveness: The Curious Resilience of the $50,000-per-QALY Threshold," The New England Journal of Medicine, 371(9) (2014), 796-797.

Ngai, L. Rachel, and Christopher A. Pissarides, "Structural Change in a Multisector Model of Growth," American Economic Review, 97(1) (2007), 429-443.

The World Health Organization, “A system of health accounts,”OECD (2011).

Pakes, Ariel, “A Reconsideration of Hedonic Price Indexes with an Application to PC’s,” The American Economic Review, 93(5) (2003), 1578-1596.

Pandya, Ankur., S. Sy, S. Cho, M. C. Weinstein, and T. A. Gaziano, “Cost-effectiveness of 10-Year Risk Thresholds for Initiation of Statin Therapy for Primary Prevention of Cardiovascular Disease,” JAMA 314(2) (2015), 142-150.

Petrin, Amil, “Quantifying the Benets of New Products: The Case of the Minivan,” Journal of Political Economy, 110(4) (2002), 705-729.

Redding, Stephen J., and David E. Weinstein, “Measuring Aggregate Price Indexes with Demand Shocks: Theory and Evidence from CES Preferences,” NBER Working Paper No. 22479, 2018.

Roehrig, Charles, “Research Spotlight: A Comparison of BEA and BLS Disease-Based Price Indexes,” Survey of Current Business, September (2017), 1-8.

Romley, John A., Dana P. Goldman, and Neeraj Sood, “US Hospitals Experienced Substantial Productivity Growth During 2002-11,” Health Affairs, 34(3) (2015), 511-58.

Schreyer, Paul, “Toward Measuring the Volume Output of Education and Health Services: A Handbook,” Organization for Economic Co-operation and Development, Working Paper No. 31, 2010.

Sheiner, Louise, and Anna Malinovskaya, “Measuring Productivity in Healthcare: An Analysis of the Literature,” Hutchins Center on Fiscal and Monetary Policy, Working Paper (2016).

Shapiro, Irving, Matthew D. Shapiro, and David Wilcox, "Measuring the Value of Cataract Surgery," In Medical Care Output and Productivity ed (The Universtity of Chicago Press, 2001).

Skinner, Jonathan, and Douglas Staiger, “Technology diffusion and productivity growth in health care,” Review of Economics and Statistics, 97(5) (2015), 951-964.

Syverson, Chad., “Challenges to mismeasurement explanations for the US productivity slowdown,” Journal of Economic Perspectives, 31(2) (2017), 165-86.

Viscusi, W. Kip, “The Value of Risks to Life and Health,” Journal of Economic Literature, 31(4) (1993): 1912-1946.

Appendix 1

Difference between the LE and TE index. The understatement of the TE quality adjustment can be shown in a few ways. First, for the index to be equal to the LE index requires the quality adjustment terms to be the same: $\frac{S_1}{H(\alpha_1 \cdot m_1)}(H(\alpha_1 \cdot m_1) - H(\alpha_0 \cdot m_0)) = \frac{U_H H_m}{U_x}(\alpha_1 \cdot m_1 - \alpha_0 \cdot m_0)$. For small changes in m we have: $H(\alpha_1 \cdot m_1) - H(\alpha_1 \cdot m_1) = H_m(\alpha_1 \cdot m_1 - \alpha_0 \cdot m_0)$. This implies that the indexes are equal when $\frac{S_1}{H(m_1)} = \frac{U_H}{U_x}$. Rearranging the equation and substituting $S_1 = p_1 m_1$ we have the indexes are equal when,

$$p_1 m_1 = \frac{U_H}{U_x} H(\alpha_1 \cdot m_1). \quad (7)$$

Equation (7) implies that the consumer is indifferent between gaining the full health benefit of treatment, $\frac{U_H}{U_x} H(\alpha_1 \cdot m_1)$, and paying for treatment, $p_1 m_1$. In other words, the consumer receives no net benefit from treatment and is indifferent to receiving any medical care. If we expect that consumers receive some value from treatment, then $p_1 m_1 < \frac{U_H}{U_x} H(\alpha_1 \cdot m_1)$.[51]

