Medical Care Price Indexes for Patients with Employer-Provided Insurance: Nationally Representative Estimates from MarketScan Data

Abe Dunn, Eli Liebman, Sarah Pack, and Adam Hale Shapiro

Objective. Commonly observed shifts in the utilization of medical care services to treat diseases may pose problems for official price indexes at the Bureau of Labor Statistics (BLS) that do not account for service shifts. We examine how these shifts may lead to different price estimates than those observed in official price statistics at the BLS.

Data Sources. We use a convenience sample of enrollees with employer-provided insurance from the MarketScan database for the years 2003 to 2007. Population weights that consider the age, sex, and geographic distribution of enrollees are assigned to construct representative estimates.

Study Design. We compare two types of price indexes: (1) a Service Price Index (SPI) that is similar to the BLS index, which holds services fixed and measures the prices of the underlying treatments; (2) a Medical Care Expenditure Index (MCE) that measures the cost of treating diseases and allows for utilization shifts.

Principal Findings. Over the entire period of study the CAGR of the SPI grows 0.7 percentage points faster than the preferred MCE index.

Conclusions. Our findings suggest that the health component of inflation may be overstated by 0.7 percentage points per year, and real GDP growth may be understated by a similar amount. However, more work may be necessary to precisely replicate the indexes of the BLS to obtain a more accurate measure of these price differences.

INTRODUCTION

The rise in health care spending has been at the forefront of recent economic policy debate. Health care's share of gross domestic product (GDP) has risen from 5.2 percent in 1960 to 16.2 percent in 2008 (Hartman et al. 2010). However, measuring real health care output is a challenging task in this industry, as

prices of the underlying procedures—the unit of analysis currently used in official statistics by the Bureau of Labor Statistics (BLS)—may not be the correct unit of analysis. A more appropriate price used to capture real output would focus on what the patient ultimately cares about, the total expenditures necessary to treat a disease (Berndt et al. 2001). This more appropriately defined price is a combination of procedural prices as well as the utilization of procedures within an episode of care. Indeed, a recent study by Aizcorbe and Nestoriak (2011) shows that not accounting for utilization shifts may result in a health care price measure that overstates price growth, and hence, understates real GDP growth. In this study, we aim to supplement this line of work by providing another measure of health care prices which takes the utilization shifts into account.

Health care service utilization has undergone many important shifts, having potentially important implications for the price of medical care. If providers are substituting away from high-priced procedures (or services) toward lower priced services, the expenditure of treating a specific episode of illness could very well decline, even if all of the prices of the individual services rise. These shifts in medical care are not reflected in official price indexes reported by the BLS, which only capture the provider prices for a fixed basket of services, not allowing for shifts in treatment.¹ Expenditure shifts have been documented and quantified for specific diseases, such as cataracts, heart attacks, and depression.² These shifts have also been shown when looking across a broad set of diseases as in the work by Aizcorbe and Nestoriak (2011), henceforth AN, and Aizcorbe et al. (2010). These studies generally compare two types of price indexes: (1) a Service Price Index, or SPI, that is similar to the BLS index because it holds services fixed (both the type of care and amount of care) and measures the prices of the underlying treatments³; (2) a Medical Care Expenditure Index or MCE index, that measures the cost of treating diseases, which is distinct from the SPI because it allows for shifts across treatment categories. These studies find that the MCE index that accounts for these shifts grows at an annual rate that is slower than that of the SPI. Using the more correctly defined MCE price measure, rather than the SPI index, these studies find that current inflation is significantly overstated and that the growth in real GDP may be understated by a similar amount.

Address correspondence to Abe Dunn, Bureau of Economic Analysis, 1441 L Street NW, Washington, DC 20230; e-mail: abe.dunn@bea.gov. Eli Liebman, B.S., and Sarah Pack, M.S., are also with the Bureau of Economic Analysis, Washington, DC. Adam Hale Shapiro, Ph.D., is with the Federal Reserve Bank of San Francisco, San Francisco, CA.

