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

by

Seidu Dauda, World Bank Group, (email: sdauda@worldbank.org)

Abe Dunn, Bureau of Economic Analysis, (email: abe.dunn@bea.gov)

Anne Hall, U.S. Department of the Treasury, (email: anne.hall@treasury.gov)

This is a preliminary draft. Please do not cite or distribute without permission of the authors.

Abstract: Health care spending has grown rapidly over the past several decades. However, whether this growth rate may be considered too high largely depends on the value of the increased length and quality of life due to medical treatment. We address this question within the framework of price measurement. Specifically, we construct quality-adjusted price indexes for three acute medical conditions for the period 2001-2011 using data for Medicare fee-for-service (FFS) beneficiaries. We find that quality adjustment has a large effect on price growth and that a substantial bias exists if one does not adjust for quality changes. As measured by a benchmark cost-of-living index (COLI), the average prices for these conditions are declining steeply. For each condition we also compare indexes created with alternative methods with a view toward exploring how the methods could be implemented in practice. We find a wide dispersion in the growth rates of different quality-adjusted price indexes. According to our theoretical model, this dispersion results from the increases in benefits exceeding the increases in spending for these conditions; only the benchmark COLI fully captures the value of the increases in benefits to patients.

1. Introduction

The value of health-care spending in the United States has been the topic of much concern among policymakers and the public for some time now. Health care has risen as a percent of GDP from 5 percent in 1960 to 17.8 percent in 2015 (Martin et al. 2016). While this rate of growth has leveled off in recent years, interest in quantifying the value of health-care spending continues to be quite high especially as the US is still an outlier among advanced economies in terms of how much it spends on health care per capita.

A major measurement challenge for health care is accounting for the quality of the services provided. The improvements in health outcomes resulting from technological advances in medicine should clearly be reflected by an increase in the measured real value of output. Given nominal spending, it is necessary to construct price indexes that appropriately account for this quality change, so that output for the treatment is measured accurately.

We expect the effect of this measurement problem may be substantial. Dunn et al. (2015) find that measuring output with the cost of treatments implies that 73 percent of the growth in per capita health-care spending relative to the economy-wide PCE deflator from 2000 to 2010 comes from growth in cost per case and only 27 percent is from growth in the number of cases treated.

Much of the growth in cost per case likely comes from technological change as new treatments tend to be more expensive than older ones.

New treatments are also often more effective than older ones so it is important to capture the value of greater effectiveness when measuring real spending in health care. There are a number of important innovations in the health care sector in recent years that have been shown to improve health outcomes and these improvements are not currently reflected in national statistics. Sovaldi and other new drugs to treat chronic hepatitis C, for example, are more effective than previous treatments, but they are also quite expensive, costing between $60,000-$100,000 for treatment.[1] New drugs to treat high cholesterol were quite costly when they entered the market, but they were also highly effective at reducing cardiovascular disease, the leading cause of death in the United States (Dunn (2012)). The latest drug treatments for rheumatoid arthritis, a disease causing pain, stiffness and loss of function in the joints, are effective at reducing pain and improving mobility, but are priced at over $30,000 per year. In addition, effective non-drug treatments for this condition, such as joint replacements, also entail substantial costs and their use has increased rapidly in recent years.[2]

In each of these cases, the improvements in the quality of life and life expectancy for these treatments are not captured in national health care price indexes. Health experts believe that the impact of these quality changes are reflected in national trends in mortality rates and rates of disability (Cutler et al. (2006) and Cutler et al. (2016)), although it is challenging to precisely separate the contribution of medical care from that of other non-medical factors for the entire population. In general, economists view health care as an area of rapid innovation where the

[1] http://www.healthline.com/health/hepatitis-c-treatment-cost#1
[2] https://www.hcup-us.ahrq.gov/reports/statbriefs/sb186-Operating-Room-Procedures-United-States-2012.jsp

potential bias in traditional measure of inflation may be large (see Lebow and Rudd (2003) and Groshen et al. (2017)).

While the field of health economics is generally devoted to measuring the value of health care against its costs, relatively little work has been done that expresses that comparison in terms of quality-adjusted price indexes for medical conditions at the national level. Hall (2016) and Sheiner and Malinovskaya (2016) review the methods for quality adjustment of medical prices that have been used in the research literature and find there is no widely accepted method for creating quality-adjusted price indexes for medical care. The goal of this paper therefore is to create a better understanding of the theoretical and practical difference across alternative methods of quality adjustment for health care price indexes that have been applied in the literature.

We first discuss potential theoretical differences between the methods used in the literature with a stylized model of health care quality change and spending. We consider four methods: a cost-of-living index with a quality adjustment based on the value of extended life (Cutler et al. 1998), a quality-constant price indexes based on the methods set out in Berndt et al. (2002), a fixed-technology index based on Frank et al. (2004), and an index with the quality adjustment based on the change in costs similar to the quality-adjusted producer price index constructed by the Bureau of Labor Statistics (BLS). The theoretical comparison shows that, if the value of the improvement in treatment of the condition is close to the increased spending on treatment, the quality-adjusted indexes deliver similar results. If the increase in value greatly departs from the increase in spending, however, the cost-of-living index constructed with the method of Cutler et al. (1998) gives very different results from the other indexes since it is the only index to incorporate the full value of the consumer's increased benefit into its formula. We also show that

this COLI is robust to the most violations of assumptions of the model and gives the correct results in the widest array of situations.

We then apply the index formulas to the creation of quality-adjusted or quality-constant price indexes for three acute conditions among fee-for-service (FFS) Medicare patients for the years 2001-2011: acute myocardial infarction or heart attack (AMI), congestive heart failure (CHF), and pneumonia. These conditions all saw significant improvements in outcomes as measured by post-hospitalization life expectancy during this period. By measuring mortality immediately following an acute episode, we are able to attribute changes in observed health outcomes to medical care conditional on observable risk factors.

The three unadjusted condition-based indexes rise by similar amounts above general inflation for each of the three conditions, having an excess growth rate of around 2 percent per year. Not surprisingly, given the improvements in outcomes, the quality-adjusted indexes all give lower growth rates in prices with the COLIs showing steep declines. Applying the COLI quality-adjustment method, and using the typical value of a life year used in the literature of $100,000 we find that annual medical price inflation declines by -2.7 percent relative to aggregate inflation rates, using an average of select conditions studied here. We regard this index as the benchmark because, as we will show, it is robust to violations of many assumptions underlying the formulas of the other indexes. Therefore quality-adjusted medical prices still appear to be declining for the subset of conditions studied here. The divergence between the COLI and the other quality-adjusted indexes is caused by the benefits of the increases in life expectancies following hospitalizations exceeding the increase in spending. For AMI and pneumonia, for example, the increased benefits are worth more than three times as much as the increase in spending per patient. Despite the sensitivity of all of the indexes to key assumptions, we find that even under

the most conservative assumptions, the quality-adjusted indexes grow at about 1 percentage points less per year than the unadjusted indexes.

Our results suggest a few guidelines for quality adjustment that may be helpful for national accounting purposes. First, one important feature of quality adjustment is that the adjustment should be tied to some increase in quality that has an actual impact on health or is somehow valued by the consumer. The COLI explicitly incorporates the consumer's value of a quality improvement, while the other approaches do not necessarily share this feature. We suggest that even when implementing a non-COLI measure, it is important to tie an adjustment to the health benefits of the consumer. Another feature is that researchers should be able to calculate the index in a relatively timely fashion. The COLI applied by Cutler et al. (1998) requires a number of years of post-treatment data to evaluate the impact on life expectancy. Here we implement an alternative version of a COLI that may be calculated with more limited post-treatment data. Finally, there should be transparency regarding the quality adjustment and its impact on the inflation rate. In this paper we show how quality-adjusted price indexes may be sensitive to the assumptions made by the researchers. Therefore, researchers should explicitly show how the quality change impacts inflation rates and perhaps report inflation rates under alternative assumptions.