The equality (7) also contradicts with what we would expect in a typical market. The first order condition (5) implies that the "market value" of the quality change should be measured at the price of purchasing a marginal change in health, $\frac{p_1}{H_m \cdot \alpha_1}$, which is larger than the average cost of producing health. This can be shown by substituting in the first order condition (5), $\frac{U_H}{U_x} = \frac{p_1}{H_m \cdot \alpha_1}$, into equation (6). In this case, the treatment endpoint index is equal to the LE index if

[51] In an idealized market the consumer receives a benefit greater than its price for all units of medical care service, except the last unit of medical care, m_1. That is, if the first order condition (5) holds we should expect $p_1 < \frac{U_H}{U_x} H_m(\alpha_1 \cdot m) \cdot \alpha_1$ for all $m < m_1$. If all infra-marginal units of consumption provide positive welfare, then so should total consumption.

$\frac{p_1 m_1}{H(m_1 \cdot \alpha_1)} = \frac{p_1}{H_m(\alpha_1 \cdot m_1) \cdot \alpha_1}$. This may be re-written as $\frac{p_1}{H(m_1 \cdot \alpha_1)/m_1} = \frac{p_1}{H_m(\alpha_1 \cdot m_1) \cdot \alpha_1}$. These terms are equal if the marginal gain in health, $H_m(\alpha_1 \cdot m_1) \cdot \alpha_1$, is equal to the average gain in health, $\frac{H(m_1 \cdot \alpha_1)}{m_1}$. However, because $H(m)$ is concave, we know that $H_m(m_1 \cdot \alpha_1) \cdot \alpha_1 < \frac{H(m_1 \cdot \alpha_1)}{m_1}$, which shows that the treatment endpoint approach provides a lower bound for the quality-adjustment term. The TE quality adjustment term is only similar to the LE adjustment term when health, $H(m)$, increases linearly with additional medical care inputs, m. In addition, equation (7) suggests that the costs would need to be equal to the benefit.

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, Dalton, and Homes (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 setting 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).[52] 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), AMI, and Heart failure and for the following HCC groupings: Unstable angina, Chronic atherosclerosis, Cardiopulmonary-

[52] Hierarchical Condition Categories (HCC) is a grouping of the over 15, 000 ICD-9-CM codes into 189 clinically coherent groups

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[53] 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.

LE Index Details

Life Expectancy Window

We begin with the assumption that there is a point in time after the acute event, γ, where survival of the event up to that point γ can be attributed to medical care. However, after point in time γ it is determined by other factors such as lifestyle and medical care for other conditions. However, life expectancy 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 life expectancy of those who die before γ and that of those who die after. If we let B_t = the share of patients who die before γ, $LE_{\gamma,t}$ = the life expectancy of patients who die before γ in period t, and $LE_{p,t}|\,\gamma$ = life expectancy of survivors who die post-γ, then:

$$LE_{MC,t} = B_t LE_{\gamma,t} + (1 - B_t) LE_{p,t}|\gamma \quad (10)$$

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

$$\Delta LE_1 = \left[B_1 LE_{\gamma,1} + (1 - B_1) LE_{p,1}|\gamma\right] - \left[B_0 LE_{\gamma,0} + (1 - B_0) LE_{p,0}|\gamma\right] \quad (11)$$

$$\Delta LE_1 = B_1 LE_{\gamma,1} - B_0 LE_{\gamma,0} + (1 - B_1) LE_{p,1}|\gamma - (1 - B_0) LE_{p,0}|\gamma \quad (12)$$

[53] 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.

B_t and $LE_{\gamma,t}$ 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 $LE_{p,t}|\gamma$ constant at its 2001 level. Then:

$$\Delta LE_1 = m_1 LE_{\gamma,1} - m_0 LE_{\gamma,0} + (m_0 - m_1) LE_{p,0}|\gamma(\#\#)$$

Because it is unclear at what point medical care for the event ceases to influence 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 assume 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.[54]

Table A5 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

[54] 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.

