Implications of Utilization Shifts on Medical-Care Price Measurement*

Abe Dunn, Eli Liebman, and Adam Hale Shapiro

September 23, 2013

Abstract

The medical-care sector often experiences changes in medical protocols and technologies that cause shifts in treatments. However, the commonly used medical-care price indexes reported by the BLS hold the mix of medical services fixed. In contrast, episode expenditure indexes, advocated by many health economists, track the full cost of disease treatment, even as treatments shift across service categories (e.g., inpatient to outpatient hospital). In our data, we find that these two conceptually different measures of price growth show similar aggregate rates of inflation over the 2003 to 2007 period. Although aggregate trends are similar, we observe differences when looking at specific disease categories.

*The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the Bureau of Economic Analysis, the Federal Reserve Bank of San Francisco, or the Board of Governors of the Federal Reserve System.

1 Introduction

The rapid rise in health-care costs has led many researchers and policy makers to search for statistics to help inform policy decisions. The passage of the Patient Protection and Affordable Care Act in 2010 has further added to the urgency to develop more meaningful statistics to assess the impact of this momentous reform. Current national health statistics report spending and prices for specific services (e.g. physician or hospital prices), but provide no information on spending by disease. This is a considerable omission, since the goal of health spending is to treat diseases and improve health.

This limitation in our national statistics has been noted by many health economists, who have advocated for reporting national statistics that track expenditures by disease (see Berndt et al. (2000) and Accounting for Health and Health Care (2010)), with a particular interest in tracking the disease price (i.e., expenditure per disease episode). There are many reasons for focusing on the disease price. Policy makers, consumers, and industry participants are increasingly interested in whether changes in the cost of treatment are worth the health benefit. By focusing on spending by disease rather than by service, researchers will be better able to connect expenditures for specific diseases with the associated health outcomes. Tracking disease expenditures also provides a more relevant unit of price for patients, since patients ultimately seek treatment for a disease regardless of the point of service (e.g., physician office, clinic or hospital). In fact, researchers have noted and documented several important shifts in treatment. For example, Shapiro and Wilcox (1996) have documented technological advances that led to shifts from inpatient to outpatient services for the treatment of cataracts.[1] Traditional price measures that restrict substitution patterns across service categories may have a substantial impact on price measurement and change in real output in this sector, which accounts for almost one-fifth of U.S. economic output.

Currently the BLS does not collect the necessary data to track prices at the disease level. To check if measures of health care inflation may be different using a disease-price

[1] Other case studies include Berndt et al. (2002) showing that drugs for depression may substitute for talk therapy; Cutler et al. (1998) looking at innovations in heart attack treatment; Lucarelli and Nicholson (2009) examining colorectal cancer treatments; and Dunn (2012) studying anti-cholesterol drugs.

index, researchers have used historic medical claims data to compare disease-price indexes to more traditional indexes that hold the basket of medical services constant. In their pioneering work, Aizcorbe and Nestoriak (2011) documented shifts in treatments across a broad range of medical conditions. Their paper measured several important shifts in treatment at the disease episode level, such as shifts away from inpatient services for many conditions. They also found a clear divide between a disease-episode expenditure index, referred to as a medical care expenditure index (MCE), which allows for shifts in encounters (i.e., a visit to a physician or facility) across service categories; and an index that holds the number of encounters fixed for each service category, an encounter-based service price index (SPI-encounter). In particular, they find that the SPI-encounter measure grows faster than an episode-based MCE measure, implying that the SPI-encounter measure would overstate inflation in the health sector. This result appears to be quite robust and has been replicated in other studies, including Dunn et al. (2012a) and Aizcorbe et al. (2011). Overall, this research hints that official price indexes may not be an accurate measure for tracking the cost of disease treatment.[2]

While the recent work looking at a broad range of medical conditions is suggestive of potential shortcomings in the Bureau of Labor Statistics (BLS) health-care price measures, those assessments are in fact incomplete. When measuring service prices, the prior literature has used an encounter-based measure (i.e., a measure based on expenditure per encounter), while the prices reported by the BLS are for precisely defined services often priced at the more granular procedural level. We will refer to pricing methods that focus on a more granular unit as "procedure-based" measures. This distinction is potentially important when considering the possible discrepancy between the BLS price measure and a cost-of-treatment price measure. Specifically, there could be within-industry changes in the intensity of treatments *per encounter* (i.e., treatment per visit), which could lead to differences in the BLS's procedure-based service price measure and an encounter-based service price measure. For example, if there is an increase in intensity of treatment per encounter at a doctor's office (e.g., going from an X-ray to a MRI) this will tend to cause the encounter-based measure to report faster price growth, while there would be

[2] This concept is related to the contributions of Berndt, Cockburn and Griliches (1996) and Griliches and Cockburn (1994), who demonstrate that pricing branded and generic drugs as equivalent products, rather than distinct products, may have a large impact on price measurement.

no effect on the BLS price measure.

To study the components of episode expenditure growth, we start by estimating an MCE index, similar to that reported in Aizcorbe and Nestoriak (2011). The MCE index is then compared to two indexes: (1) a new procedure-based methodology for measuring service prices (SPI-procedure); and (2) an encounter-based approach for measuring service prices (SPI-encounter), applied in Aizcorbe and Nestoriak (2011). The procedure-based approach for measuring service price indexes (SPI-procedure), developed here, accounts for the potentially large differences in intensity of treatment across procedures. The procedure-based approach for measuring service prices is closer to that performed by the BLS. Indeed, we show empirically that the growth in the SPI-procedure is closer to the BLS growth rates, relative to price measurements using an encounter-based approach. In this study we examine the 2003-07 time period using MarketScan data, which is a convenience sample of the commercial health insurance population.

The main finding is that we observe little difference between the MCE and SPI-procedure indexes in the aggregate. In other words, it appears that utilization shifts do not cause any difference between the SPI-procedure and MCE index, at least for the 2003 to 2007 period in our data. To reconcile our result with the previous literature, we dig deeper into the shifts in utilization that affect the relative growth of each index. Consistent with prior work, we find that *across-industry shifts in encounters* (i.e., from inpatient to outpatient) may lead the SPI-encounter to grow faster than the MCE, but we also find that within-industry growth in *intensity of treatment per encounter* tends to offset this effect in the SPI-procedure. In other words, the number of encounters per episode is falling, but the intensity of treatment per encounter is rising causing overall service utilization (i.e., intensity of treatment per episode) to remain almost constant. Consequently, the SPI-procedure and the MCE grow at similar rates.

Although we find offsetting factors that net out any aggregate differences in utilization shifts, we do find important differences in the MCE index and SPI-procedure for specific disease categories. These differences have potential implications for inflation in the health sector. For some condition categories, such as cardiology, we see the SPI-procedure grow faster than the MCE index, indicating that changes in inflation implied by the MCE is lower than what is implied by service prices alone. For other categories, such as orthopedic conditions, we see the reverse.[3]

There are a couple of points that are important to highlight. First, the findings in this paper are specific to the time period studied and for the commercial insurance market, and different patterns may be found for other time periods or markets (e.g., Medicare and Medicaid). In general, there is no golden rule about the relationship between the SPI and MCE indexes. Second, it is important to note that neither the SPI-procedure nor the MCE index controls for changes in the quality of treatment. However, the disease-price index offers an important first step in producing the ideal price index for the health sector, since it is more amenable to adjusting for quality changes.

The remainder of this paper is divided into four sections. The following section discusses the methodology of the index construction. Next, we present the data followed by the results. In the results section we focus primarily on one approach for measuring the components of disease price growth. However, after presenting our main results, we briefly discuss some of the relevant results from some companion papers, which demonstrate the robustness of our findings. The last section concludes.

2 Methodology of Index Construction

2.1 A Motivating Example

To help motivate our methodology, we start with a simple example and focus on a procedure-based index.[4] Consider a time period, t, where people are treated for a knee injury (k). Also suppose there exists only one type of treatment available— an X-ray. Let

[3]Some case studies have documented instances where new technologies do not lead to lower expenditures (e.g., Duggan (2005) and Frank, McGuire and Normand (2006)).