Our paper has several limitations. First, our results are based solely on data for elderly Medicare fee-for-service beneficiaries. Most of these methods require data on treatments, spending, and outcomes and we do not have any data tying all of those together for individuals for other populations. Second, we measure outcomes solely with mortality as we have no information in the claims data on quality of life. A third limitation of this paper is that the improvements in outcomes observed for these conditions are not necessarily representative of the medical sector

as a whole. For other conditions where we do not witness such substantial improvements in outcomes, such as Alzheimer's disease, quality adjustment is potentially not as important. However, until we conduct analysis of quality adjustment for other condition categories, we will not know the true inflation rate for the health care sector.

1. Background on quality-adjusted price indexes in health care

In this section we give some background on the state of research and current practice in quality-adjusted price indexes in health care and set out some specifications for what an ideal quality-adjusted price index for health care would be. First, throughout this paper we will be discussing the quality adjustment of medical price indexes based on the costs of a specific condition. Previous research on quality adjustment of price indexes in medical care has all been performed on condition-based indexes (Cutler et al. 1998, Berndt et al. 2002, Frank et al. 2004). The BEA has already taken the initial steps to redefine output in health care with the release of its Health Care Satellite Account (HCSA) (Dunn et al. 2015). The HCSA offers an alternative measure of inflation in health care where the price is measured by the average yearly spending on a specific condition, rather than the prices of specific goods and services (as is the current practice), thus allowing for treatment patterns to shift flexibly. The change in real output for a condition is calculated as the change in total expenditure for that condition divided by the price of treating that condition.

The Health Care Satellite Account does not, at this point, address the second major measurement challenge in health care, that of adjusting for quality. However, creating such indexes is necessary to fit health care into the framework of the national statistics and to better measure output and productivity in the health care sector. Many analysts and policymakers would like to

make valid comparisons between health care and other industries in terms of inflation, growth of real output, and productivity, but in order to do that, statistics for health care must be estimated in the same way.

We view the ideal quality-adjusted index as a conditional cost-of-living index (COLI), in accordance with the guidelines laid out in "At What Price?" (National Research Council 2002). A conditional COLI measures the change in the cost of living from the perspective of the consumer while utility and the environment are held constant. There are other ways to value quality and we will also briefly discuss an alternative. In the general price index literature, two primary concepts have been applied to create quality adjustments (Triplett 1982). The first approach is based on the monetized value of the quality change to consumers or "user value" and this approach is the basis for a conditional cost-to-living index. As discussed below, however, quality can also be measured from the perspective of the producer or in terms of "resource costs."

Measuring quality in terms of user value is likely preferable, however, because the BEA uses the "final expenditures" method for measuring GDP (Landefeld et al. 2008). In this approach, GDP is measured indirectly by measuring the value of consumer, government, and business spending and then subtracting out imports. Health care is a component of personal consumption expenditures (PCE) or consumer spending. Our goal therefore is to appropriately deflate total nominal consumer spending on health care to measure the real value of consumer spending. The imports within this spending (which probably consist largely of prescription drugs) would be subtracted out before total GDP is calculated but that step is not considered in this paper.

In their surveys of the research literature in this area, both Hall (2016) and Sheiner and Malinovskaya (2016) find that there is currently no standard method for quality adjustment in health care. Hall (2016) notes that quality adjustments in the research literature have been based on either observed medical outcomes or on the imputed medical value of the treatments observed being given. When measuring aggregate outcomes, the challenge is to separate the effects of medical care (which should be included in the quality adjustment) from the effects of other factors such as behavior, risk factors and demographics, since those are part of the environment that is to be held constant. Because of this issue, many economics papers in this literature choose to measure quality based on observed short-term mortality outcomes of acute illnesses, both because those outcomes are more likely to be observed in data and because measuring them is relatively straightforward without medical expertise. We follow this approach in this paper as we calculate quality-adjusted price indexes for three acute high-mortality inpatient illnesses among Medicare beneficiaries based on short-term mortality outcomes during or after the hospitalization.

Cutler et al. (1998) proposed a method for creating cost-of-living indexes for medical conditions by adjusting prices with the value of life extended by medical treatment, while holding constant environmental factors that also affect health outcomes such as demographics. They found that increases in this value exceeded increased costs by a considerable margin for AMI (acute myocardial infarction) patients. Their results show that the value of life extended by improved AMI care was around $17,000 per patient between 1984 and 1991 but that costs for those patients only increased by about $3,200 per patient. Life expectancy for AMI patients post-AMI rose by four months relative to non-AMI patients during this period and even, at conservative estimates of the value of life, that increase represents a considerable amount of value.

The other method cited by Triplett (1982) for adjusting for quality in price indexes is to adjust for the "resource cost" of the producer (e.g., hospital or physician) necessary to produce the change in quality. This resource cost approach of adjusting for quality is typically applied in the Producer Price Index (PPI) by the BLS.Triplett argues that this approach is the theoretically preferred measure when adjusting for quality for output price indexes. While the resource-cost approach is applied to many goods when constructing the PPI index, one technical challenge is that we must make strong assumptions regarding the cost of inputs and their relationship to quality in the absence of detailed cost data.

Triplett (1982) notes that there are other problems with the resource-cost approach. First, it implies that quality cannot improve without costs going up. He cites the counterexample of birth control pills where reducing the dose by half lowered costs while delivering the same value to the consumer. Furthermore, in health care, quality can rise even if the price decreases, for example if low-value treatments and diagnostic tests are reduced.

No price index has yet been constructed in the research literature for health care by adjusting for resource costs. However, the Bureau of Labor Statistics (BLS) in general constructs quality adjustments in the Producer Price Index with this method and does calculate a cost-based quality adjustment in the hospital Producer Price Index (PPI) based on quality indicators for hospitals collected by the Department for Health and Human Services. The index has some limitations, however. First, the BLS calculates a composite quality score index for each condition where the weights for each quality score are based on how many people the score applies to. As they note themselves, it would be preferable to weight the scores by medical importance but data for that is not easily available to them. The percent change in the composite score is then multiplied by an estimate of costs formed from the median charges for the Diagnosis-Related Group (DRG) and

the cost-to-charge ratio for operating and capital costs. Again, as they note, it would be preferable to use more direct proxies for costs of each quality score but they were unable to identify ones for each score. Their approach illustrates the potential difficulties of constructing a quality-adjusted price index in health care based on changes in resource costs.

Another approach is to redefine the price units by focusing on the price per *successful* treatment. Berndt et al. (2002) construct a price index for major depression by identifying a treatment endpoint (remission) and then measure the incremental price of the probability of reaching that endpoint relative to no treatment at all. This is an intuitively appealing price index that holds quality constant by measuring price per successful treatment. Frank et al. (2004) construct an index for schizophrenia by holding the market shares of key treatments or technologies fixed. Holding treatments fixed should hold quality constant in theory, assuming the treatment baskets are constant in quality. The remainder of the paper is devoted to discussing all of these methods in more detail both in theory and in practice.

2. Alternative Quality-Adjustment Methods

In this section, we outline a simple model to illustrate the differences among quality-adjusted, quality-constant, and resource-cost-based indexes. From the consumer's perspective, the theoretical ideal for a price index is a COLI conditional on the environment and population. As is well known, a COLI is written as $\frac{e(p_1,U_0)}{e(p_0,U_0)}$ where $e(\cdot)$ is the expenditure function that expresses the minimum expenditure to achieve a certain level of utility given a certain set of prices p. It is the ratio of expenditure needed at new prices to maintain base-period utility to the expenditure in the base period given base-period utility and prices in both periods.

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

$$U(H_1(m_1), Y - p_1 m_1 - CV) = U(H_0(m_0), Y - p_0 m_0) \quad (2)$$

In this equation, m_t is the quantity of medical care, p_t is the price of medical care and $H_t(\cdot)$ translates medical care into health H. Taking a first-order Taylor approximation at period 0 yields:

$$CV = \frac{U_H H_m}{U_x}(m_1 - m_0) - (p_1 m_1 - p_0 m_0) \quad (3)$$

where U_H is the marginal utility of health, H_m is the marginal effect on health of medical care and U_x is the marginal utility of non-health consumption ($x = Y - p_0 m_0$).