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. 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 A5 shows that the bulk of the increases were between 2001 and 2007, with little improvements between 2007 and 2014.

Table A1
Rates of comorbidities

	Acute myocardial infarction	Congestive heart failure	Pneumonia
Race			
White	88.6%	85.9%	88.6%
Black	7.3%	10.3%	7.1%
Asian	1.0%	0.9%	1.1%
Hispanic	1.7%	1.8%	1.8%
Others	1.4%	1.2%	1.5%
Other comorbidities:			
Protein-calorie malnutrition	4.8%	7.3%	10.8%
Dementia	14.3%	17.1%	25.3%
Hemiplegia, paraplegia, paralysis, functional disability	6.2%	7.1%	8.5%
Peripheral vascular disease	30.6%	36.2%	31.2%
Metastatic cancer	3.8%	4.3%	8.4%
Trauma in last year	4.9%	6.6%	7.6%
Major psychiatric disorders	4.4%	5.8%	8.1%
Chronic liver disease	1.1%	2.0%	1.6%
Cerebrovascular disease		29.9%	29.4%
Severe hematological disorders			3.5%
Iron deficiency and other blood disease			54.7%
Depression			21.2%
Parkinson's and Huntington's diseases			4.4%
Seizure disorders and convulsions			6.0%
Fibrosis of lung and other chronic lung disorders			58.5%
Asthma			15.5%
Vertebral fractures			5.2%
Number of observations for each condition	173,277	314,560	340,675

Table A2: The number of events for each condition in each year.

Year	Acute myocardial infarction	Congestive heart failure	Pneumonia
2001	15,839	24,596	27,184
2002	16,224	25,030	29,097
2003	15,942	26,683	30,393
2004	14,953	26,653	27,955
2005	13,703	25,744	30,230
2006	12,753	24,945	26,557
2007	12,066	23,023	24,299
2008	11,719	21,956	24,276
2009	10,699	21,569	21,766
2010	10,830	21,012	21,312
2011	10,099	19,799	21,462
2012	10,164	18,862	20,297
2013	9,539	18,113	19,643
2014	8,747	16,575	16,204
Total	173,277	314,560	340,675

Table A3

Mean total spending per patient

	Days after hospitalization		
	30	60	90
Acute myocardial infarction			
2001	$24,693	$28,593	$31,185
2007	$25,901	$30,159	$33,129
2014	$24,430	$28,322	$30,966
Congestive heart failure			
2001	$14,613	$18,736	$21,864
2007	$16,829	$21,561	$25,227
2014	$17,521	$22,685	$26,479
Pneumonia			
2001	$14,351	$17,725	$20,047
2007	$14,807	$18,570	$21,177
2014	$15,966	$19,986	$22,883

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

Table A4. Risk-adjusted survival rates and life expectancy following hospitalizations

	Number of days after hospitalization		
	30	60	90
Year	Survival rates		
Acute myocardial infarction			
2001	83.8%	79.5%	76.8%
2007	88.7%	85.5%	83.2%
2014	90.1%	87.0%	85.2%
Congestive heart failure			
2001	90.0%	84.6%	80.7%
2007	91.8%	87.1%	83.5%
2014	92.4%	87.1%	83.3%
Pneumonia			
2001	86.2%	80.8%	77.4%
2007	91.1%	86.6%	83.5%
2014	91.7%	87.0%	83.9%

Notes: Survival rates and life expectancy are risk-adjusted as described in Appendix 2.