[4]The methodological framework is similar to Dunn et al. (2013) that looks at geographic differences in expenditures.

$$N_{k,t} = \text{Number of treated knee injury episodes at time } t.$$

$$c_{k,t} = \text{Average expenditure for a knee injury per episode.}^{5}$$

$$q_{k,t} = \text{Number of X-rays per episode.}$$

$$p_{k,t} = \text{Price per X-ray (i.e., } \frac{c_{k,t}}{q_{k,t}}\text{).}$$

Let t=0 be the base time period, where the price for an X-ray for a knee injury in the base period is $p_{k,0}$. In this simple case, the relative price level of t to 0 is simply $\frac{p_{k,t}}{p_{k,0}}$. Clearly, this ratio reflects only differences in the contracted prices, not the number of X-rays. Similarly, the relative utilization level is $\frac{q_{k,t}}{q_{k,0}}$ which depends only on the number of X-rays performed per episode. It follows that the relative expenditure per episode between t and 0 may be expressed as:

$$\frac{c_{k,t}}{c_{k,0}} = \left(\frac{p_{k,t} \cdot q_{k,0}}{p_{k,0} \cdot q_{k,0}}\right) \cdot \left(\frac{p_{k,t} \cdot q_{k,t}}{p_{k,t} \cdot q_{k,0}}\right). \tag{1}$$

The first term in (1) is a price index, and the second term is a utilization index. Expanding on this example, now suppose that a knee injury may be treated with two types of services, prescription drugs and physician office services, where the service categories correspond to the subscripts (D) and (O). That is, $q_{k,t,O}$ and $p_{k,t,O}$ are the utilization and price for physician office services, and $q_{k,t,D}$ and $p_{k,t,D}$ are the utilization and price for prescription drugs (e.g., pain medication). Continuing with the index decomposition that is parallel to (1), but with two services, the decomposition becomes:

$$\frac{c_{k,t}}{c_{k,0}} = \frac{p_{k,t,O} \cdot q_{k,t,O} + p_{k,t,D} \cdot q_{k,t,D}}{p_{k,0,O} \cdot q_{k,0,O} + p_{k,0,D} \cdot q_{k,0,D}} \tag{2}$$

$$= \left(\frac{p_{k,t,O} \cdot q_{k,0,O} + p_{k,t,D} \cdot q_{k,0,D}}{p_{k,0,O} \cdot q_{k,0,O} + p_{k,0,D} \cdot q_{k,0,D}}\right) \cdot \left(\frac{p_{k,t,O} \cdot q_{k,t,O} + p_{k,t,D} \cdot q_{k,t,D}}{p_{k,t,O} \cdot q_{k,0,O} + p_{k,t,D} \cdot q_{k,0,D}}\right) \tag{3}$$

Again the first term corresponds to the price index and the second term corresponds to the utilization index.

2.1.1 Procedure-Based and Encounter-Based Indexes

The essential difference between the indexes is how the quantity of treatment is measured. While the procedure-based approach focuses on the intensity of services and procedures for each service category, the encounter-based approach uses the number of encounters, where an encounter at a physician's office is defined as a day of care from that provider. Expanding on the example above using only physician services may help illustrate the differences between the encounter-based and procedure-based indexes. Consider the following hypothetical case where physicians only perform one of two procedures to treat an injured knee, an X-ray (a low intensity procedure) or an MRI (a high intensity procedure). Suppose the number of physician office encounters is the same in period 0 and t, but there is a shift from physicians performing X-rays in period 0, to performing MRIs in period t. Also suppose that the contracted prices for both services do not change. In this case, the procedure-based index will show a increase in quantity, $\frac{q_{k,t}}{q_{k,0}}$, and no change in price, $\frac{p_{k,t}}{p_{k,0}} = 1$. In contrast, using an encounter-based index, both $q_{h,t}$ and $q_{h,0}$ are simple encounter counts, so there would be no change in $q_{h,0}$ and there would be an increase in price.[6]

2.2 The General Case

In the general case, we define the medical-care expenditure for the treatment of an episode of a disease (that is, a specific condition) as the total dollar amount of medical care used until treatment is completed, including *all* service categories.[7] Formally, denote the average expenditure paid to medical providers for an episode of treating disease d at time period t as $c_{d,t}$. The MCE index is a measure of the relative medical-care expenditures for an episode of care for a certain disease. The MCE index for disease d is:

[6] Alternatively, rather than a difference in intensity of procedures, the two indexes could also diverge if there is a change in the number of procedures conducted per encounter with a physician.

[7] For medical diseases that are chronic, we interpret an episode as the total expenditure for services used to treat the chronic disease over a one-year period.

$$MCE_{d,t} = \frac{c_{d,t}}{c_{d,0}}. \tag{4}$$

Thus, similar to the example above, if the $MCE_{d,t}$ is larger than one, it signifies that the expenditure for treating disease d is larger than initial period 0, and if the index is less than one it signifies that the expenditure is less than the base period expenditure.

Next, we decompose the MCE index into two distinct components: a service price and service utilization component. This can be seen more easily by showing that the average expenditure is calculated by totaling dollars spent on all services to treat the disease and dividing those dollars by the number of episodes: $c_{d,t} = \sum_s p_{d,t,s} Q_{d,t,s}/N_{d,t}$, where $Q_{d,t,s}$ is the quantity of services for service type, s; $p_{d,t,s}$, is the service price for service type s; and $N_{d,t}$ is the number of episodes treated. To simplify, let $q_{d,t}$ be a vector of services utilized for the typical treatment of diseases in an MSA, $q_{d,t} = Q_{d,t}/N_{d,t}$, where the component of the utilization vector for service type s is, $q_{d,t,s} = Q_{d,t,s}/N_{d,t}$. Similarly, let $p_{d,t}$ be a vector of service prices, where the the price for a particular service type and disease can be calculated by dividing its average expenditure by the average quantity of services provided: $p_{d,t,s} = \frac{c_{d,t,s}}{q_{d,t,s}}$ where $c_{d,t,s}$ is the average episode expenditure for disease d for service type s in time period t. This decomposition allows us to create a service price and service utilization index. The service price index (SPI) is then calculated as:

$$SPI_{d,t} = \frac{p_{d,t} \cdot q_{d,0}}{c_{d,0}}, \tag{5}$$

which holds the utilization of services fixed at the initial period. The SPI measures the compensation necessary to purchase a fixed utilization of medical goods when moving from the initial period to period t. The service utilization index (SUI) may be defined as:

$$SUI_{d,t} = \frac{p_{d,0} \cdot q_{d,t}}{c_{d,0}}, \tag{6}$$

which holds the price of services fixed while allowing the utilization of services to vary. The SUI measures the compensation necessary to purchase medical goods at base period prices when moving from the initial period to time t. We choose to apply Laspeyres indexes for price and quantity, so that the estimates may be compared to a national "base" amount: essentially answering the question, how much are disease expenditures growing relative to the national average due to price differences or due to utilization differences? With these indexes the decomposition that relates these three indexes is additive, rather than multiplicative.[8] The relationship between these three indexes is described by the following decomposition:

$$MCE_{d,t} = SPI_{d,t} + SUI_{d,t} + \frac{(q_{d,t} - q_{d,0})(p_{d,t} - p_{d,0})}{c_{d,0}} - \frac{p_{d,0} \cdot q_{d,B}}{c_{d,B}}. \tag{7}$$

Here the MCE index is equal to the service price index, $SPI_{d,t}$, plus the service utilization index, $SUI_{d,t}$, plus a cross term, $(q_{d,t} - q_{d,0})(p_{d,t} - p_{d,0})/(c_{d,0})$, minus $\frac{p_{d,0} \cdot q_{d,0}}{c_{d,0}}$ (which is close to one). The term, $(q_{d,t} - q_{d,0})(p_{d,t} - p_{d,0})/(c_{d,0}) - \frac{p_{d,0} \cdot q_{d,0}}{c_{d,0}}$, accounts for joint changes in price and utilization and, in practice, the term is near minus 1. In the case where there are few changes in utilization per episode over time, $SUI_{d,t}$ is fixed near 1, and the $MCE_{d,t}$ will be determined entirely by service prices. Similarly, if there are few differences in service prices across markets, $SPI_{d,t}$, is near 1, and the $MCE_{d,t}$ will be entirely determined by utilization.

3 Data

We use retrospective claims data for a sample of commercially insured patients from the MarketScan® Research Database from Truven Health. The specific claims data used are the Commercial Claims and Encounters Database, which contains data from employer and health plan sources containing medical and drug data for several million commercially insured individuals, including employees, their spouses, and dependents. Each observation in the data corresponds to a line item in an "explanation of benefits"

[8] This approach follows others in the health literature that also apply additive decompositions (e.g., Roehrig and Rousseau (2011) and Rosen et al. (2013)), which leaves a cross-term. As another possibility, we could have used a Laspeyres index for the price index and a Paasche index for the quantity index, which provides an exact decomposition (e.g., $SUI^{Laspeyres} \cdot SPI^{Paasche} = MCE$). These alternative estimates are included in the Online Appendix. It is also worth noting that the alternative Paasche index may be computed from the reported estimates: $SPI^{Paasche} = \frac{MCE}{SUI^{Laspeyres}}$. A national base was selected, since it is intuitive to think about the "typical individual" in the United States, rather than the diseases that are observed in a particular MSA.

form; therefore each claim can consist of many records and each encounter can consist of many claims.