Cutler et al. (1998) note that the first term in this equation is the increased benefit in monetary terms of medical care for the condition being treated ($\Delta B = B_1 - B_0$) and the second term is the change in spending on the condition ($\Delta S = S_1 - S_0 = p_1 m_1 - p_0 m_0$) where $S_t = p_t m_t$. The COLI for a patient with the condition is therefore:

$$COLI_{pat} = \frac{Y-(\Delta B - \Delta S)}{Y} \quad (4)$$

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

$$COLI_{pop} = \frac{Y - r(\Delta B - \Delta S)}{Y} \quad (5)$$

From this expression they then derive the quality-adjusted price index for the individual condition. Let α be the per capita income share of spending on the condition in the initial period: $\alpha = \frac{rS_0}{Y}$. Then the change in quality-adjusted prices for the condition multiplied by its income share will be equal to the change in the COLI, holding all other prices constant:

$$\alpha\left(\frac{\widehat{S_1} - S_0}{S_0}\right) = \frac{-r(\Delta B - \Delta S)}{Y}$$

Solving for $\widehat{S_1}$, we obtain $\hat{S}_1 = S_1 - \Delta B$. So the quality-adjusted price index across the population is equal to:

$$\frac{S_1 - \Delta B}{S_0} \quad (6)$$

This formula in fact corresponds to the formula for a direct quality adjustment in the CPI by the BLS (Bureau of Labor Statistics 2015). If we can obtain an accurate measure of the additional quality-adjusted life expectancy added by medical treatment, it can serve as the direct quality adjustment.

Rather than adjusting for quality, an alternative is to hold quality constant. Berndt et al. (2002) define the output as the number of successful treatments rather than as the number of treatments. Then the price to be measured is the incremental price of achieving a certain endpoint or goal of treatment relative to receiving no treatment at all. This achievement can be measured either prospectively based on the medical value of treatments given and calculating the probability of achieving the endpoint based on them or retrospectively based on whether the endpoint was

observed being achieved. Berndt et al. (2002) take the former approach when they calculate a quality-constant price index for major depression by measuring the price per probabilistic incremental remission based on expert opinion of the treatments being given. Romley et al. (2015) take the latter approach when they measure the output of hospitals by measuring how many patients meet the treatment goals of certain acute illnesses by surviving for a specified period without an unplanned readmission.

Frank et al. (2004) create a price index for schizophrenia by defining treatment baskets for schizophrenia based on treatment guidelines issued by medical experts. They place patients into those treatment baskets and then measure the changes in prices of the treatment baskets while holding the share of patients receiving each basket constant over time. Rather than holding quality constant directly, this approach holds constant the relevant technologies or treatments. Quality is thus held constant on the assumption that a treatment basket is constant in quality over time. They then calculate the index as a Fisher index which is the geometric average of a Laspeyres index where the shares of patients receiving each treatment are held constant at the level of the initial period while the prices of the treatment baskets change each year and a Paasche index where the shares are held at the level of the final period.

Finally, in contrast to indexes based on the quality of the treatment to the patient or indexes that hold quality constant, the producer price index that is typically used in measuring output is constructed based on the producer's problem (Fisher and Shell 1972) of how to set prices that maximize revenue given a fixed set of inputs. If quality of the output changes, this implies that the inputs have changed, and a quality adjustment is necessary. Therefore, to hold inputs fixed in a producer price index, it is recommended that practitioners adjust for the resource cost of those inputs by subtracting the change in costs from the second-period price (Triplett 1982). The BLS

uses this method when constructing quality-adjusted Producer Price Indexes (PPIs) (Bureau of Labor Statistics 2014).

A simple model for comparing across methods To compare these methods for creating quality-adjusted or quality-constant price indexes for medical care, consider the following simple model for a condition that has two treatments T_1 and T_2 and has an endpoint that delivers B of benefit as valued in dollars.

- T_i has cost C_{it} in period t and patients receiving T_i reach the endpoint with a 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 represents the case where no medical care is received, in this model, everyone receives medical care in practice.
- $C_{1t} > C_{2t}$ 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 equal to $S_t = q_t R_{1t} + (1 - q_t) R_{2t}$.
- 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 $\frac{S_1 - \Delta q * (\pi_1 - \pi_2) * B}{S_0} = UI - \frac{\Delta q * (\pi_1 - \pi_2) * B}{S_0}$. Since B is measured with life expectancy in Cutler et al. (1998), we will call this type of index the LE index.

Treatment endpoint (TE) index: A constant-quality index that measures the relative change in price of meeting the treatment endpoint, such as that created by Berndt et al. (2002), will be written $\frac{\frac{S_1}{q_1\pi_1 + (1-q_1)\pi_2 - \pi_3}}{\frac{S_0}{q_0\pi_1 + (1-q_0)\pi_2 - \pi_3}} = UI * \frac{q_0\pi_1 + (1-q_0)\pi_2 - \pi_3}{q_1\pi_1 + (1-q_1)\pi_2 - \pi_3}$. We will call this type of index the TE index.

Basket price (BP) index: A constant-technology index that measures the changes in the prices of treatment baskets and aggregates these prices holding the shares receiving the treatment or technology constant using a Fisher index formula, such as that created by Frank et al. (2004),

will be written $\sqrt{\frac{q_0R_{11}+(1-q_0)R_{21}}{q_0R_{10}+(1-q_0)R_{20}} * \frac{q_1R_{11}+(1-q_1)R_{21}}{q_1R_{10}+(1-q_1)R_{20}}} = \sqrt{UI * \frac{q_0R_{11}+(1-q_0)R_{21}}{q_1R_{10}+(1-q_1)R_{20}}}$. We will call this type of index the BP index.

Resource cost (RC) index: An index based on the change in costs coming 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 last two terms measure the changes in the reimbursements of the same treatments over time and therefore capture pure inflation. 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, 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).

Next we examine how the different indexes may deviate from a COLI estimate of a quality change. By construction, the LE index is a COLI which gives an accurate measure of price changes from the perspective of the consumer under all conditions and we regard it as the benchmark index. We then explore how the other indexes perform relative to the LE index under alternative scenarios where the assumptions of the model given above are violated:

1. If $q_1 = q_0$, there are no changes in treatment patterns and therefore no need for quality adjustment. In that case, all four indexes are appropriately equal to the unadjusted index.

2. If B = 0, that is, if achieving the treatment endpoint does not deliver any benefit at all in reality, the LE index will be appropriately equal to the unadjusted index but the other three indexes will not. The TE index, for example, will still measure the changes in the price of achieving the treatment endpoint whether or not achieving that endpoint has any meaning. It is essential therefore when constructing this type of index to choose a treatment endpoint that is medically meaningful.
3. If $\pi_1 = \pi_2$, that is, if both treatments are equally effective and there is therefore no actual change in quality, the LE index and the TE index are both appropriately equal to the unadjusted index. The BP and RC indexes, however, will differ from the unadjusted index. This reflects a weakness of these indexes, that whether or not they are meaningful depends on whether the shifts in qHowever, it is questionable whether shifts to newer, more expensive treatments or increases in intensity of treatment always reflect actual differences in efficacy in health care.
4. If both treatments cost the same in both periods but $q_1 \neq q_0$, so there is quality change but no change in spending other than general inflation, the BP and RC indexes are inappropriately equal to the unadjusted index. These indexes assume quality changes are only reflected in changes in spending. However, as noted above, quality in health care can improve (decline) without increases (decreases) in spending.

In general, the other indexes approximate the LE index most closely when the value of the changes in quality lines up with the changes in spending.

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.

Similarly, if we set the LE and BP indexes equal, we find they are equal when $\Delta q * (\pi_1 - \pi_2) * B = S_0 * (UI - BP) = S_0$*(%Δ*spending* - %Δ*quality-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.

In standard economic theory and in specialized conditions, the difference between the monetized value of quality change and the change in the prices or in the costs of production deriving from the quality change is zero in equilibrium. However, in sectors with rapid product innovation and new goods, such as health care, the benefits received by consumers for new goods will likely be higher than the cost of production. Even when the benefits to the marginal consumer equal the marginal cost, there may be substantial surplus from new goods for inframarginal consumers whose benefit of a new product greatly exceed its cost (Pakes (2003) and Redding and Weinstein (2017)).[3]

[3] An alternative way to view this problem, discussed in Sheiner and Malinovskaya (2016), is to considers consumers of medical care as being at a corner solution prior to the introduction of a new technology. That is, they purchase all the health that is possible under existing technologies and their budget constraint, but the price of purchasing additional health could be prohibitively high. When new effective technologies enter, the price of additional health falls. This technology shift implies a jump in the level of utility, which does not necessarily relate to the price or cost of the new technology.