Table A5

Syntheic life expectancy post-event holding long-term life expectancy constant at its 2001 level, 60-day window

Year	Life expectancy (days) before 60 days	60-day survival rate	Life expectancy (days) conditional on surviving 60 days in 2001	Synthetic life expectancy (days)
Acute myocardial infarction				
2001	13.3	79.5%	1941.9	1547.1
2007	14.5	85.5%	1941.9	1662.0
2014	15.3	87.0%	1941.9	1691.1
Congestive heart failure				
2001	22.2	84.6%	1254.2	1064.6
2007	23.2	87.1%	1254.2	1094.9
2014	24.8	87.1%	1254.2	1095.9
Pneumonia				
2001	19.3	80.8%	1418.7	1150.3
2007	22.0	86.6%	1418.7	1230.8
2014	23.1	87.0%	1418.7	1237.5

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

Table A6. Prices per incremental successful outcome 2001-2014

Year	Mean total 60-day spending per patient	Rate of successful treatment (survival to 60 days without an unplanned readmission)	Assumed success rate of untreated cases	Price per incremental successful treatment
Acute myocardial infarction				
2001	$28,593	59.7%	20.0%	$72,022
2007	$30,159	66.6%	20.0%	$64,760
2014	$28,322	70.1%	20.0%	$56,565
Congestive heart failure				
2001	$18,736	61.4%	20.0%	$45,290
2007	$21,561	64.7%	20.0%	$48,289
2014	$22,685	66.2%	20.0%	$49,144
Pneumonia				
2001	$17,725	63.0%	20.0%	$41,259
2007	$18,570	68.7%	20.0%	$38,170
2014	$19,986	70.1%	20.0%	$39,900

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

Table A7. Hypothetical Adjustment to BLS Multifactor Productivity Estimate for Hospitals and Nursing and Residential Care Facilities (NAICS 622, 623)

	BLS (current)								Alternative Adjusted Productivity		
	Real output	Price indexes	Nominal output	Annual Price Index Growth	New Price Index	Real combined inputs	Productivity	Productivity growth	New price index (rebased)	New real output	New productivity
2001	74.66	79.87	5963.49	1.00	51.80	73.92	1.01	-0.01	1.02	58.30	0.79
2002	79.44	81.90	6506.37	0.99	51.50	78.81	1.01	0.00	1.02	63.98	0.81
2003	81.96	84.32	6911.45	1.00	51.41	81.58	1.00	0.00	1.02	68.08	0.83
2004	84.00	87.16	7321.44	1.00	51.52	82.96	1.01	0.01	1.02	71.97	0.87
2005	88.77	90.01	7989.74	1.00	51.59	88.97	1.00	-0.01	1.02	78.44	0.88
2006	91.27	92.93	8480.89	1.00	51.64	91.27	1.00	0.00	1.02	83.18	0.91
2007	93.82	95.66	8974.81	1.00	51.54	95.03	0.99	-0.01	1.02	88.19	0.93
2008	95.87	98.77	9469.09	1.00	51.59	96.33	1.00	0.01	1.02	92.95	0.96
2009	100.00	100.00	10000.00	0.98	50.64	100.00	1.00	0.00	1.00	100.00	1.00
2010	103.41	101.94	10541.61	0.99	50.05	104.11	0.99	-0.01	0.99	106.66	1.02
2011	106.24	105.00	11154.78	1.00	49.99	107.84	0.99	-0.01	0.99	113.01	1.05
2012	111.13	106.07	11787.47	0.98	48.96	113.74	0.98	-0.01	0.97	121.93	1.07
2013	114.05	108.06	12324.27	0.99	48.36	117.31	0.97	0.00	0.95	129.06	1.10
2014	117.17	110.06	12894.83	0.99	47.75	121.08	0.97	0.00	0.94	136.76	1.13
2015	124.61	110.96	13826.60	0.98	46.68	128.89	0.97	0.00	0.92	150.02	1.16

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 -3.1 percent per year, assuming a 60 day window and a value of a life of $50,000 per year.