This study contributes to the literature by comparing the medical care expenditure and service price indexes to quantify the presence of shifts in the utilization of services. A comparison of the SPI, which approximates the price measure currently used to calculate real GDP, and the MCE index, which is preferred by most health economists, allows us to assess the contributions of service shifts to overall expenditure growth. To do this, we quantify the potential importance of service shifts for price indexes across a broad range of conditions by calculating and comparing expenditure growth using these two types of indexes. The methodology for measuring these shifts is similar to AN; however, we use an alternative data source from Thomson Reuters that is retrospective claims data for a sample of commercially insured patients; the AN study uses commercially insured data from Pharmetrics. Documenting these shifts using alternative data sources is important for understanding the potential bias in real GDP for several reasons. First, the underlying data in both studies are based on convenience samples, but because MarketScan data are based on a different sample of enrollees (including data from both large employers and insurers), it offers an important check to determine whether the patterns discovered in AN can be applied to the general U.S. population. Second, because neither dataset is nationally representative, we apply weights based on region, age, and gender to make the analysis representative of the commercially insured population and examine how these weights affect our results. Third, while AN study the period between 2003 and 2005, we examine patterns in spending over the extended period from 2003 to 2007.⁴

We find that over the 2003 : 1–2005 : 1 period, the MCE index grew at a compound annual rate (CAGR) of about 3.6 percent per year, which is 0.9 percentage points slower than the SPI. A literal interpretation of this result is that the expenditure growth in our sample would have been even higher in the absence of service shifts. This pattern extends over the period 2003–2007 and we find that applying the nationally representative weights has a minimal effect on our results. Over the entire period of study from 2003 : 1 to 2007 : 4 the difference in CAGR is 0.7 percentage points. Although the qualitative finding is similar to that of AN, we find a smaller difference between the MCE and SPI indexes, implying a smaller inflation bias. Related work (Aizcorbe et al. 2010) also conducts a similar study using the nationally representative Medical Expenditure Panel Survey (MEPS) data and finds results both qualitatively and quantitatively similar to those found using the MarketScan data. That is, it finds an inflation bias that is smaller than that presented by AN and more aligned with the estimates presented in this study.

DATA SOURCE

Because the MarketScan data are a convenience sample, it is not representative of the commercially insured or national populations. To account for the lack of representativeness, in this study we apply weights based on region, age, and gender to make the analysis more representative of the commercially insured population and examine how these weights affect our results.⁵

We use a sample of enrollees that are not in capitated plans from the MarketScan database for the years 2003–2007.⁶ MarketScan data draw disproportionately from the southern region of the United States. This difference from the true nationwide distribution of the population implies that it may be important to apply weights that control for the enrollment distribution across geographic areas. We limit our sample to enrollees with drug benefits because drug purchases will not be observed for individuals without drug coverage. The MarketScan database tracks claims from all providers using a nationwide convenience sample of patients and is therefore not representative of all commercially insured patients. MarketScan collects data from employers, health plans, and state-level Medicaid agencies; all claims have been paid and adjudicated. Each enrollee has a unique identifier and can be identified at the three-digit zip code level (Adamson, Chang, and Hansen 2008). This study uses the Commercial Claims and Encounters Database portion of the MarketScan Databases, which includes health care utilization and cost records at the encounter level. This portion of the database provides patient identifiers that may be used to sum expenditures to the patient level.

The Commercial Claims and Encounters Database contains data from employer and health plan sources concerning medical and drug data for several million employer-sponsored insurance (ESI)-covered individuals, including employees, their spouses, and dependents. Each observation in the data corresponds to a line item in an “explanation of benefits” form; therefore, each claim can consist of many records and each encounter can consist of many claims. To obtain the numbers of services and treated patients (“x” and “N”, respectively, throughout this study), we aggregate the individual records assigned to a specific condition. For services that last only 1 day (e.g., pharmacy visit, office visit), we define the service as a day of care from the provider for a particular disease. For inpatient stays, we define the service as the medical confinement. We group providers by industry (using an identifier for the place of service), by the medical disease (using the ETG code assigned by the Symmetry grouper), and by day (using the date). We measure the number of

patients treated as the number of people who received treatment for a disease, d , in a given period. Expenditures are measured as the amount received by all providers of the services (including both out-of-pocket payments and amounts paid by insurance firms).

Claims datasets such as MarketScan are used in some other studies that explored problems in medical care price indexes (Berndt et al. 2001; Song et al. 2009) and in other studies that document shifts in utilization (Bundorf, Royalty, and Baker 2009; Chernew and Fendrick 2009). Although these indexes are not nationally representative, their advantage is that the large number of observations provides a better representation of spending at the high end of the distribution and the use of administrative records avoids undercount issues typical of household expenditure surveys (Aizcorbe et al. 2010).

METHODOLOGY

In this study, we construct both MCE and SPI indexes and then apply the decomposition developed by AN, both of which require measures of spending by disease. In addition, one must link services to the diseases or conditions that are being treated and then choose fixed periods of time over which to identify industry shifts.