We use a sample of enrollees that are not in capitated plans from the MarketScan database for the years 2003 to 2007. The MarketScan database tracks claims from all providers using a nationwide convenience sample of enrollees. Each enrollee has a unique identifier and includes age, sex, and region information that may be used when calculating patient weights. All claims have been paid and adjudicated. While the full MarketScan database has been growing substantially due to the addition of data contributors, we focus on a subset of the data that is provided by the same contributors in each year, which limits potential changes caused by new or exiting data contributors (see Dunn et al. (2012b) and Dunn et al. (2012c) for additional discussion). We also limit our sample to enrollees with drug benefits because drug purchases will not be observed for individuals without drug coverage.

The claims data have been processed using the Symmetry grouper from Optum. The grouper assigns each claim to a particular Episode Treatment Group (ETG) disease category.[9] The grouper uses a proprietary algorithm, based on clinical knowledge, that is applied to the claims data to assign each record to a clinically homogeneous episode. The episode grouper allocates all spending from individual claim records to a distinct condition; the grouper also uses other information on the claim (e.g., procedures) and information from the patient's history to allocate the spending. An advantage of using the grouper is that it can use patients' medical history to assign diseases to drug claims, which typically do not provide a diagnosis. However, these algorithms are also considered a "black box" in the sense that they rely entirely on the grouper software developer's expertise. The ETG Symmetry grouper is applied to one calendar year of data at a time. Although this limits the amount of information used for each person (since we often

[9] The ETG grouper allocates each record into one of over 500 disease groups. To ensure that we observe full episodes, we limit the sample to those enrollees that have a full year of continuous enrollment. In addition, we require that enrollees have one year of enrollment in the prior year and one year of enrollment in the following year to make sure that episodes occurring at the beginning or the end of a year are not truncated. This may be an overly conservative constraint on the sample of enrollees, and we are currently working on examining the sensitivity of our analysis to alternative assumptions on enrollment.

observe multiple years), it also avoids potential biases that may occur if the grouper is not applied symmetrically across all years (see Dunn et al. (2012d)).

Population weights are applied to each individual to adjust for differences in age, sex, and region across populations, so the expenditure estimates may be comparable across years (see Dunn et al. (2012c)).[10] Our aim is to make these data representative of the entire commercially insured population, but it is important to remember that the data comes from primarily large employers, which may represent a distinct population.[11]

To better control for the severity of the diagnosis, we incorporate additional severity measures provided by the ETG grouper to further classify each episode. The availability of severity classifications vary by the ETG disease category, and range from 1 (the least severe) to 4 (the most severe). For instance, the most severe case of diabetes will be given a severity level of 4 while the least severe case will be given a severity level of 1. The ETG severity level is determined for each episode based on a variety of additional information including age, gender, comorbidities, and other potential complications.[12]

3.1 Service Price, Utilization, and Episodes

The number of episodes is a simple count of the total number of episodes of a medical disease that end in the sample period. Total episode expenditures are measured as the total dollar amount received by all providers for the services used to treat an episode of a specific disease (including both out-of-pocket payments and amounts paid by insurance firms). The expenditure information is based on actual payments, not provider charges.

Service utilization measures are created for each type of service based on the definition of a service within that service type. The service-type categories are inpatient hospital, outpatient hospital, general physician, specialist physician, prescription drug, and other. Measuring service utilization is not a straightforward task since the definition of "service"

[10] Specifically, using the enrollment data in each MSA, weights are applied to different age and sex categories so that the total enrollment files match the population for commercially insured individuals in the U.S. for 2007. Information on the population is obtained from the Current Population Survey.

[11] Although this is a potential problem, Dunn et al. (2012c) show that the expenditure growth and service price growth in our weighted sample appear to match national estimates.

[12] If this severity adjustment does not appropriately account for the change in severity over time, then the estimates may be biased.

is a bit ambiguous and there are a variety of ways that one could define it across various service types. Ideally, we would like the definition of a specific service to depend on how the price of that service is typically set and paid. For example, for physician services, individuals pay a unique price for each procedure done to them (that is, the insurer and the patient together pay this amount), whereas the prices paid to facilities are often set based on the treated disease. Next we describe how the quantity of services is measured for each service type.

3.1.1 Measuring the Quantity of Service by Service Type

For each claim line in the data, we first categorize it by place of service, which determines the service-type category. For each category, the following steps describe how the amount is determined for each encounter, where a encounter is defined by the enrollee and the date of service or admission:

Physician office - Physician services are priced based on procedures performed in a physician's office. Since not all procedures are equivalent, each procedure is weighted to reflect the intensity of the service. For the Medicare payment system, Relative Value Units (RVUs) define reimbursement rates and are intended to capture the intensity of the services provided. In that spirit, we proxy for the intensity of service by using the average prices for each Current Procedural Terminology (CPT-4) code and modifier code. The total quantity of services performed in an office is then computed by summing over these RVU amounts. More precisely, the total amount of services performed during an encounter at a physician office is computed as $q_{office} = \sum_{cpt \in Encounter} \overline{p}_{cpt,office}$, where $cpt \in Encounter$ is a complete list of CPT procedures performed during the encounter in an office setting and $\overline{p}_{cpt,office}$ is the base price for procedure code, cpt. The base group price, $\overline{p}_{cpt,office}$, is computed as the average price in the data for that procedure code and modifier code. Since most insurers set prices from a base price schedule (e.g., 10 percent above Medicare rates), one can think of the price level as the base price multiplied by a scalar price, α_t, where $p^t_{cpt} = \alpha_t \overline{p}_{cpt}$. For instance, if a CPT code indicates an X-ray has an average price of \$100, its value will be 100 RVUs (i.e., $\overline{p}_{99213} = 100$). It should be clear that the RVU amount is a measure of utilization and not price. To see this, if the fee on an X-ray is \$120 in time period t ($p^t_{99213} = \$120$), then the price of the service will be calculated as \$120/100RVU=1.2 \$/RVU (i.e., $\alpha_t = \frac{p^t_{cpt}}{\overline{p}_{cpt}}$).

Hospital inpatient - Inpatient hospital stays not only consist of facility fees paid to the hospital, but also fees paid to the physician. A variable in the claims data distinguishes these two types of payments. For the portion of fees paid to the hospital, the amount of services is measured as the average dollar amount for an inpatient stay for the observed disease. For the portion of fees paid to the physician, we assign an RVU in the same way that we calculate an RVU in an office setting. The total amount of services performed in an inpatient setting is calculated by adding the physician and facility amounts. Specifically, $q_{inpatient} = \overline{p}_{d,inpatient} + \sum_{cpt \in Encounter} \overline{p}_{cpt,inpatient}$ where $\overline{p}_{d,inpatient}$ is the base price for inpatient facility claims for disease d, where the base price is the average price in the data for an encounter (i.e., admission) to an inpatient facility for treating disease d. The term $\sum_{cpt \in Encounter} \overline{p}_{cpt,inpatient}$ is the amount calculated for the physician portion of the bill and is computed in a manner identical to the physician office category, but is based on only physician claims in an inpatient setting.

Hospital outpatient - Outpatient hospital services are calculated in an identical fashion to the inpatient hospital services. That is, the facility amount is calculated based on the average expenditure per encounter on outpatient services for that disease, and the doctor's portion of the total amount is calculated based on the average payment for the procedure codes in an outpatient setting.

Prescription drugs - The amount of the prescription drug varies based on the molecule, the number of pills in the bottle, the strength of the drug, and the manufacturer. An 11-digit National Drug Code (NDC) uniquely identifies the manufacturer, the strength, dosage, formulation, package size, and type of package. To capture these differences, we calculate the average price for each NDC code. This means we treat branded and generic products that contain the same active molecule as distinct drugs. The average price for each NDC code represents the amount of the service used. Specifically, the amount of drug services used is $q_{drug} = \sum_{NDC \in Encounter} \overline{p}_{NDC}$, where $NDC \in Encounter$ is a complete list of NDC codes purchased from an encounter at a pharmacy and $\overline{p}_{NDC}$ is the base price for a specific NDC code. The base price for each NDC is computed as the average price in the data.

All other - The other category primarily includes ambulatory care, independent labs, and emergency room visits. For these services, if no procedure code is available, the amount of each category is measured as the average cost for an encounter to that particular place of service for treating a particular disease (for example, the average cost of an ambulatory care visit to treat ischemic heart disease). For cases where procedure codes are available, we use the average cost of that procedure code for that place of service.