Table A8
LE Reservation Price Index for Innovations (VSLY $100k): By Disease Category, Type Of Intervention, Funding Sponsor, and Type of Study

	All			High Score		
Disease Category	Obs.	Mean	Median	Obs	Mean	Median
Cardiovascular	1,577	0.77	0.78	1,149	0.76	0.77
Digestive	479	0.90	0.91	277	0.85	0.90
Endocrine Disorders	700	0.71	0.74	466	0.76	0.82
Infectious	1,792	0.65	0.64	1,234	0.65	0.64
Malignant Neoplasms	1,855	0.83	0.85	1,267	0.85	0.86
Maternal/Child	68	0.74	0.89	32	0.79	0.91
Musculoskeletal/Rheumatologic	804	0.81	0.87	509	0.80	0.88
Neuro-Psychiatric/Neurological	718	0.88	0.90	500	0.92	0.92
Other	1024	0.73	0.73	519	0.77	0.81
Respiratory	278	0.76	0.74	184	0.78	0.75
Sense Organ	158	0.74	0.73	101	0.76	0.69
Intervention						
Care Delivery	342	0.75	0.80	198	0.74	0.82
Device	324	0.68	0.68	202	0.67	0.66
Diagnositc	372	0.85	0.93	253	0.91	0.96
Education	204	0.75	0.76	141	0.78	0.87
Immunization	276	0.68	0.76	195	0.74	0.80
Pharmaceutical	4,951	0.77	0.77	3,474	0.79	0.78
Procedure	1,200	0.75	0.75	746	0.73	0.75
Screening	1,196	0.83	0.96	724	0.84	0.97
Surgical	497	0.72	0.68	268	0.71	0.75
Funding Sponsor						
Foundation	679	0.72	0.86	463	0.74	0.88
Government	2,702	0.79	0.87	1,761	0.81	0.88
Health Care Organization	442	0.85	0.86	317	0.83	0.86
Other	2,525	0.76	0.76	1,609	0.78	0.80
Pharma or Device Manuf.	2,969	0.75	0.75	2,035	0.75	0.76
Prof Member Organization	136	0.71	0.66	53	0.67	0.62
Type of Study						
Meta-Analysis	565	0.82	0.82	415	0.84	0.85
Other	2,112	0.83	0.85	1,146	0.84	0.88
Randomized	3,147	0.76	0.76	2,115	0.77	0.77
Simulation	3,629	0.73	0.80	2,562	0.74	0.81

Notes: The indexes are reported along two categorical dimensions in this table: disease chapter of the illness being treated and the type of innovation being tested in the study. The reviewers of the medical studies that enter the studies in the CEA database score the quality of the research on various dimensions. An overall rating is included in the database indicating the quality of the study. Following Hult et al. we report overall estimates and estimates based only on those studies with a rating at or above the median. The indexes at the bottom and top 1 percent of the distribution have been removed for the construction of this table.

Table A9
TE Index for Innovations: By Disease Category, By Type Of Intervention, Funding Sponsor, and Type of Study