MCE AND SPI INDEXES

The MCE index is the average cost of treating a specific disease in the base period for a given patient, compared with the cost of treating a disease for a given patient in the current period; existing literature uses the concept of an episode of care as the fundamental building block for MCE price indexes. The MCE index is the ratio of period 1 prices to treat the disease, divided by period 2 prices:

$$MCE_d = \frac{C_d^2}{C_d^1} = \frac{\sum_s (c_{d,s}^2 x_{d,s}^2) / N_d^2}{\sum_s (c_{d,s}^1 x_{d,s}^1) / N_d^1} \quad (1)$$

where c_d denotes the expenditures for services used to treat disease d and N_d is the number of people with disease d .⁷ $x_{d,s}$ represents the number of encounters for service type s performed for disease d .

The MCE measure allows for shifts in both the type and number of services performed for a given patient. In contrast, the SPI represents an index

similar to those used traditionally that holds fixed the basket of services performed across time periods. Specifically, the price change from period 1 to period 2 is as follows:

$$SPI_d = \frac{\sum_s (c_{d,s}^2 x_{d,s}^1) / N_d}{\sum_s (c_{d,s}^1 x_{d,s}^1) / N_d^1} \quad (2)$$

That is, for the SPI index, the x s and the N s are held fixed from period 1 to period 2. The idea here is to track the amount of dollars needed to compensate a patient for any change in the price of a service given that she consumes the same service type mix between periods.

MEASURING SPENDING BY DISEASE

To construct the above price indexes, it is necessary to classify spending by disease. There are a number of methods that may be used to classify spending, which are discussed more extensively in AN. Here, we classify disease spending using a commercial algorithm called a grouper; specifically, we use the ETG grouper from Symmetry. A grouper applies an algorithm to the data to assign each record to an 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 entirely rely on the expertise of those that developed the grouper software.

The ETG grouper allocates each record into one of over 500 disease groups called "episode treatment groups" (ETGs). Although the price indices are constructed at the ETG level, the results will be presented at a higher level of aggregation that Symmetry calls major practice categories or MPCs.

DESCRIPTION OF THE MARKETSCAN POPULATION AND COMPARISON WITH PHARMETRICS DATA

Some descriptive statistics from the MarketScan data are reported in Table 1, along with a comparison to the Pharmetrics data. The population in the

Table 1: Descriptive Statistics for MarketScan and Pharmetrics Samples

	<i>MarketScan Unweighted</i>			<i>Pharmetrics</i>	
	<i>2003</i>	<i>2005</i>	<i>2007</i>	<i>2003</i>	<i>2005</i>
Number of plans	—	—	—	21	21
Number of enrollees (mil)	9.9	15.2	19.1	10.9	11.3
Gender (%)					
Female	51.7	51.6	51.2	54.5	53.9
Male	48.3	48.4	48.8	45.5	46.1
Age (%)					
0–18	26.6	27.3	27.6	29.7	28.0
19–24	8.5	8.2	8.5	6.6	6.7
25–34	14.0	14.0	14.2	13.9	13.7
35–54	35.9	35.4	35.1	37.1	37.1
55 and over	14.9	15.2	14.7	12.7	14.6
Region (%)					
NE	11.3	11.2	11.6	22.5	24.3
MW	28.0	26.8	23.0	26.1	27.8
S	46.2	45.8	49.6	31.5	30.4
W	14.5	16.2	15.7	19.8	17.4

MarketScan data grows substantially over the study period, starting with a population of 9.9 million (10 percent smaller than the Pharmetrics data), and nearly doubling in size to 19.1 million. This growth is from two sources: (1) the incorporation of insurance claims data starting in 2003 and expanding over the entire sample period and (2) an expansion in the number of enrollees from employers. This difference from the true nationwide distribution of the population implies that it may be important to apply weights that control for the enrollment distribution across geographic areas.

Table 2 reports health care expenditures by disease category and a comparison with the Pharmetrics data. Of the \$6.7 billion of total spending reported in the MarketScan data for 2003 : 1, about 97 percent is allocated to disease classes (Major Practice Categories [MPC] 1–19), and just over 1 percent is allocated to nondisease MPCs (preventative and administrative care). Similar to AN’s findings using Pharmetrics data, we find that most of these allocations (96 percent) are done using some combination of diagnosis (ICD-9) and procedure codes (CPT-4) on the claim record or in the patients’ histories (column 1). The remainder of the expenditures are allocated using NDC drug codes, which (column 2) represent spending by patients with ongoing prescriptions and not other medical encounters (hence, no diagnosis or procedure codes). Overall, the share of spending by disease category is quite similar to that observed in the Pharmetrics data (comparing the last two columns).