To further complicate issues, there are additional well-known market frictions in health care that can complicate the analysis. Specifically, there are moral hazard problems, information asymmetries, and potential agency problems with providers, which could make the price of a good or service deviate from its marginal value to consumers. The value of $\Delta q * (\pi_1 - \pi_2) * B$ has the potential to be much higher or much lower than the change in spending. For conditions such as cardiovascular disease where medical innovation has delivered a lot of value, as measured by increases in life expectancy, this value is likely much higher than the increases in spending from the diffusion of bypass surgery, angioplasty, or antihypertensive and antihyperlipidemic medication (Cutler et al. 2006). Yet others have found a significant amount of overuse of services of no medical value (Colla et al. 2015). In those cases, the quality adjusted price could be raised by lowering spending, which would cause a serious divergence between the LE and TE indexes and the other two indexes.

There are other advantages of the LE index. As discussed by Hall (2016), the LE index is better at incorporating highly innovative new medical treatments. Let us hypothesize a medical innovation with a new treatment endpoint that delivers $B_2 > B$ in monetized 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 $\frac{S_1 - B_2}{S_0}$ since monetized QALYs are a universal metric that can be used to compare the values of all treatments. However, the other three indexes all require treatments to be comparable across periods to be constructed. The TE and BP indexes cannot be calculated without identical endpoints or treatment baskets across periods. For the TE index, it is important to choose an outcome metric that is common across all technologies. The RC index is challenging to calculate as well since

$S_1 - S_0 = R_{31} - (q_0 R_{10} + (1 - q_0) R_{20})$ 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 BP and RC indexes, however, is that they can be constructed without knowing B or observing outcomes. That information is often unknown to the economist constructing the index. They do, however, require creating treatment baskets so cannot be computed without any medical expertise at all.

In the next section, we will construct all four of these indexes for three acute illnesses where outcomes have improved in recent years and show how and why the indexes diverge in value from each other.

1. Data and methods

For this paper, our goal is to construct all four of the indexes described above with quality based on observed mortality outcomes. To do so, we need data on treatments received, on spending, and on death dates if death occurred. We therefore use Medicare fee-for-service (FFS) claims as this is one of the few datasets where spending and details of treatments can be reliably connected to death dates. Our sample consists of elderly FFS Medicare beneficiaries who had an inpatient admission in 2001-2011 for one of the following conditions: acute myocardial infarction (AMI, or heart attack), congestive heart failure (CHF), or pneumonia. The full details of how the sample was put together and how risk adjustment was performed are in the appendix. Beneficiaries were included if they had a full year of FFS enrollment prior to the index admission (for risk adjustment based on diagnoses) and a full year after the admission or death within the year after the admission, to measure outcomes. Enrollment and death dates are taken from the enrollment file.

Our study has several limitations, due to well-known data constraints. First, as noted above, we needed to be able to tie diagnoses and treatments to mortality outcomes up to a year after the index event and the Medicare claims files are virtually the only data source available that allows this capability. However, as a result, our study is limited to creating price indexes for these conditions for elderly FFS Medicare beneficiaries. We have no information on the commercially insured, Medicare Advantage enrollees, the non-Medicare publicly insured or the uninsured, so these price indexes are not a comprehensive measure for the US population as a whole. Moreover, in common with other papers in this literature, we are only able to measure health outcomes with mortality. We have no data on quality of life for these beneficiaries. Finally, we lack spending and treatment data on outpatient pharmaceuticals for all beneficiaries in our sample.

With these data, we create quality-adjusted and quality-constant price indexes for the three acute conditions, AMI, CHF and pneumonia, using the methods described in the previous section. The quality adjustment and the measure for holding quality constant in the first two indexes are based solely on retrospective mortality outcomes of the patients; we make no process-based adjustments.

2. Descriptive Statistics

The selected health conditions in the paper tend to afflict older individuals. Over 70 percent of the events in our sample are for individuals over the age of 75, even though half of the population in this age range is between 65 and 75. The high average age of individuals afflicted with these conditions, suggests that they are likely to have a lot of comorbidities, highlighting the

importance of adjusting for severity. Although we are working with just a 5 percent sample of Medicare enrollees, it is also worth noting that the sample is sufficiently large for each of these condition categories to form an appropriate index.

Table 1. Distribution of patients across condition cohorts.

	Heart attack	Heart failure	Pneumonia
Total Events	62,939	96,290	109,739
	Proportion	Proportion	Proportion
Male	43.5%	36.9%	38.6%
Age group:			
Age: 65-69	11.7%	8.2%	8.9%
Age: 70-74	17.1%	13.2%	14.3%
Age: 75-79	20.0%	18.1%	18.5%
Age: 80-84	20.9%	22.4%	21.8%
Age: 85-89	17.4%	21.0%	19.6%
Age: >=90	12.8%	17.2%	16.9%
Race			
White	88.9%	86.0%	88.7%
Black	7.2%	10.2%	7.1%
Asian	0.9%	0.8%	1.0%
Hispanic	1.7%	1.8%	1.8%
Others	1.3%	1.1%	1.4%

To better gauge how adjusting for severity impacts our key estimates, we first show unadjusted estimates of spending and survival. Table 2 below shows both unadjusted expenditures and survival rates at the beginning of our sample in 2001 and the end of our sample in 2011. We see that the annual costs of these acute events are quite high and growing, with each condition exceeding $30,000 and growing by over 18 percent above general inflation. These values can be compared to health care spending per capita for the above 65 population for 2012, which was

reported to be $18,544.[4] Across each condition, we see spending rising over this period, but with the cost of treating heart attacks rising at the lowest rate. Mortality rates are falling and the number of days of survival over a one year period are rising for heart attacks and pneumonia, but mortality rates actually rise for heart failure. These trends are unadjusted, however, for demographics and comorbidities.

Table 2: Unadjusted 1-year mortality, survival, and spending estimates for each condition cohorts.

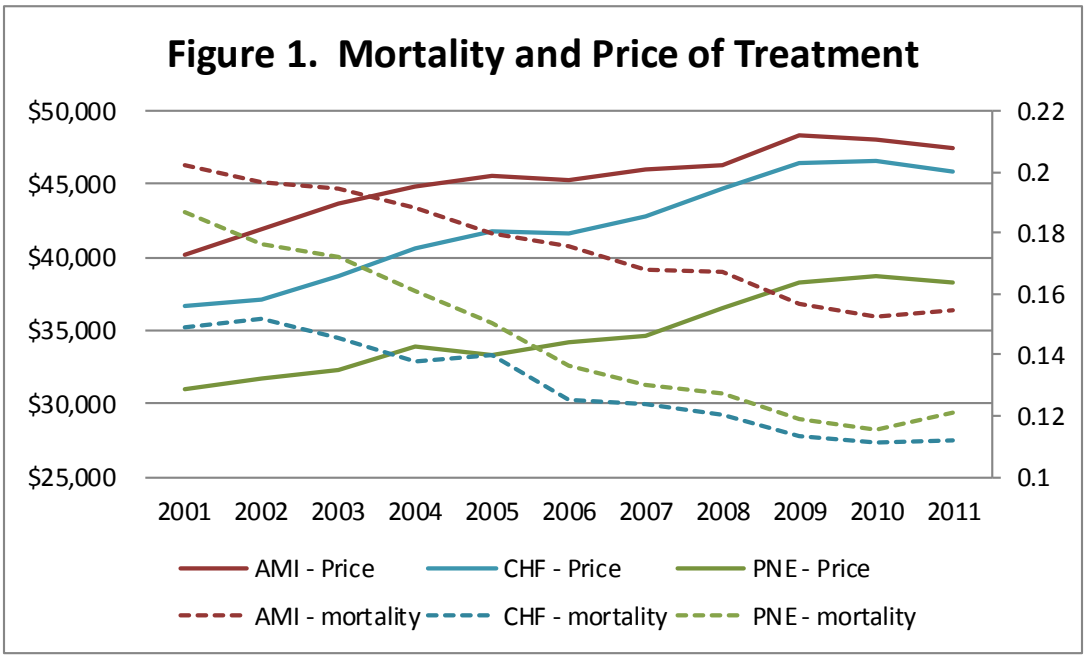
	Heart attack			Heart failure			Pneumonia		
Year	Mortality	Survival (in Days)	All spending	Mortality	Survival (in Days)	All spending	Mortality	Survival (in Days)	All spending
2001	35.1%	264	$41,069	35.8%	277	$36,271	38.2%	264	$30,791
2011	31.9%	276	$48,735	39.0%	269	$49,151	36.4%	271	$41,892
%Chg	-9.1%	4.5%	18.7%	9.0%	-2.9%	35.5%	-4.7%	2.7%	36.1%

Note: Spending is reported in 2011 dollars using the GDP deflator, which grew by 23 percent over this time period.