Our decomposition relies on the institutional feature that insurers and providers typically negotiate from a percentage of a base fee schedule (for example, 10 percent above Medicare rates).[13] As our measure of service price can be intuited as the expenditures from an encounter divided by a proxy for a "RVU", it can also be thought of as a percentage amount from a base (or average) payment—a measure close to how prices are actually set. For this reason, these measures of service quantity subsequently allow us to create service prices that correspond well with how fees are negotiated in the marketplace. In other words, our approach attempts to construct a unit value index that reflects the heterogeneity in how goods and services are actually priced. It can also be shown that if pricing is set based on a percentage of a set fee schedule then our index is equivalent to an index that prices specific procedures. See Dunn et al. (2013) and the associated technical appendix for additional details.

3.2 Descriptive Statistics

Table 1 reports descriptive statistics for the MarketScan data. Prior to computing these statistics, sample weights were applied, so that the enrollment counts and the age and sex distribution are fixed to 2007 levels. Although the demographics of the population are held constant, we see expenditure per capita grow from $2,583 to $3,229, an increase of 25 percent. This growth is due to both an increase in the number of episodes, from 3 to 3.2 per enrollee, but also due to growth in expenditures per episode, which increased about 15 percent. For purposes of this paper, we are particularly interested in the bottom portion of Table 1, which shows how expenditure shares have shifted across service categories. The table shows expenditures have shifted away from inpatient hospital services and

[13]In a survey of 20 health plans conducted by Dyckman & Associates, all 20 health plan fee schedules were influenced by the Medicare fee schedule. That is, a resource-based relative value scale (RBRVS), essentially adopting Medicare's base fee schedule.

toward physician offices and other service categories. We also see a shift away from branded drugs toward generic drugs. We focus on the implications for these types of shifts on price growth measures. In particular, we examine whether these service shifts lead to differences in disease price and service price indexes.

Table 1. Summary Statistics

	2003	2007
Expenditure (Billions $)	$471.5	$589.3
Enrollees (Millions)	182.5	182.5
Expenditure per Capita	$2,583	$3,229
Number of Episodes (Millions)	3.0	3.2
Expenditures per Episode	$158	$182
Expenditure Share		
Inpatient Hospital	28.1%	26.2%
Outpatient Hospital	24.1%	24.2%
Office	22.2%	22.9%
Other	9.8%	10.9%
Brand Drugs	13.1%	11.6%
Generic Drugs	2.7%	4.3%

4 Results

Recall from the methodology section that there are two ways for the MCE index to be divided into service price and service utilization components, an encounter-based approach and a procedure-based approach. Table 2 reports the results for these two types of decompositions. The left-hand side of Table 2 shows the aggregate MCE index for each year from 2003 to 2007. The next two columns show the procedure-based decomposition. The procedure-based measure shows the prices of the underlying services grow at a pace that is very close to the growth in the MCE index. In addition, it shows that the utilization per episode declines only slightly (just 0.4 percent) over the period of study. When conducting the encounter-based decomposition, we noticed that outliers had a larger influence on measurements. Therefore, when comparing the procedure-based and encounter-based approach we focused on only those diseases with 10,000 or more episodes in the right half of Table 2.[14] Using this more limited sample, we see that

[14]This sample selection drops about 9 percent of expenditures.

the procedure based decomposition results are practically the same. In contrast, using the encounter-based measure, we find that the SPI grows very rapidly, by 25.3 percent, while the utilization falls quickly, by 8 percent. Therefore, the SPI-encounter greatly overstates inflation relative to the MCE, while the growth rate in the SPI-procedure is quite close to the MCE.[15]

Table 2. Components of Episode Expenditure Growth

	All Disease Categories			Disease Categories with 10,000 or More Episodes				
		Procedure-Based			Procedure-Based		Encounter-Based	
Year	MCE	SPI-proc.	SUI-proc.	MCE	SPI-proc.	SUI-proc.	SPI-enco.	SUI-enco.
2003	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2004	1.052	1.056	0.998	1.050	1.053	0.998	1.076	0.988
2005	1.075	1.087	0.993	1.072	1.083	0.993	1.128	0.959
2006	1.121	1.126	1.002	1.115	1.120	1.002	1.200	0.945
2007	1.139	1.159	0.996	1.132	1.152	0.995	1.253	0.920

Notes: We found that the encounter-based approach is more influenced by outliers than the procedure-based approach when measuring service prices. To overcome this problem, encounter-based and procedure-based estimates are only conducted on diseases with more than 10,000 episodes in the data. The basic findings shown above hold for numerous alternative specifications and samples, as outlined in greater detail in the robustness section of this paper.

To better understand the relationship among these indexes, an alternative decomposition of the MCE growth for the 2003-07 period is shown in Table 3. The top of the table shows the total growth in the MCE index from 2003-07, which is 13.2 percent. As earlier, this amount may be decomposed into a price component that increased by 15.2 percent and a utilization component that fell by 0.5 percent. To connect the procedure-based approach to the encounter-based approach, one may think of the service utilization component as being composed of two distinct parts. One piece is encounters per episode, which is the SUI-encounter measure that has declined by 8.0 percent. The second piece is the amount of RVUs per encounter, which measures the intensity of treatment for each encounter.[16] The RVUs per encounter have grown by 7.5 percent, implying more

[15] We find that when we adjust for severity and use the sample of constant data contributors (as in Dunn et al. (2012b)), the encounter-based approach produces relatively noisy estimates. Therefore, the decomposition presented in this paper shows only those diseases with at least 10,000 observations in the data. It is worth noting that if we do not apply severity adjustment and use the full MarketScan sample, as in Dunn et al. (2012a), we obtain very similar results to those presented here.

[16] RVUs per encounter is calculated as the difference between the SUI-encounter index and the SUI-procedure index.

intense treatments per encounter over time.[17] In other words, encounters per episode are falling, but the intensity of treatment per encounter is rising causing service utilization (i.e., SUI-procedure) to remain almost constant. Consequently, the SPI-procedure and the MCE grow at similar rates.

Table 3. Accounting for the Change in Procedure-Based and Encounter-Based Indexes from 2003-07

Index	Percent Change
MCE	**13.2%**
SPI-procedure	**15.2%**
SUI-procedure	**-0.5%**
Encounters per Episode (SUI-Encounter)	-8.0%
"RVUs" per Encounter	7.5%
Cross Term	-1.5%

Taking another look at the discrepancy between the procedure-based and encounter-based measures, Table 4 breaks out the components of the MCE growth by service category.[18] The difference between the procedure-based and encounter-based measures of price and utilization is large across several service categories. For instance, the table shows that the service price index for physician office services grows by 8.4 percent using the procedure-based index, but grows by 15.7 percent using the encounter-based index. Together these figures imply that the intensity of treatment per encounter at the physician office has grown by about 7.3 percent. Similarly, we see large differences for both branded and generic drugs, suggesting a utilization shift for each prescription filled (e.g., individuals are purchasing larger bottles or greater strength pills).[19]

[17] Note that if the intensity of services per encounter did not change, then there would be no difference between the encounter-based and procedure-based measures.

[18] Here, the MCE, SPI, and SUI are decomposed as above, but the focus is only a single service category s. For example, $SPI_{d,t,s} = \frac{p_{d,t,s}q_{d,0,s}}{c_{d,0,s}}$. When we aggregate over disease d, we weight by the expenditure share for disease d for place of service s.

[19] It should be noted that in prior work Aizcorbe and Nestoriak (2011) did not separately price pharmacy encounters for branded and generic drugs. We find that when we do not price them separately, the SPI-encounter grows more slowly, but still considerably faster than the SPI-procedure index.

Table 4. Episode Expenditure Growth by Service - 2003-07

		Procedure-Based		Encounter-Based	
	MCE	SPI-proc.	SUI-proc.	SPI-enco.	SUI-enco.
Inpatient Hospital	1.054	1.203	0.878	1.273	0.842
Outpatient Hospital	1.136	1.137	0.997	1.164	0.974
Office	1.166	1.084	1.077	1.157	1.008
Other	1.272	1.167	1.094	1.207	1.057
Brand Drug	0.980	1.242	0.784	1.569	0.628
Generic Drug	1.762	0.867	2.058	1.258	1.387

A key reason for estimating a procedure-based index is that it more closely follows how services are actually priced in the marketplace. Another important advantage of this approach is that our estimates are more comparable to the BLS price indexes, allowing us to better evaluate any possible discrepancy in official statistics with a cost-of-treatment type measure. First, we must check how the procedure-based price index compares in value to national official price statistics. To make this first comparison, we turn to an overall health-care price measure, Bureau of Economic Analysis (BEA) personal consumption expenditure deflator for health care (PCE deflator). This index grows by 13.7 percent over the 2003 to 2007 time period, which is closer to the SPI-procedure measure (growth of 15.2 percent), relative to the encounter-based measure (growth of 25.3 percent).[20] Our SPI-procedure and PCE deflator provide two independent estimates constructed in a similar manner and arrive at similar rates of inflation. This similar rate of growth helps to substantiate the growth rate reported by BEA and confirms that our estimate falls in a reasonable range. Comparing the PCE deflator to the MCE index, we find that MCE index growth (i.e., 13.2 percent) is only slightly lower than growth in the PCE for health care, indicating no difference between these two aggregate price statistics.