	All			High Score		
Disease Category	Obs.	Mean	Median	Obs	Mean	Median
Cardiovascular	1,592	1.29	1.03	1,156	1.23	1.03
Digestive	487	1.35	1.06	280	1.26	1.04
Endocrine Disorders	702	1.20	1.02	466	1.21	1.03
Infectious	1,788	1.37	1.08	1,229	1.36	1.08
Malignant Neoplasms	1,859	1.44	1.09	1,273	1.49	1.10
Maternal/Child	64	1.26	1.00	29	1.12	1.00
Musculoskeletal/Rheumatologic	809	1.30	1.03	518	1.33	1.03
Neuro-Psychiatric/Neurological	721	1.24	1.01	502	1.17	1.02
Other	1005	1.45	1.03	521	1.40	1.04
Respiratory	273	1.44	1.08	180	1.53	1.08
Sense Organ	155	1.57	1.10	100	1.66	1.13
Intervention						
Care Delivery	340	1.26	1.01	198	1.13	1.02
Device	328	1.32	1.04	204	1.26	1.04
Diagnositc	379	1.16	1.01	256	1.16	1.01
Education	196	1.27	1.02	138	1.24	1.03
Immunization	272	1.62	1.05	192	1.38	1.04
Pharmaceutical	4,975	1.39	1.05	3,493	1.35	1.05
Procedure	1,203	1.34	1.06	755	1.35	1.06
Screening	1,174	1.28	1.04	712	1.40	1.06
Surgical	499	1.29	1.06	270	1.37	1.10
Funding Sponsor						
Foundation	682	1.41	1.06	466	1.41	1.08
Government	2,720	1.38	1.06	1,772	1.40	1.06
Health Care Organization	431	1.37	1.06	306	1.41	1.05
Other	2,519	1.44	1.06	1,614	1.42	1.07
Pharma or Device Manuf.	2,970	1.24	1.02	2,046	1.20	1.02
Prof Member Organization	133	1.40	1.12	50	1.42	1.02
Type of Study						
Meta-Analysis	569	1.28	1.03	420	1.26	1.03
Other	2,119	1.37	1.05	1,155	1.38	1.04
Randomized	3,125	1.31	1.04	2,105	1.30	1.05
Simulation	3,642	1.39	1.05	2,574	1.37	1.05

Notes: The indexes are reported along two categorical dimensions in this table: disease chapter of the illness being treated and the type of innovation being tested in the study. The reviewers of the medical studies that enter the studies in the CEA database score the quality of the research on various dimensions. An overall rating is included in the database indicating the quality of the study. Following Hult et al. we report overall estimates and estimates based only on those studies with a rating at or above the median. The indexes at the bottom and top 1 percent of the distribution have been removed for the construction of this table.

Table A10. Process Measures of Quality from Hospital Compare

	Percent of patients given the following recommended treatment						
Process Measure for Patients Given:	2004	2005	2006	2007	2008	2009	% Increase
Condition: Heart Attack							
ACE Inhibitor or ARB for Left Ventricular Systolic Dysfunction (LVSD)	82	83	87	91	94	96	16.8%
Aspirin at Arrival	94	95	97	97	98	98	4.2%
Aspirin at Discharge	94	96	97	97	98	98	4.2%
Beta Blocker at Discharge	93	95	96	97	98	98	6.0%
Smoking Cessation Advice/Counseling	87	92	97	98	99	99	14.9%
Condition: Heart Failure							
ACE Inhibitor or ARB for Left Ventricular Systolic Dysfunction (LVSD)	81	83	86	90	92	94	16.2%
Assessment of Left Ventricular Function (LVF)	87	90	93	94	96	98	11.7%
Discharge Instructions	52	58	71	76	82	88	68.9%
Smoking Cessation Advice/Counseling	74	83	92	95	97	98	33.5%
Condition: Pneumonia							
Patients Assessed and Given Pneumococcal Vaccination	52	62	78	83	88	93	78.0%
Initial Antibiotic(s) within 4 Hours After Arrival	72	75	81	93	94	95	31.1%
Smoking Cessation Advice/Counseling	70	79	90	92	95	97	38.5%
The Most Appropriate Initial Antibiotic(s)	77	80	87	89	89	91	18.3%
Blood Culture Performed Prior to First Antibiotic Received in Hospital	82	83	90	91	93	95	15.7%
Surgical Infection Prevention							
Received Preventative Antibiotic(s) One Hour Before Incision	77	81	86	89	93	96	24.7%
Preventative Antibiotic(s) are Stopped Within 24 hours After Surgery	64	69	78	84	90	94	45.7%

Notes: The estimates from this table come from authors calculation from the Hospital Compare database archives from the Center for Medicare and Medicaid services (https://data.medicare.gov/data/archives/hospital-compare). The estimates are based on a simple weighted averages across all hospitals in the database where the weight is determined by the sample size at each hospital. Quality measures that were discontinued or continued in the middle of the sample range are not shown. The year reported in this table is based on the year the information was gathered from the hospital, which is typically lagged one year in the database. For instance, the process measures for 2004 are from the 2005 hospital compare database.