Table 2: Allocation of Spending by Disease, 2003 : 1 (Million Dollars)

Major Practice Category	MarketScan				Pharmetrics	
	Spending Allocated Using		Total Spending		Total Spending	
	Diagnoses and Procedure Codes	Drug Codes (NDC)	Millions	Percent	Millions	Percent
1 Infectious diseases	\$72.4	\$8.1	\$80.5	1.2	\$80.5	1.0
2 Endocrinology	\$414.5	\$36.7	\$451.2	6.7	\$451.2	6.0
3 Hematology	\$139.9	\$2.0	\$141.9	2.1	\$141.9	2.3
4 Psychiatry	\$254.2	\$36.8	\$291.1	4.4	\$291.1	5.3
5 Chemical dependency	\$27.9		\$27.9	0.4	\$27.9	0.6
6 Neurology	\$353.0	\$20.8	\$373.8	5.6	\$373.8	5.5
7 Ophthalmology	\$91.4	\$0.3	\$91.7	1.4	\$91.7	1.7
8 Cardiology	\$847.6	\$11.3	\$858.9	12.8	\$858.9	9.3
9 Otolaryngology	\$404.0	\$24.4	\$428.4	6.4	\$428.4	7.6
10 Pulmonology	\$365.9	\$13.5	\$379.4	5.7	\$379.4	4.8
11 Gastroenterology	\$559.6		\$592.3	8.9	\$592.3	8.5
12 Hepatology	\$158.7		\$158.7	2.4	\$158.7	2.3
13 Nephrology	\$99.1		\$99.1	1.5	\$99.1	1.0
14 Urology	\$215.1	\$3.0	\$218.1	3.3	\$218.1	2.9
15 Obstetrics	\$242.0		\$242.0	3.6	\$242.0	4.4
16 Gynecology	\$442.0	\$0.4	\$442.4	6.6	\$442.4	6.5
17 Dermatology	\$225.7	\$6.7	\$232.4	3.5	\$232.4	3.9
18 Orthopedics and rheumatology	\$980.6	\$8.9	\$989.5	14.8	\$989.5	14.7
19 Neonatology	\$103.9		\$103.9	1.6	\$103.9	1.9
20 Preventive and administrative	\$126.5	\$45.3	\$171.8	2.6	\$171.8	3.7
21 Late effects, environmental trauma, and poisonings	\$32.8		\$32.8	0.5	\$32.8	0.5
22 Isolated signs and symptoms	\$73.7	\$8.3	\$82.1	1.2	\$82.1	1.5
Other	\$195.4		\$195.4	2.9	\$195.4	4.1
Total	\$6,426.0	\$259.3	\$6,685.3	100.0	\$6,685.3	100.0

Perhaps, the most notable difference in the two datasets is the larger share of expenditures allocated to cardiology spending in the MarketScan data. This is at least partly due to the higher share of individual above 55 in the MarketScan sample.⁸

Before proceeding to the results, it is worth noting that one known difference between the MarketScan and Pharmetrics samples is that the MarketScan draws information from both large employers (70 percent of the data and 77 contributing employers) and health insurers (30 percent of the data and 12 contributing health plans),⁹ whereas the Pharmetrics data used in AN are from 21 commercial health plans of various types (PPO, POS, and HMO). However, additional details regarding the underlying samples in both the MarketScan and Pharmetrics datasets are unknown, as the identity of the data contributors is confidential in both datasets.¹⁰

STUDY FINDINGS

Chart 1 illustrates a comparison of the SPI, which measures the prices of the underlying treatments, and the MCE, which measures the cost of treating diseases. Using the MarketScan data, the quarterly indexes are constructed for 2003 : 1 through 2007 : 4, with 2003 : 1 being the baseline of 1.0. Based on the MarketScan results for the period 2003 : 1–2007 : 4, the MCE exhibits slower growth than the SPI: the difference between the two indexes is notable, with the SPI growing 18.9 percent over the 3-year period and the MCE growing at a markedly slower 15.3 percent. There is a difference of about 0.7 percentage points between the compound annual growth rates, with the MCE CAGR being 3.0 percentage points and the SPI CAGR being 3.7 percentage points.