[20] For hospital services, our figures show lower price growth than the BLS figures. This is because we combine professional and hospital services for services conducted at a hospital, as in Aizcorbe and Nestoriak (2011). If we were to separate these components, the hospital prices in our data would be closer to the price growth reported by the BLS.

Table 5. Benchmark Price Growth Measures - 2003-07

	Service Price Growth
Overall, PCE health care deflator - BEA	1.137
Hospital - BLS PPI	1.176
Hospitals (Non-Medicare & Non-Medicaid) - BLS PPI	1.211
Physician Offices - BLS PPI	1.091
Pharmaceutical Drugs (Branded and Generic) - BLS PPI	1.172

Next, we compare our indexes with official price indexes for specific service categories. We find that the procedure-based measure tends to be much closer to the corresponding BLS indexes for each service category. For instance, physician office services grow by 8.4 percent based on the SPI-procedure and the corresponding producer price index (PPI) from the BLS grows by 9.1 percent. In contrast, the SPI-encounter for physician offices grows by 15.7 percent. Next we turn to the drug price indexes. Branded drugs account for a greater share of expenditures in 2003 (13.1 percent from Table 1, compared to 2.7 percent for generics), so that a weighted average price growth is around 18.4 percent, which is quite similar to price growth for the BLS. The encounter-based approach shows considerably faster drug price growth for both branded and generic drugs.[21]

For hospital price growth, the comparison is less straightforward. To be consistent with Aizcorbe and Nestoriak (2011), we combine professional and hospital payments, while the BLS includes only hospital facility payments. To provide a more direct comparison, we calculate additional price index measures for hospital services using only the facility component of the payment. For the procedure-based approach, we find that the facility component of the price grows by 25.6 percent for the inpatient price and 17.4 percent for the outpatient price, so the average is about 22 percent, which is quite close to the BLS PPI.[22] A very similar growth rate is found looking at facility payments using

[21] Note that both the encounter-based measure and the procedure-based measure do not average over branded and generic drugs of the same molecule type, as is the current practice of the BLS. Therefore, one should expect both of these measures to have some positive bias relative to the BLS measure. We did not replicate this aspect of the BLS price index due to insufficient information on how specific molecules may be matched.

[22] In our sample, around 53 percent of hospital facility expenditures were for inpatient services in 2003.

the encounter-based approach. Overall, the primary difference between the encounter-based approach and procedure-based approach appears to be for professional services and prescription drugs.

4.1 A Decomposition of MCE and SPI-procedure Differences by Disease

To see how results vary by disease category, we report growth rates for the 2003-07 period by Major Practice Category (MPC) in Table 6,[23] applying both the procedure-based and encounter-based decompositions. One can see that the SPI-encounter tends to rise faster than the SPI-procedure for nearly every category, suggesting that the intensity of services per encounter is growing for all MPCs.

In contrast to the aggregate results, we find that the SPI-procedure and MCE do not have similar growth rates across all disease categories. That is, the finding that the MCE and SPI-procedure measures move together in the aggregate does not hold at the MPC level. For some categories, such as cardiology, we see the MCE index grow more slowly than the SPI-procedure, while for other categories, such as orthopedics, we see the MCE index grow more rapidly. These differences have implications for productivity growth. For instance, these results suggest that the change in real output in treating cardiology conditions, measured using an MCE, is greater than what is implied by the service price index; meanwhile, for orthopedics, the change in real output, as measured by the MCE, is actually less than what is implied by the service price index.

[23] The Major Practice Category is a categorization of ETG disease episodes into related disease groups, as defined by Ingenix.

Table 6. Components of Disease Category Growth, 2003-07

			Procedure-Based		Encounter-Based	
Major Practice Category	Expenditure (in Billions)	MCE	SPI-proc.	SUI-proc.	SPI-enco.	SUI-enco.
Orthopedics & rheumatology	$73.29	1.158	1.144	1.032	1.262	0.946
Cardiology	$54.19	1.056	1.154	0.928	1.236	0.863
Gastroenterology	$37.21	1.132	1.147	0.997	1.220	0.933
Gynecology	$31.91	1.202	1.191	1.012	1.258	0.958
Endocrinology	$29.79	1.067	1.157	0.932	1.284	0.854
Otolaryngology	$29.44	1.108	1.135	1.001	1.238	0.915
Neurology	$23.92	1.184	1.202	0.993	1.302	0.915
Psychiatry	$21.14	1.114	1.147	1.008	1.295	0.885
Pulmonology	$20.35	1.167	1.203	0.976	1.290	0.912
Dermatology	$20.21	1.179	1.155	1.038	1.357	0.938
Obstetrics	$19.45	1.175	1.148	1.026	1.213	0.972
Urology	$16.01	1.126	1.149	0.990	1.229	0.924
Preventive & administrative	$10.91	1.261	1.141	1.106	1.312	0.999
Hepatology	$9.48	1.122	1.163	0.968	1.219	0.926
Ophthalmology	$7.60	1.054	1.082	0.980	1.115	0.949
Nephrology	$5.84	0.876	0.869	1.013	1.047	0.850
Hematology	$5.17	1.049	1.166	0.905	1.213	0.874
Neonatology	$4.76	1.153	1.111	1.033	1.143	1.004
Infectious diseases	$4.19	1.142	1.116	1.025	1.234	0.949
Isolated signs & symptoms	$3.41	1.112	1.102	1.020	1.220	0.939
Late effects, environ. trauma & poisonings	$2.57	1.339	1.295	1.044	1.365	0.994
Chemical dependency	$2.09	1.062	1.098	0.988	1.098	0.979

The faster growth in the MCE (shown for some MPC categories), relative to the SPI-procedure, implies that inflation using the SPI-procedure is understated. However, this assessment hinges on the assumption that quality is fixed, which is probably not the case. For instance, the MCE may be rising more quickly than the SPI-procedure for orthopedic conditions because more technologically advanced treatments are being applied. Future research may entail determining whether this pattern is attributable to technological changes and whether these changes affect quality. These questions are crucial to policy makers, consumers, and other industry participants, but fall outside the scope of this study. Instead, here we present only the changes in the cost of treatment (i.e., the disease price), which we view as an important first step in a more complete analysis of changing productivity in the health sector.

Whether quality is fixed or changing, it will be important to understand how utilization shifts drive a wedge between the SPI-procedure and MCE indexes. To better analyze this wedge, we apply an additional decomposition that reports the difference between the SPI-procedure and MCE index by service-type, s. We follow the decomposition approach outlined in Aizcorbe and Nestoriak (2011), which we adapt to the procedure-based approach. The decomposition equation is

$$MCE_{d,t} = SPI_{d,t}+(MCE_{d,t}-SPI_{d,t}) = SPI_{d,t}+\sum_{s}(MCE_{d,t,s}-SPI_{d,t,s}) \left(\text{Expenditure Share}_{d,0,s}\right) \tag{8}$$

$$= SPI_{d,t} + \sum_{s}(MCE_{d,t,s} - SPI_{d,t,s}) \left(\frac{q_{d,0,s} \cdot p_{d,0,s}}{\sum_{s} q_{d,0,s} \cdot p_{d,0,s}} \right).$$

The term $(MCE_{d,t,s} - SPI_{d,t,s}) \left(\text{Expenditure Share}_{d,0,s}\right)$ represents service category s's contribution to the difference between the MCE and SPI index. To gain some additional intuition for this equation, we substitute $MCE_{d,t,s} - SPI_{d,t,s}$ with the approximation $SUI_{d,t} - 1 \approx MCE_{d,t,s} - SPI_{d,t,s}$, which is taken from the decomposition (8) but removes the cross term. After substituting, the decomposition by service category is $MCE_{d,t} \approx SPI_{d,t} + \sum_s (SUI_{d,t,s} - 1) \left(\text{Expenditure Share}_{d,0,s}\right)$. From this approximate decomposition, one can see that the difference between the two indexes will primarily depend on the change in utilization of the different services and the corresponding expenditure share of the service category. Table 7.1 shows the contribution of each service-type, s, to the difference between the MCE and SPI-procedure (applying the exact decomposition 8).