As discussed above, we want to hold the environment constant since our goal is a conditional cost-of-living index. In this application that means adjusting measures of spending and outcomes for patient demographics and comorbidities to accurately capture the changes in health care technology and quality conditional on those factors. We therefore adjust for severity by applying standard regression techniques that control for the demographic and other health conditions of individuals and details of these methods are outlined in the appendix.

[4] This figure of above 65 Personal Health Care expenditures per capita is taken from the National Health Expenditure Accounts (NHEA) estimates by age and gender, Table 7 for 2012.

After adjusting for severity, Figure 1 shows a clear and consistent pattern across the three conditions. The prices of treatments are rising for the three conditions, but the risk-adjusted mortality rates are falling. Figure 1 highlights the goal of this paper: combining the price and quality information (i.e., the mortality estimates) into quality adjusted price indexes. The next section discusses more specifically how we take the theoretical concepts from the previous section and apply them.



Results

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

LE index: As described in section 2, the LE index approximates a COLI by adjusting the numerator of the price index by the change in value (ΔB) received from medical care. We construct the COLI for the entire population as derived in the previous section:

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

The key challenge of this index is evaluating the monetary benefit of the quality change, ΔB. The benefit of the change is from the increased life expectancy following the hospitalization induced by improvements in treatment technologies and practices. If we measure this increase simply with the observed change in life expectancy, however, we run into two issues. The first issue is how to isolate the benefits for this condition, when other factors such as shifts in the mortality rate for cancer 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. A problem with this approach for the purposes of national accounting, however, is that it may be difficult to apply when looking at a broader set of conditions, as it would not be clear then how to define the general population.[5]

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

[5] In addition, there are likely to be improvements for other health conditions, leading to a reduction in relative benefits.

time after the event, either 30, 60 or 90 days. After the 30, 60 or 90 day period, we make the assumption that the health of the population that experienced the event is identical to the health of the population that survived the event in the initial period of the data. 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 10 year period.[6] 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 the 30 day mortality rate, 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 improvements in treatments only affect mortality after the window. For example, a new treatment may not affect 30-day mortality, but could improve the 60-day mortality rate.

We must also settle on a monetary value for a year of life. 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. As they note, the $150,000 amount has been justified as an upper threshold by the World Health Organization since it is approximately three times GDP per capita (Neumann et al. 2014).

In calculating our LE indexes therefore, we use a range of values for both the length of the window over which we measure the short-term benefits of treatment and for the monetary value of a life-year. We allow the window to be 30, 60 or 90 days and we allow the value of a life year to be $50,000, $100,000, or $150,000. The estimates of unadjusted indexes and indexes adjusted

[6] Specifically, for those that survived the acute event in 2001, we look at their mortality rate in the following 10 years. For all subsequent years, we assume that individuals that survive the event have the same post-event survival rate as those that survived in 2001. Ideally, we would compare mortality of the population that has an AMI with a comparable population that does not, but this is not possible for very recent mortality estimates. Our analysis provides a practical alternative.

for the value of change life expectancy are reported in Table 3. The unadjusted indexes report the growth of the average annual costs for each condition, deflated with the GDP deflator.

Table 3. Summary of Alternative Annual Growth Rates Using the Life Expectancy Approach Across Different Assumptions - Disease Specific Index

	30 days			60 days			90 days		
Annual Value of A Life	$50,000	$100,000	$150,000	$50,000	$100,000	$150,000	$50,000	$100,000	$150,000
AMI									
Unadjusted index	1.7%								
COLI	-0.7%	-3.1%	-5.6%	-1.1%	-3.9%	-6.8%	-1.3%	-4.3%	-7.4%
Congestive heart failure									
Unadjusted index	2.3%								
COLI	1.2%	0.1%	-1.1%	0.8%	-0.8%	-2.3%	0.5%	-1.3%	-3.1%
Pneumonia									
Unadjusted index	2.1%								
COLI	-0.7%	-3.6%	-6.5%	-1.6%	-5.4%	-9.3%	-2.1%	-6.5%	-11.0%

Notes: Estimates are computed as compound annual growth rates. The COLI estimates are computed by rebasing the amounts in each year. Price index is calculated after applying GDP deflator to 2011 dollars, which grew by 23 percent, or about 2 perecent annually over this time period. About 2 percentage points would need to be added to the index to remove this deflator effect.

We make a few observations about the results in Table 3. First, quality adjustment turns out to be important across all of the assumptions. For each scenario we observe the quality adjustment as having a large impact, relative to the unadjusted index. The unadjusted indexes show annual price increases of around 2 percent a year across conditions, while the growth rates of the quality-adjusted indexes are lower and often negative.

Table 3 shows that the estimates are much more sensitive to the variations in the value assigned to a life than to variations in the time period over which we are measuring benefits. Fixing the value of a life at $100,000, the table shows that the time period over which benefits are measured has a moderate impact for these conditions, with a difference of 2-3 percentage points. However,

assigning the value of a year of life to be $50,000 compared to a value of $150,000, can change the inflation rate by a larger amount. For instance, holding days of measured benefit to be 60 days, we find the annual inflation rate is lower by between 5-8 percentage points more if the value of a life is placed at $50,000 compared to $150,000.

When we compare our results to those of Cutler et al. (2001), they are similar although our methods and assumptions are not identical. They find an annual inflation rate for the treatment of AMI of around -1 to -2 percentage points below general inflation. They used relatively conservative estimates of the value of a year of life of $25,000 in 1991 dollars which is $36,000 in 2009 dollars. Using our conservative value of a life year of $50,000 and allowing benefits to change up to a 60-day window we find that the average inflation rate across conditions, weighting by expenditure share, is -0.5 percentage points below general inflation. With the $100,000 per life year estimate and 60 day window we arrive at an average inflation rate of -3.2 percentage points below general inflation. For the AMI condition studied in Cutler et al. (1998, 2001), we find an annual inflation rate of between -1.1 and -3.9 relative to general inflation. Based on our indexes, the quality-adjusted prices for acute conditions like AMI are still declining. Even when we use our most conservative estimate of a value of a life-year of $50,000, we find inflation declines at a -1.1 percent rate, which falls on the lower end of the Cutler et al. (2001) estimate.

TE index: We construct the treatment endpoint (TE) index in the same way as Berndt et al. (2002) construct their index but with the endpoints for the conditions as defined by Romley et al. (2015) who study the same acute inpatient conditions that we consider here. Berndt et al. (2002) construct their index by measuring the price of the expected achievement of an endpoint (remission of major depression). Their prospective measurement is based on the estimated

medical value of the treatments being given. Romley et al. (2015) meanwhile measure hospital output retrospectively, defining successful output as meeting the endpoint of survival to 30 days without an unplanned readmission.

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

$$\frac{S_1}{\sigma_1} \Big/ \frac{S_0}{\sigma_0}$$

where S_t is average spending as defined above and σ_t is the percent of treatments that are successful relative to no treatment. Using the stylized model discussed earlier, $\sigma_t = q_t\pi_1 + (1 - q_t)\pi_2 - \pi_3$. Following Romley et al. (2015), we define "successful" treatment as surviving up to 365 days without an unplanned readmission within 30 days of discharge, with unplanned readmissions identified with the algorithm used by the Centers for Medicare & Medicaid Services (CMS). Survival rates with "successful" treatment are risk-adjusted as indicated above.