Appendix 3

Additional Theoretical Discussion

A simple model for comparing across methods for discrete technologies: To further 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 differentiated treatments (T_1 and T_2) and has an endpoint that delivers a mean value of *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 mean probability of π_i.
- The proportion of patients in period *t* receiving T_1 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 π_3. While π_3 in this model represents the case without medical care, in practice everyone receives medical care.
- $C_{1t} > C_{2t}$ in each period *t* and $\pi_1 > \pi_2 > \pi_3$. T_1 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 $R_{it} = C_{it} * m_t$ 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 $\Delta q = q_1 - q_0$.
- 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 $\frac{S_1-\Delta q*(\pi_1-\pi_2)*B}{S_0} = UI - \frac{\Delta q*(\pi_1-\pi_2)*B}{S_0}$.

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 $\frac{\frac{S_1}{q_1\pi_1+(1-q_1)\pi_2-\pi_3}}{\frac{S_0}{q_0\pi_1+(1-q_0)\pi_2-\pi_3}} = UI * \frac{q_0\pi_1+(1-q_0)\pi_2-\pi_3}{q_1\pi_1+(1-q_1)\pi_2-\pi_3}$.

Hedonic 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 $\sqrt{\frac{q_0R_{11}+(1-q_0)R_{21}}{q_0R_{10}+(1-q_0)R_{20}} * \frac{q_1R_{11}+(1-q_1)R_{21}}{q_1R_{10}+(1-q_1)R_{20}}} = \sqrt{UI * \frac{q_0R_{11}+(1-q_0)R_{21}}{q_1R_{10}+(1-q_1)R_{20}}}$.

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:

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

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: $\frac{S_1-\Delta q*(C_{11}m_1-C_{21}m_1)}{S_0} = UI - \frac{\Delta q*(C_{11}m_1-C_{21}m_1)}{S_0}$. 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: $UI - \frac{\Delta q*(R_{11}-R_{21})}{S_0}$ (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. We explore how the other indexes perform relative to the LE index under alternative scenarios:

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, 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 hedonic 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 hedonic 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.

If we set the LE and TE indexes equal, for example, and solve the value of the change in quality $\Delta q * (\pi_1 - \pi_2) * B$, 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. This equality holds when consumers are indifferent between receiving health benefits or paying medical care expenditures. In a more realistic scenario, consumers actually receive some net benefit from treatment, so we should expect $B > \frac{S_1}{q_1\pi_1+(1-q_1)\pi_2-\pi_3}$. Specifically, one could think of consumers as sorted across treatments based on their preferences and the preferences of their doctors, which may be viewed as random. Under this scenario, all the inframarginal consumers receive a benefit from treatment and only the marginal consumer pays an amount equal to her benefit.

Similarly, if we set the LE and hedonic indexes equal, we find they are equal when $\Delta q * (\pi_1 - \pi_2) * B = S_0 * (UI - hedonic) =$ *S_0*(%Δspending - %Δquality-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 $(\pi_1 - \pi_2) * B = R_{11} - R_{21}$ 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 $\frac{S_1 - B_2}{S_0}$ and the QALY index can be calculated as $\frac{\frac{S_1}{B_2}}{\frac{S_0}{q_0\pi_1 + (1-q_0)\pi_2 - \pi_3}}$ because monetized QALYs are a universal metric that can be used to compare the values of all treatments. However, constructing the other indexes require treatments to be comparable across periods. The TE and hedonic indexes cannot be calculated without identical endpoints or treatment baskets across periods. The RC index is challenging to calculate as well because $S_1 - S_0 = R_{31} - (q_0 R_{10} + (1 - q_0) R_{20})$, 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 hedonic and RC indexes, however, is that they can be constructed without knowing B or observing outcomes, information which is often unknown to the economist constructing the index. They do, however, require creating treatment baskets or characteristics which cannot be computed without medical expertise.