Table 7.1 shows several clear patterns across services. First, for nearly every disease category, there is a shifting away from spending on inpatient services. This is consistent with the results of Aizcorbe and Nestoriak (2011) and Dunn et al. (2012a), who show that substitution away from inpatient services generally leads to a lower MCE relative to an SPI. This savings from reduced utilization on inpatient services is partly offset by a strong increase in the utilization of physician services for most disease categories. For drug services, we observe a shifting away from branded drugs, leading to a relative decline in the MCE, and we see an increase in generic drugs, contributing to an increase in the MCE. Combined, the shifting away from branded drugs toward generics causes a net decline in the MCE relative to the SPI-procedure for most disease categories.[24] Also note that there is a positive shift toward "Other" services for many of the MPC categories. We conducted additional analysis to better understand what is happening in the "Other" category. Specifically, we broke out the positive shifts in the "Other" category in Table 7.1 into more disaggregate service categories (i.e., emergency room care, ambulatory surgical centers, and independent laboratory) and we noticed a couple of patterns. For many of the key instances (i.e., Orthopedics, Gastroenterology, Otolaryngology, and Ophthalmology), the positive shift in the "Other" category is due to a growth in services at ambulatory surgical centers. For "Urology" and "Late effects, environ. Trauma & pois.", the positive shift in the "Other" category is toward hospital emergency room services.

Table 7.1 Comparison of MCE and SPI-procedure Price Indexes and Sources of Differences, 2003 - 2007

		Contribution to MCE-SPI-procedure difference					
Major Practice Category	Diff. MCE-SPI-proc	Inpat. Hospital	Outpat. Hospital	Office	Other	Brand Drug	Generic Drug
Orthopedics & rheumatology	0.014	-0.027	-0.002	0.040	0.021	-0.038	0.020
Cardiology	-0.097	-0.100	-0.006	0.017	-0.001	-0.030	0.022
Gastroenterology	-0.015	-0.035	-0.018	0.010	0.045	-0.034	0.017
Gynecology	0.011	-0.046	0.029	0.034	0.002	-0.017	0.009
Endocrinology	-0.090	-0.091	-0.003	0.001	0.008	-0.054	0.049
Otolaryngology	-0.026	-0.010	-0.005	0.011	0.014	-0.089	0.052
Neurology	-0.018	-0.035	-0.001	0.014	0.000	-0.028	0.032
Psychiatry	-0.033	-0.004	0.006	-0.014	-0.011	-0.096	0.086
Pulmonology	-0.036	-0.056	0.004	0.009	0.007	-0.018	0.019
Dermatology	0.024	-0.010	0.000	0.011	0.006	-0.029	0.047
Obstetrics	0.026	0.002	0.004	0.016	0.001	-0.002	0.005
Urology	-0.023	-0.046	0.001	0.024	0.015	-0.034	0.017
Preventive & administrative	0.120	-0.010	0.009	0.113	0.007	-0.001	0.002
Hepatology	-0.041	-0.022	0.009	0.002	0.006	-0.049	0.013
Ophthalmology	-0.028	-0.009	-0.040	0.016	0.018	-0.016	0.003
Nephrology	0.006	-0.009	0.013	0.003	0.009	-0.005	-0.005
Hematology	-0.117	-0.078	-0.009	-0.011	0.001	-0.028	0.007
Neonatology	0.042	0.062	-0.002	-0.007	-0.010	0.001	-0.001
Infectious diseases	0.025	0.000	0.000	0.001	0.001	0.013	0.011
Isolated signs & symptoms	0.009	-0.011	-0.011	0.014	0.004	-0.018	0.031
Late effects, environ. trauma & pois.	0.045	-0.063	0.040	0.008	0.045	0.007	0.007
Chemical dependency	-0.036	-0.061	0.045	-0.014	-0.033	0.019	0.007

Table 7.2 is identical to Table 7.1, but shows the decomposition for the difference between the MCE and $SPI - Encounter$ indexes. One of the main differences be-

[24]To provide a more complete picture of the various components of spending growth, Tables A1.1, A1.2 and A1.3 in the Appendix report the changes in the MCE, SPI, and SUI, respectively, by service type for the top five diseases.

tween Tables 7.1 and 7.2 is from the physician office category, where $(MCE_{d,t,s} - SPI_{d,t,s})\left(\text{Expenditure Share}_{d,0,s}\right)$ tends to be larger using the procedure-based approach. We also see that the difference for branded drugs tends to be more positive using the procedure-based approach.

Table 7.2 Comparison of MCE and SPI-Encounter Price Indexes and Sources of Differences, 2003 - 2007

		Contribution to MCE-SPI-encounter difference					
Major Practice Category	Diff. MCE-SPI-encounter	Inpat. Hospital	Outpat. Hospital	Office	Other	Brand Drug	Generic Drug
Orthopedics & rheumatology	-0.103	-0.044	-0.008	0.024	0.001	-0.084	0.007
Cardiology	-0.180	-0.116	-0.008	0.000	0.001	-0.067	0.010
Gastroenterology	-0.088	-0.048	-0.023	0.000	0.033	-0.057	0.007
Gynecology	-0.056	-0.057	0.016	0.008	0.004	-0.031	0.005
Endocrinology	-0.217	-0.097	-0.005	-0.004	0.009	-0.165	0.045
Otolaryngology	-0.130	-0.013	-0.009	-0.013	0.008	-0.139	0.035
Neurology	-0.118	-0.048	-0.003	-0.001	-0.001	-0.075	0.012
Psychiatry	-0.181	-0.005	-0.002	-0.019	0.004	-0.216	0.057
Pulmonology	-0.124	-0.078	-0.002	-0.004	0.009	-0.057	0.008
Dermatology	-0.178	-0.082	-0.001	-0.007	0.008	-0.117	0.021
Obstetrics	-0.038	-0.058	0.007	0.010	0.003	-0.001	0.000
Urology	-0.103	-0.057	-0.008	-0.001	0.008	-0.054	0.008
Preventive & administrative	-0.051	-0.062	0.005	0.008	0.001	-0.006	0.003
Hepatology	-0.096	-0.044	0.004	-0.001	0.007	-0.065	0.003
Ophthalmology	-0.060	-0.011	-0.041	0.001	0.013	-0.022	0.000
Nephrology	-0.171	-0.010	-0.130	0.000	-0.012	-0.018	0.000
Hematology	-0.164	-0.091	-0.015	-0.014	0.003	-0.050	0.002
Neonatology	0.010	0.020	-0.003	0.000	-0.008	0.000	0.000
Infectious diseases	-0.092	-0.008	-0.002	-0.004	0.002	-0.089	0.007
Isolated signs & symptoms	-0.108	-0.019	-0.014	-0.003	0.002	-0.090	0.017
Late effects, environ. trauma & pois.	-0.026	-0.085	0.039	-0.003	0.045	-0.025	0.004
Chemical dependency	-0.036	-0.070	0.030	-0.018	0.001	0.018	0.003

4.2 Robustness Checks

The results presented thus far have looked at just one approach for measuring the SPI-procedure and MCE, so additional analysis is necessary to check whether these results hold up to further scrutiny. To briefly recap the methodology, we decompose price and utilization by first constructing a measure of utilization that reflects the intensity of services given to an individual. Recall the utilization measure is defined to capture how a good is typically priced, such as by procedure for physician services. Next, we use the measure of utilization to calculate the price for each service category by disease. Although this approach is arguably very close to the BLS method of pricing medical services, it is distinct. In particular, we aggregate across CPT codes to estimate an amount of services, q_{office}, that is used to calculate price. In contrast, the BLS prices specific CPT codes and holds the quantity of each CPT code fixed. If providers are pricing on a percentage of a typical fee schedule or Medicare prices (which we believe is quite common), then these two approaches will yield very similar results.[25] If not, then the price estimates from these two methods could potentially diverge. In any case, it may be useful to estimate an alternative service price index that more closely follows the BLS methodology.