A challenge of constructing a TE index is that, since it measures the change in the incremental price relative to no treatment (i.e., π_3), it is necessary to know the rate of reaching the endpoint without any medical treatment at all. Berndt et al. (2002) were able to estimate the rate of remission of major depression without any treatment based on expert opinion because it was not uncommon for major depression to go untreated. However, for the conditions we are considering, every patient we observe receives treatment so it is difficult to know what the rate of success for patients that go untreated should be. At one extreme, the illnesses studied here are sufficiently severe that one may view non-treatment as a complete failure, so that the rate of

success for untreated cases is arguably zero. However, there is the potential for survival without treatment for all of these conditions. For example, prior to the 1960s when none of the modern treatments were available, 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 (Lee 2011). We view these estimates as providing an approximate baseline for "non-treatment." Below we show alternative indexes based on differing assumptions for untreated cases. As the assumption regarding the success rate of untreated cases increases, the value of the quality changes increases because it suggests that the relative incremental change in the quality is larger. For example, if the untreated successful rate of treatment is 40 percent, then a increase in the treated success rate going from 60 percent to 65 percent would be (65-40)/(60-40)=25/20=1.25 or a 25 percent increase in quality. However, if the untreated success rate is 0 percent, then the quality change is (65-0)/(60-0)=1.083 or just 8.3 percent. In our study, moving from a 0 percent success rate for untreated cases to a 40 percent success rate, leads to an annual inflation rate that is lower by about 2-3 percentage points.

Table 4. Treatment Endpoint Index Under Alternative Assumptions Regarding Untreated Success Rate

	Rate of Success for Untreated Cases		
	0%	20%	40%
AMI			
Unadjusted index	1.7%		
Average Success Rate with Treatment	58.4%		
Treatment Endpoint Index (CAGR)	0.6%	0.0%	-1.7%
Congestive heart failure			
Unadjusted index	2.3%		
Average Success Rate with Treatment	54.1%		
Treatment Endpoint Index (CAGR)	1.2%	0.6%	-1.7%
Pneumonia			
Unadjusted index	1.7%		
Average Success Rate with Treatment	57.8%		
Treatment Endpoint Index (CAGR)	0.7%	-0.1%	-2.6%

BP index: The next method constructs a technology-constant index following Frank et al. (2004). This method holds the quality of care constant by holding key treatment technologies fixed. Specifically, this method holds the share of patients receiving specific treatments constant over time while allowing the prices those treatments to change. Specifically, we run the following GLM regression, separately for each condition and year:

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

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

construct a Paasche-type index using the same method on the 2011 data. The final index is a Fisher index, the geometric average of the two.

For the heart attack 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 is 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 heart failure 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.

Pneumonia treatment mostly relies on antibiotics. Given the difficulty in using ICD-9 codes in Medicare claims data available to us to identify the many different antibiotic consensus recommendations for treating pneumonia, we did not create BP indexes for the pneumonia cohort.

When we apply the BP method to AMI and CHF, we find that there is very little difference between the BP indexes and the unadjusted indexes. Given the limited change in these estimates relative to the unadjusted figures, we do not report these estimates separately but show them in the next section when we compare across methods (Figures 1-3).

The BP indexes also diverge substantially from the outcomes-based LE and TE indexes both of which fall much faster than the unadjusted and BP indexes. This divergence suggests that the shift in the shares of the treatment baskets that we have defined are not actually related to the changes in observed outcomes captured in the two outcomes-based indexes. As discussed in section 3, a potential hazard of the BP index is defining treatment baskets that do not actually cause a difference in outcomes; in that case, the BP index does not capture the true quality shift. The improvements in mortality outcomes of AMI and CHF that we observe may have been caused by shifts among other treatments that we did not count separately, or by improvements in quality not captured in the claims data, such as improved coordination among hospital staff.

RC Index: This approach is often preferred by BLS when adjusting the PPI for quality and it is used by BLS to adjust the current hospital and nursing home PPIs. A challenge in the health care field is that it may be difficult to assign a cost to a quality change. For pharmaceuticals, for example, much of the costs associated with the innovation are fixed costs of development that may be difficult to measure. In the hospital setting many costs are also fixed and would need to be dispersed across a wide range of conditions (e.g., the hospital structure or purchases of capital equipment like MRI machines). As mentioned previously, there may also be quality changes where there is no or little associated cost (e.g., receiving an aspirin within 24 hours of admission for an AMI).

Section 3 defined a formula for the RC index based on a model with defined treatment baskets with differing and known levels of quality: $RCI = UI - \frac{\Delta q*(R_{11}-R_{21})}{S_0}$. As discussed above, however, the concordance of the BP index with the unadjusted index and its divergence from the

LE and TE indexes mean that the shifts in the shares of the treatment baskets we have defined may not be related to the changes in outcomes we are observing.

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

The average cost of attaining a successful health outcome can be calculated as $\frac{S_t}{\sigma_t}$ .We also write the change in the share of health outcomes that are successful over time as: $(\sigma_1 - \sigma_0)$. If cost is linear in successful treatments, then the estimated cost of the quality change can then be written as the average cost per outcome times the change in the outcome: $\frac{S_t(\sigma_1-\sigma_0)}{\sigma_t}$. That is, rather than using the change in the value of lives saved to adjust the index, this term is capturing the approximate change in the cost to achieve an outcome. The price index can then be written as:

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

Substituting $\frac{S_t}{\sigma_t}$ with period 1's price per quality change, $\frac{S_1}{\sigma_1}$, the result then becomes:

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

Therefore, applying these assumptions, we see that the resource cost method is identical to the treatment endpoint method.

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

$$RC = \frac{S_1 - \frac{\alpha S_1(\sigma_1 - \sigma_0)}{\sigma_1}}{S_0} = \frac{S_1(1-\alpha) + S_1\alpha\frac{\sigma_0}{\sigma_1}}{S_0} =$$

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

The assumptions that we would make to calculate the RC index lead to an index that is extremely close to the TE index, so we do not report calculations for this index separately. However, as the above equation shows, quality adjustment would tend to have a dampening effect on the RC index relative to the TE index.

Violation of the assumption that quality is related to cost has different implications for this approximation of the RC index from the theoretical RC index. Recall that one criticism of the RC index is that if quality changes without a corresponding change in the cost, there would be no

change in the index. Because we are now assuming that changes in costs are directly related to the quality change and can therefore be inferred from a change in observed outcomes, this re-formulation of the RC index assumes that if the quality changes by an amount $(\sigma_1 - \sigma_0)$, there is necessarily a quality adjustment using the RC formula. The assumption that changes in costs are linear in changes in outcomes is quite strong, however, and not necessarily typical of health care cost/quality functions.

Across-Method Comparison

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

Figures 1 to 3 show the differences between the indexes graphically. All calculations were made with amounts measured in 2009 dollars so the growth rates pictured are relative to economy-wide inflation as measured by the GDP deflator. Across the three conditions we find similar patterns. We find that the unadjusted index and the BP index are nearly identical within conditions (recall it is not applied to pneumonia). We find that inflation is considerably lower when measured by outcomes in both the TE index and LE index, relative to the unadjusted index, but the amount of adjustment is much larger for the LE index.