As a robustness check, we track the price by specifically defined service. For instance, we price each individual CPT code plus modifier code for physician office services by calculating average price changes for each, and we use the expenditure share of each of those precisely defined services in the base period to construct our service price index. A very similar approach was applied in Bundorf, Royalty, and Baker (2009), who also looked at price growth for the commercial sector. Applying this alternative methodology, the price growth measurements by service type are shown in Table 8. The price trends in Table 8 match quite closely with the price trends using our RVU methodology, although slightly higher. These results both confirm the robustness of the RVU methodology applied in this paper, but also show that the MCE index is quite close to an alternative service price index measure that entirely ignores the disease mix.[26] However, it is

[25] Let the the price and quantity for CPT code *cpt* in time period t be denoted $P_{cpt,t}$ and $Q_{cpt,t}$. In this case, the Laspeyres price index for time period t for physician services may be computed as:

$\text{SPI}_{Lasp} = \frac{P_{1,t}\cdot Q_{1,0}+P_{2,t}\cdot Q_{2,0}...+P_{N,t}\cdot Q_{N,0}}{P_{1,0}\cdot Q_{1,0}+P_{2,0}\cdot Q_{2,0}...+P_{N,0}\cdot Q_{N,0}}$. Assuming that physicians change prices from a base fee schedule, then the prices in time t can be computed as α_t times the base fee schedule. That is, $P_{1,t} = \alpha_t\ P_{1,0}$, $P_{2,t} = \alpha_t\ P_{2,0}, ...,$ and $P_{N,t} = \alpha_t\ P_{N,0}$, so

$\text{SPI}_{Lasp} = \frac{P_{1,t}\cdot Q_{1,0}+P_{2,t}\cdot Q_{2,0}...+P_{N,t}\cdot Q_{N,0}}{P_{1,0}\cdot Q_{1,0}+P_{2,0}\cdot Q_{2,0}...+P_{N,0}\cdot Q_{N,0}}$
$= \frac{\alpha_t(P_{1,0}\cdot Q_{1,0}+P_{2,0}\cdot Q_{2,0}...+P_{N,0}\cdot Q_{N,0})}{P_{1,0}\cdot Q_{1,0}+P_{2,0}\cdot Q_{2,0}...+P_{N,0}\cdot Q_{N,0}} = \alpha_t$. In this example, our index is the same as a price index that tracks prices at the procedural level. Of course, to the extent that physicians price procedures individually, rather than based on a schedule, this result would not hold.

[26] Constructing the alternative BLS-type service price index is informative, but one should also note some of the advantages of the RVU pricing methodology applied in this paper. One advantage of the RVU approach is that it allows for unique trends by disease, so that cardiologist price trends may differ from those of orthopedic doctors. In contrast, it may be challenging to construct disease-specific service prices using a BLS-type methodology, since there are thousands of procedure and drug codes, but it is

also worth noting that the service price growth reported here is slightly larger than the MCE, both using the RVU methodology and based the fixed basket of services, reported in Table 8. As we explore alternative robustness checks, we find that this slight difference does not necessarily hold when alternative methodologies are applied.

Table 8. Service Price Growth Measures - Fixed Basket - 2003-07

	Expenditure Share	Service Price Growth
Overall	100.0%	1.166
Inpatient Hospital	28.1%	1.239
Outpatient Hospital	24.1%	1.115
Office	22.2%	1.067
Other	9.8%	1.234
Brand Drug	13.1%	1.303
Generic Drug	2.7%	0.763

We conduct a number of additional robustness checks to investigate how changing various aspects of our analysis may affect our results. This additional analysis is conducted in two companion pieces to this paper.[27] Overall, the robustness checks provided

likely impossible to observe sufficent observations for each disease to price each CPT code and drug. It is possible to price by disease using the RVU methodology because we exploit the fact that providers typically price based on a percentage of a fee schedule. In this sense, there is only a single price that is relevant, the percent deviation from the fee schedule.

[27] Dunn et al. (2012d) investigate how applying different methodologies to allocate expenditures to disease episodes may affect the various components of expenditure growth. Specifically, we analyze disease decompositions by applying different grouper software, including the ETG Symmetry grouper from Optum (used here) and the Medical Episode Grouper (MEG) from Truven Health. We also explore how different ways of running each grouper or defining disease episodes may affect these results. For instance, we compare the results when severity adjustments are applied with alternative estimates when severity adjustments are not applied. We also explore alternatives that do not rely on grouper algorithms. For instance, we use the primary diagnosis and also apply regression techniques to allocate spending across disease categories. In all cases, the basic qualitative findings presented here appear to hold. Namely, the MCE grows at about the same rate as the SPI (either the SPI-procedure or the alternative procedure-based SPI presented here) and the BLS price index. It is worth noting that when we apply the MEG grouper, we actually see the MCE grow slightly faster than the SPI.

In another paper, Dunn et al. (2012c), we study how different weights or sampling strategies may by these two companion papers support the main results presented here. The appendix reports some robustness findings. Table A2 shows how decompositions from Table 2 change when various alternative methodologies are applied. For instance, when we do not adjust for severity or when we apply different population weights. In general, the main findings change very little.

5 Conclusion

Shifts in technologies and protocols used to treat diseases could drive a wedge between the types of service price indexes that the BLS and BEA currently report and the disease prices, which reflect what individuals actually pay for treatment. Given the known shifts in utilization in the health sector, many health economists have advocated for tracking and reporting disease prices that arguably provide a more meaningful measure of inflation.

To investigate if these utilization shifts lead to a different rate of inflation than official price indexes for medical care, we compare growth rates from an aggregate disease price measure (i.e., the MCE) and an aggregate service price measure (i.e., the SPI-procedure) that is constructed in a similar manner to official price statistics. We find that the MCE and the SPI-procedure indexes grow at similar rates, indicating that utilization shifts cause no aggregate differences in these indexes. Moreover, the growth in the MCE and SPI-procedure indexes are similar to the corresponding BEA PCE deflator for medical care. This finding indicates that, over this time period, the BEA PCE deflator may provide a reasonable proxy to the cost of treatment. This result is robust to numerous

affect the various components of expenditure growth. For instance, we compare estimates from using the full sample with estimates from a sub-sample of the data that consists only of data contributors that contribute to the data in each year of the sample period. In addition, we compare unweighted estimates to weighted estimates that hold the age, sex, and geographic distribution constant. Although we find that applying alternative weights and samples may have a measurable impact on the components of growth, we generally find that the SPI-procedure index and MCE index tend to grow at about the same rate, indicating no aggregate discrepancy between the two types of measures. We also find that, when weights are applied that make the data representative of U.S. totals, both per capita spending and service price estimates fall close to the corresponding national statistics.

alternative ways of estimating both the MCE and SPI-procedure indexes. Therefore, our finding suggests that shifts in utilization patterns across service categories do not create any large discrepancy between service price growth and cost-of-treatment growth over the 2003 to 2007 time period studied.

The aggregate SPI and MCE indexes appear to grow at about the same rate, but looking at specific disease categories we uncover some important differences, which have implications for health care inflation and productivity. For instance, for cardiology conditions, we find that the MCE grows more slowly than the SPI-procedure. This implies that the SPI-procedure overstates inflation relative to the MCE and understates real output growth by the same amount. We observe the reverse pattern for orthopedic conditions, where we find that the MCE grows more quickly than the SPI-procedure. These findings are likely to lead to speculation about the changes in treatment patterns that may cause these differences. For instance, for cardiology conditions, one may note the wider use of hypertension and high cholesterol drugs that may prevent costly inpatient admissions. For orthopedic conditions, one may think of new technologies, such as the growing trend toward the use of spine surgeries to treat back pain, which some have argued are potentially wasteful and lead to excessive growth in utilization (Dartmouth Atlas (2012)).[28] This real output interpretation may be controversial, since the assumption of fixed quality is likely not to hold for many treatments and in many instances technological improvements are quite visible (e.g., drugs for treating depression, cancer, and cholesterol; cataract treatments; and heart attack treatments). Moreover, health experts will likely have differing views regarding quality changes. For instance, some health experts may argue that the trend toward a greater number of spine surgeries is, indeed, beneficial.[29]

Whatever view one has about the causes of relative trends in disease and service prices, simply examining the disease-price indexes focuses attention on whether the disease expenditures are worth it. Addressing this question is beyond the scope of our study. However, we anticipate that providing these disease price statistics will contribute to this

[28] For example, Kallmes et al. (2009) show no benefit from surgery relative to a control group for the treatment of certain back fractures.

[29] Although they could not show that all spine surgeries were highly cost-effective, Tosteson et al. (2012) show benefits of certain spine surgeries after two years.

line of research by prompting questions in this area.

There are several areas for future research related to disease-price indexes. First, more research is needed to investigate whether the patterns observed in this paper are observed for other markets (e.g., Medicare and Medicaid) and time periods. Second, although we attempt to account for the severity of patient illnesses, future work should check the robustness of these results to alternative severity adjustments. Third, future research is needed to better account for the changing quality of treatments over time.

References

[1] Aizcorbe, Bradley (BLS), Herauf, Kane, Liebman, Pack, Rozental (BLS), (2011), "Alternative Price Indexes for Medical Care: Evidence from the MEPS Survey", BEA Working Paper (WP2011-01).

[2] Aizcorbe, Ana and Nicole Nestoriak, (2011), "Changing Mix of Medical Care Services: Stylized Facts and Implications for Price Indexes", *Journal of Health Economics*, 30(3) pgs 568-574.

[3] Berndt, Ernst, David Cutler, Richard Frank, Zvi Griliches, Joseph Newhouse, and Jack Triplett, (2000), "Medical Care Prices and Output", Handbook of Health Economics, Chapter 3.

[4] Berndt, Ernst, Anupa Birb, Susan H. Buschc, Richard G. Frank, ,and Sharon-Lise T. Normande, (2002), "The Medical Treatment of Depression, 1991–1996: Productive Inefficiency, Expected Outcome Variations, and Price Indexes", *Journal of Health Economics*, 21(3) pages 373-396.

[5] Berndt, Ernst R., Iain M. Cockburn, and Zvi Griliches, (1996), "Pharmaceutical Innovations and Market Dynamics: Tracking Effects on Price Indexes for Antidepressant Drugs", Brookings Papers on Economic Activity: Microeconomics 2 (1996): 133-88.

[6] Bundorf, Kate, Anne Royalty, and Laurence Baker, (2009), "Health Care Cost Growth Among the Privately Insured", *Health Affairs*, September/October.

[7] Chernew, Michael, Joseph Newhouse, (2012), "Health Care Spending Growth", *Handbook of Health Economics*, Chapter 1, pgs 1-43.

[8] Committee on National Statistics, (2011), *Accounting for Health and Health Care: Approaches to Measuring the Sources and Costs of Their Improvement*, The National Academies Press, Washington, D.C.

[9] Cutler, David M., Mark McClellan, Joseph P. Newhouse, and Dahlia Remler, (1998), "Are Medical Prices Declining? Evidence from Heart Attack Treatments", *Quarterly Journal of Economics*, 113 pgs 991-1024.

[10] Dartmouth Atlas, (2012), "A Report by the Dartmouth Atlas of Health Care", CMS-FDA Collaborative.

[11] Duggan, Mark, (2005), "Do New Prescription Drugs Pay for Themselves? The Case of Second-Generation Anti-Psychotics", *Journal of Health Economics*, 24, pgs 1-31.

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

[13] Dunn, Abe, Eli Liebman, Sarah Pack, and Adam Shapiro, (2012a), "Medical Care Price Indexes for Patients with Employer-Provided Insurance: Nationally-Representative Estimates from MarketScan Data", *Health Services Research*, Forthcoming.

[14] Dunn, Abe, Eli Liebman, and Adam Shapiro, (2012b), "Decomposing Medical-Care Expenditure Growth", *Working Paper.*

[15] Dunn, Abe, Eli Liebman, and Adam Shapiro, (2012c), "Developing a Framework for Decomposing Medical-Care Expenditure Growth: Exploring Issues of Representativeness", Chapter 17 in Dale Jorgenson, Steven Landefeld, and Paul Schreyer, eds., *Measuring Economic Sustainability and Progress*, NBER Book Series Studies in Income and Wealth, Forthcoming

[16] Dunn, Abe, Eli Liebman, Adam Shapiro, and Lindsey Rittmueller, (2012d), "Defining Disease Episodes and the Effects on the Components of Expenditure Growth", Bureau of Economic Analysis, *Working Paper.*

[17] Dunn, Abe, Adam Shapiro, and Eli Liebman, (2013), "Geographic Variation in Commercial Medical-Care Expenditures: A Framework for Decomposing Price and Utilization", *Journal of Health Economics, Forthcoming.*

[18] Frank, R.G., T.G. McGuire, and S.-L. Normand, (2006), "Cost-Offsets of New Medications for Treatment of Schizophrenia." National Bureau of Economic Research, Cambridge, MA.

[19] Griliches, Zvi and Iain Cockburn, (1994), "Generics and New Goods in Pharmaceutical Price Indexes", *American Economic Review*, 84(5) pgs 1213-1232.

[20] Kallmes, David et al., (2009), "A Randomized Trial of Vertebroplasty for Osteoporotic Spinal Fractures", *New England Journal of Medicine*, 361(6) pgs 569-579.

[21] Lucarelli, Claudio and Sean Nicholson, (2009), "Quality-Adjusted Price Index for Colorectal Cancer Drugs", NBER Working Paper No. 15174

[22] Matthew D. Shapiro and David W. Wilcox, (1996), "Mismeasurement in the Consumer Price Index: An Evaluation," in Ben S. Bernanke and Julio J. Rotemberg, Editors, NBER Macroeconomics Annual 1996, Volume 11 MIT Press

[23] Tosteson, Anna, J.D. Lurie, T.D. Tosteson, J.S. Skinner, H. Herkowitz, T. Albert, S.D. Boden, K. Bridwell, M. Longley, G.B. Andersson, E.A. Blood, M.R. Grove, J.N. Weinstein(2008), "Surgical Treatment of Spinal Stenosis with and without Degenerative Spondylolisthesis: Cost-Effectiveness after 2 Years", *Annals of Internal Medicine*, 149 pgs 845-853.

6 Appendix

6.1 Decomposition of MCE by Service Category

Table A1.1 MCE by Service Category

	Inpat. Hospital	Outpat. Hospital	Office	Other	Brand Drug	Generic Drug
Orthopedics & rheumatology	1.189	1.109	1.183	1.335	0.816	1.603
Cardiology	0.967	1.083	1.158	1.275	0.990	1.461
Gastroenterology	1.044	1.091	1.241	1.448	0.860	1.720
Gynecology	0.987	1.308	1.282	1.242	0.994	1.377
Endocrinology	0.753	1.148	1.081	1.264	1.086	1.679

Table A1.2. SPI by Service Category

	Inpat. Hospital	Outpat. Hospital	Office	Other	Brand Drug	Generic Drug
Orthopedics & rheumatology	1.325	1.117	1.040	1.161	1.232	0.815
Cardiology	1.199	1.109	1.042	1.295	1.232	0.857
Gastroenterology	1.186	1.136	1.148	1.112	1.198	0.799
Gynecology	1.191	1.229	1.139	1.211	1.228	0.815
Endocrinology	1.124	1.173	1.071	1.155	1.247	1.004

Table A1.3. SUI by Service Category

	Inpat. Hospital	Outpat. Hospital	Office	Other	Brand Drug	Generic Drug
Orthopedics & rheumatology	0.906	0.993	1.139	1.152	0.673	1.977
Cardiology	0.805	0.974	1.111	0.990	0.803	1.701
Gastroenterology	0.881	0.957	1.080	1.316	0.718	2.117
Gynecology	0.830	1.060	1.128	1.024	0.807	1.650
Endocrinology	0.666	0.979	1.008	1.091	0.865	1.750

6.2 Alternative Robustness Checks

In the main text, we apply regional weights and the ETG grouper with severity adjustments. In companion pieces to this paper, we explain in greater depth the effect of applying alternative grouper methodologies and population weights (see Dunn et al. (2012d) and Dunn et al. (2012c)). Although a more complete discussion of both weights and groupers is relegated to other papers, here we briefly present some key robustness results. Three of these estimates are reported in Table A2. The estimates in the first row are the same estimate as Table 2, but we do not apply ETG's severity adjustment. The results are nearly identical, but the MCE and SUI-procedure increase slightly. The estimates in the second row are the same as in Table 2, but we apply weights at the county level. The county weights are discussed in greater detail in Dunn et al. (2012c), but they essentially hold demographics constant and also hold constant each county's contribution to the national estimate to be proportional to each county's population. (Note that only those counties with at least 2,000 enrollees in each year are kept for this analysis). The county weights are applied to control for fluctuations in the geography of the sample within a region. The results remain very similar to those reported in the paper, despite the unique weighting. Finally, using the same county weighting strategy, we apply the MEG grouper with severity adjustment. Again the MCE and SPI-procedure have a similar growth pattern. However, note that when we apply the MEG grouper we see that the difference between the SPI-encounter and SPI-procedure is diminished.

Table A2. Components of Episode Expenditure Growth 2003-07 - Alternative Methodologies

		Procedure-Based		Encounter-Based	
	MCE	SPI-proc.	SUI-proc.	SPI-enco.	SUI-enco.
ETG - Not Severity Adj., Regional Weights	1.146	1.155	1.004	1.256	0.928
ETG - Severity Adj., County Weights	1.120	1.132	1.003	1.233	0.926
MEG - Severity Adj., County Weights	1.146	1.139	1.009	1.187	0.976