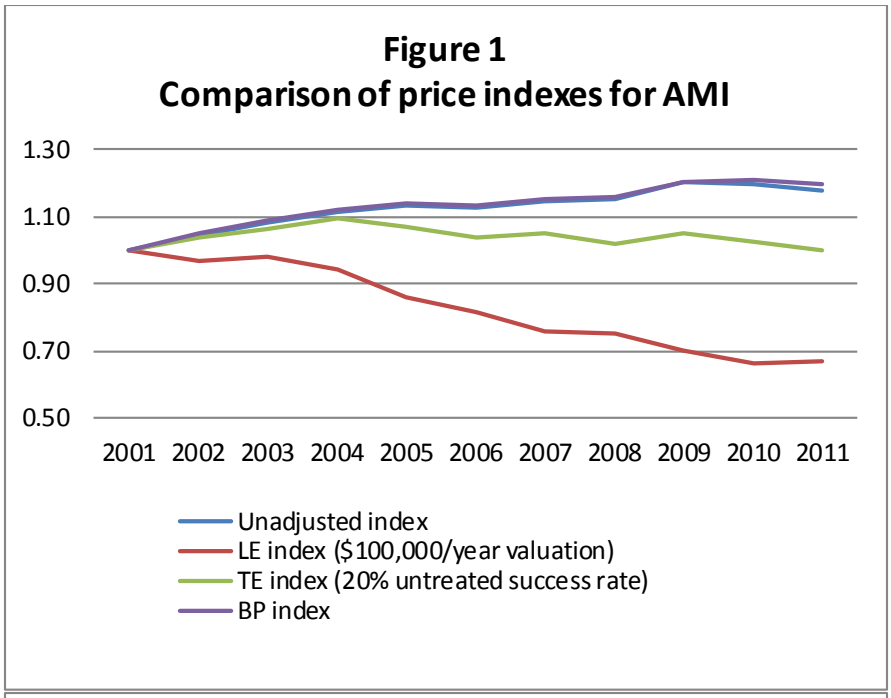


Figure 1
Comparison of price indexes for AMI
1.30
1.10
0.90
0.70
0.50
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
Unadjusted index
LE index ($100,000/year valuation)
TE index (20% untreated success rate)
BP index

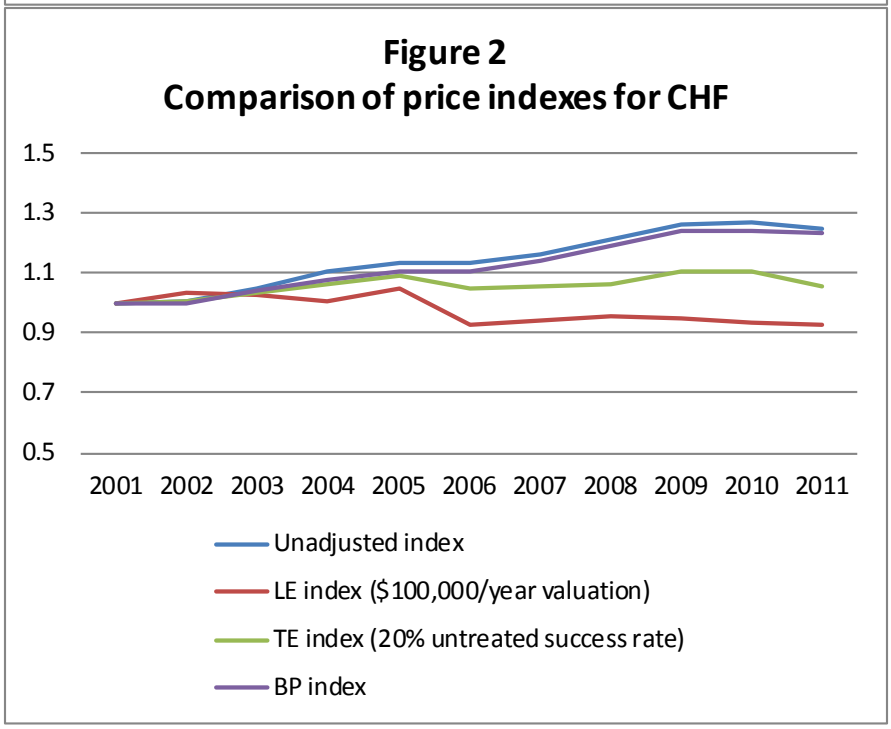
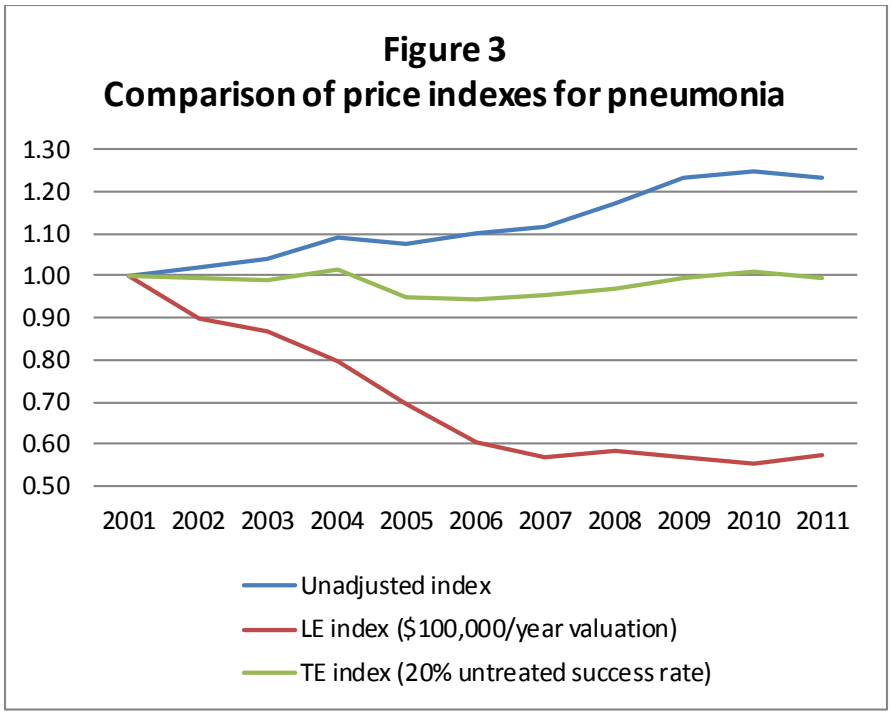


Figure 2
Comparison of price indexes for CHF
1.5
1.3
1.1
0.9
0.7
0.5
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
Unadjusted index
LE index ($100,000/year valuation)
TE index (20% untreated success rate)
BP index



Also recall that the TE index is nearly identical to the re-formulation of the RC index suggested in the previous section. The only difference will be a small adjustment due to a profit margin of the hospital, which is estimated to be between 2 and 6 percent.[7] All else equal, we expect the inflation rate of the TE index to be only slightly below that of the re-formulated RC index.

Adjusting for quality is clearly important since outcomes improved markedly during this period. Across the TE and LE methods that adjust for quality based on mortality outcomes, the adjusted inflation rate is more than 1 percentage point below the unadjusted rate. Although quality adjustment matters, the method used by researchers to make the adjustment is also quite important. We find a very large range in quality adjustments depending on the assumptions applied, from between 1 and 13 percent below the adjusted estimate.

[7] These estimates are based on AHA survey data. See http://www.aha.org/research/reports/tw/chartbook/2016/chapter4.pdf.

As discussed in section 2, differences can arise between the LE index and other indexes when the increase in the benefits of treatment diverge greatly from the increase in spending. That seems to be the case for the conditions under study here. For AMI, for example, risk-adjusted spending per patient increased by $7,214 while increased benefits (with short-term benefits measured through a 60-day window and with a year of life valued at $100,000) add up to $24,710, or more than three times the increased spending. For CHF, the difference between benefits and costs is not as great; under the same assumptions, the increase in monetized life expectancy only exceed increased spending by about $3,000 and, accordingly, the divergence between the different price indexes is not as great. For pneumonia, the results are similar to AMI; under the same assumptions as given for AMI above, spending per pneumonia patient increased by $7,169 while the monetized value of the increase in life expectancy increased by $24,511.

3. Conclusion

This paper reviews different methods for quality adjustment of medical price indexes. We contrast these methods theoretically and empirically for three acute conditions. In the theoretical model, we find that a COLI-type index whose quality adjustment is based on the monetized value of the increase in the health benefits of treatment such as that constructed by Cutler et al. (1998, 2001) gives the most theoretically accurate results and that the other indexes are closer to the COLI when the increase in medical spending is closer to the value of the increase in health benefits.

We conducted an empirical application by constructing these indexes for three acute conditions in the Medicare population in the period from 2001 to 2011: acute myocardial infarction (AMI),

congestive heart failure (CHF), and pneumonia. All of these conditions showed improvements in post-hospitalization life expectancy during this period, AMI and pneumonia more so than CHF. In this empirical application, we find that quality adjustment has a significant effect on the levels of the price indexes. Since, however, we also found that the monetized increases in life expectancy exceeded the increases in spending by considerable margins, the other quality-adjusted indexes diverged greatly from the benchmark COLIs for each condition.

There were several limitations to this analysis. First, our sample was limited to elderly fee-for-service Medicare patients as we did not have appropriate data for other types of patients. Second, our estimate of the benefits of medical treatment is limited to the mortality benefits; we have no estimates of the quality of life of these patients. The lack of data on quality of life has two potentially biasing effects. First, we are not measuring any improvements or declines in quality of life from medical treatment. Second, we are also potentially overestimating the value of the additional quality-adjusted life since our range of assumptions for this value is based on the average for the population while the sample we are studying is mostly very elderly Medicare patients.

The third limitation is that we are only considering three conditions and the results should therefore not necessarily be generalized to the rest of medical care. Shapiro et al. (2001) found that inflation measured at the medical condition level was considerably lower than inflation measured at the service level even without any quality adjustment when they measured it for one specific condition, cataracts. The Health Care Satellite Account, however, showed that this result did not hold across conditions in a later time period since it found that inflation measured by condition was actually slightly higher than inflation measured by current methods. The previous research on condition-based medical price indexes was generally conducted on conditions for

which there was considerable technological advancement in the period being studied, which would bias towards finding lower inflation. There are many conditions and time periods, however, in which there is little technological advancement in medical care.

Our working examples show that, as predicted by Triplett (1982) and others, the benefits and costs of health care can diverge considerably, both because of the measurement challenges related to new innovations as well as well-known market frictions in health care. It was already known that standard methods for quality adjustment of price indexes do not work for health care; here we have shown that even methods developed specifically for health care have the potential to give biased results in certain circumstances. Going forward, there is considerable room for further development of quality-adjusted price indexes for medical conditions as measurements of quality of life are improved, more detailed data on treatments and outcomes become available, and valuations of the quality and amount of life become more certain. The need for assessing the true value of medical spending is only going to become more pressing in upcoming years.

References

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

Bradley, R., J. Hunjan, and L. Rozental. 2015. Experimental Disease-Based Price Indexes. BLS working paper. Available at http://www.bls.gov/pir/journal/rb03.pdf. Last accessed 10/19/16.

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. 2011. "An Administrative Claims Model for Profiling Hospital 30-Day Mortality Rates for Pneumonia Patients," *PLoS ONE* 6(4): e17401.

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

Bureau of Labor Statistics. 2013. Quality Adjustment in the Producer Price Index. Available at http://www.bls.gov/ppi/qualityadjustment.pdf. Last accessed 10/19/16.

Bureau of Labor Statistics. 2015. Handbook of Methods, Chapter 17: The Consumer Price Index. Available at http://www.bls.gov/opub/hom/pdf/homch17.pdf. Last accessed 11/15/16.

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

Cutler, D. M., McClellan, M., Newhouse, J. P., and Remler, D. 2001. Pricing Heart Attack Treatments. in *Medical Care Output and Productivity*. David Cutler and Ernst Berndt, eds. Chicago: U. Chicago Press, pp. 305-62.

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

Dunn, A., Rittmueller, L., and Whitmire, B. 2015. Introducing the New BEA Health Care Satellite Account. *Survey of Current Business* 95(1). Available at http://www.bea.gov/scb/pdf/2015/01%20January/0115_bea_health_care_satellite_account.pdf. Last accessed 10/19/16.

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

Hall, A. E. 2016. "Adjusting the Measurement of the Output of the Medical Sector for Quality: A Review of the Literature." *Medical Care Research and Review*. Published online ahead of print August 11, 2016. doi:10.1177/1077558716663388.

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

Lee, R. 2011. "Surving a heart attack: a success story." Harvard Health Letter. http://www.health.harvard.edu/heart-health/surviving-a-heart-attack-a-success-story.

Lucarelli, C. and Nicholson, S. (2009). A Quality-Adjusted Price Index for Colorectal Cancer Drugs. NBER working paper 15174 (http://www.nber.org/papers/w15174).

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

Neumann, P. J., J. T. Cohen, and M. C. Weinstein. 2014. "Updating Cost-Effectiveness: The Curious Resilience of the $50,000-per-QALY Threshold," *The New England Journal of Medicine* 371: 796-797. DOI: 10.1056/NEJMp1405158.

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

Romley, J. A., Goldman, D. P. and Sood, N. (2015). US Hospitals Experienced Substantial Productivity Growth During 2002-11. *Health Affairs* 34(3): 511-58.

Sheiner, L. and A. Malinovskaya. 2016. "Measuring Productivity in Healthcare: An Analysis of the Literature." Working paper, Hutchins Center on Fiscal and Monetary Policy, The Brookings Institution.

Shapiro, Irving, Matthew Shapiro, and David Wilcox, (2001), "Measuring the Value of Cataract Surgery,". In Medical Care Output and Productivity, David M. Cutler and Ernst R. Berndt, eds. Chicago: University of Chicago Press

Appendix

Data sources

This study uses 2000-2012 Medicare claims data from the inpatient, outpatient and carrier (physician) files. However, we perform the analysis only for the period 2001-2011. The 2000 data sets were used to identify a 365-day history of certain conditions for index admissions occurring in 2001 and the 2012 data sets were used to get the full 365-day spending and survival measures for index admissions occurring in 2011. We obtain patient demographic, enrollment and mortality information from the enrollment files.

Patient cohorts

The analytical sample includes a cohort of Medicare beneficiaries aged at least 65 years with an inpatient hospitalization and a primary discharge diagnosis for acute myocardial infarction (AMI or heart attack), congestive heart failure and pneumonia between 2001 and 2011. 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 congestive heart failure 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 either based on those used by (CITE) or based on the literature (CITE).

We restrict the samples to fee-for-service beneficiaries who were continuously enrolled for 365 days before the index admission and 365 days (or until death) after the index admission. The 365 days prior to the index admission requirement is to ensure that we have a full 1-year history of certain conditions that we use as risk adjusters and the 365 days post index admission requirement is to ensure that we are able to capture the full 1-year spending and survival measures after the index admission. We allow at least a 365-day window after an index admission of a particular cohort before a patient can reenter that cohort. However, a patient can appear in a different cohort during the 365-day window of one cohort. In effect, a patient can

appear multiple times within a particular cohort or appear in different cohorts during the sample period.

Outcome variables

The key outcome measures are one-year survival days, one-year survival rates and one-year spending. The spending variable encapsulates all medical care expenses incurred in an inpatient, outpatient or physician office settings during and after the index admission and is inflation-adjusted to 2011 dollars using the U.S. gross domestic product implicit price deflator.

Risk adjusters

To obtain predicted average yearly survival days, survival rates and spending for each condition 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 race groups (i.e., Whites, Blacks, 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 (see Krumholz et al. 2006a, 2006b; Bratzler et al. 2011).[8] 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 the following HCC groupings: Unstable angina, Chronic atherosclerosis, Cardiopulmonary-respiratory failure and shock, Valvular heart disease, Hypertension, Stroke, Renal failure, COPD, Pneumonia, Diabetes, Protein-calorie malnutrition, Dementia, Hemiplegia-paraplegia-paralysis-functional disability, Peripheral vascular disease, Metastatic cancer, Trauma in last year, Major psychiatric disorders, and Chronic liver disease. Additional cohort-specific covariates include two dummy variables for

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

the AMI locations[9] in the heart attack cohort, Cerebrovascular diseases in the heart failure 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.

Appendix Page:

Table A1: The number of observations for each condition cohort.

	Heart attack	Heart failure	Pneumonia
2001	15838	24597	27184
2002	16223	25030	29099
2003	15941	26684	30387
2004	14953	26653	27951
2005	13703	25746	30217
2006	12753	24946	26551
2007	12065	23023	24300
2008	11720	21957	24269
2009	10698	21570	21760
2010	10829	21013	21312
2011	10099	19802	21458
Total	144822	261021	284488

Note: For each estimation sample.

[9] 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.