The results presented in Chart 1 are for an unweighted population. Some additional results are reported in Table 3, including estimates for a weighted population and standard errors for the MarketScan indexes. Table 3 reports estimates that include population weights, so that the MarketScan population is nationally representative of the commercially insured population.¹¹ We find that the weighted results are nearly identical to the unweighted results. Although both the SPI and MCE appear to grow at slightly faster rates when using population weights, the difference in levels between the SPI and MCE indexes is quite similar. Therefore, even when population weights are applied, the implied bias of using an MCE index relative to an SPI index remains unchanged. More precisely, over the entire period and focusing on the

Table 3: Comparison of Weighted and Unweighted Price Indices with AN Unweighted Results

Year/ Quarter	AN Results		MarketScan							
			Unweighted				Weighted			
	SPI	MCE	SPI	SE	MCE	SE	SPI	SE	MCE	SE
2003-1	1.00	1.00	1.00	(0.006)	1.00	(0.007)	1.00	(0.005)	1.00	(0.006)
2003-2	1.02	1.00	1.02	(0.003)	1.02	(0.003)	1.02	(0.003)	1.02	(0.003)
2003-3	1.03	0.99	1.03	(0.005)	1.02	(0.005)	1.03	(0.005)	1.02	(0.005)
2003-4	1.07	1.03	1.05	(0.003)	1.04	(0.003)	1.05	(0.003)	1.04	(0.003)
2004-1	1.08	1.04	1.05	(0.003)	1.05	(0.003)	1.05	(0.002)	1.05	(0.003)
2004-2	1.10	1.06	1.06	(0.003)	1.05	(0.003)	1.06	(0.003)	1.05	(0.003)
2004-3	1.11	1.06	1.06	(0.003)	1.05	(0.003)	1.06	(0.003)	1.04	(0.003)
2004-4	1.10	1.05	1.08	(0.003)	1.06	(0.003)	1.09	(0.003)	1.06	(0.004)
2005-1	1.11	1.06	1.08	(0.003)	1.06	(0.003)	1.09	(0.003)	1.06	(0.003)
2005-2	1.14	1.09	1.09	(0.003)	1.08	(0.003)	1.10	(0.003)	1.08	(0.003)
2005-3	1.16	1.10	1.11	(0.003)	1.08	(0.003)	1.11	(0.003)	1.09	(0.003)
2005-4	1.18	1.11	1.13	(0.003)	1.10	(0.003)	1.14	(0.004)	1.11	(0.004)
2006-1			1.13	(0.003)	1.11	(0.004)	1.14	(0.004)	1.12	(0.004)
2006-2			1.14	(0.003)	1.11	(0.003)	1.15	(0.003)	1.13	(0.004)
2006-3			1.14	(0.003)	1.11	(0.004)	1.16	(0.004)	1.12	(0.004)
2006-4			1.16	(0.003)	1.12	(0.003)	1.18	(0.003)	1.14	(0.004)
2007-1			1.16	(0.003)	1.12	(0.003)	1.17	(0.003)	1.14	(0.003)
2007-2			1.17	(0.003)	1.13	(0.003)	1.19	(0.003)	1.16	(0.003)
2007-3			1.17	(0.003)	1.12	(0.003)	1.19	(0.003)	1.15	(0.004)
2007-4			1.19	(0.003)	1.15	(0.004)	1.21	(0.004)	1.18	(0.004)
CAGR										
2003-1 to 2005-4	1.062	1.040	1.045		1.036		1.048		1.038	
2006-1 to 2007-4			1.031		1.024		1.035		1.027	
2003-1 to 2007-4			1.037		1.030		1.042		1.035	

Note. Standard errors are computed by applying the delta method.

weighted results, we find the CAGR for the MCE is 3.5 percent and the CAGR for the SPI is 4.2 percent, so the overall implied bias is 0.07 percent. It is also worth highlighting that for each of the indexes reported in Table 3 the standard errors are relatively small, indicating that each of these indexes is precisely estimated.¹² We do not have standard error estimates for the AN

analysis, although given the size of their sample, it is likely that their standard errors are also quite small.¹³

Overall, our findings are similar to those (Aizcorbe and Nestoriak 2010; Aizcorbe et al. 2010) who also found that the MCE grows slower than the SPI. However, there are two key differences. As indicated in Table 3, the overall growth rate using the Pharmetrics sample is faster for both types of indexes, with compound annual growth rates of 6.2 percentage points for the SPI (compared with 4.5 percent in the MarketScan data) and 4.0 percent for the MCE (compared with 3.6 percent in the MarketScan data). Second, the difference between the MCE and SPI is notably smaller in the MarketScan data even when focusing on the same 2003 : 1–2005 : 4 period. Using the weighted or unweighted analysis, the difference in the CAGR between in MCE and SPI indexes is 0.9 percentage points over this 3-year period, compared with a difference of 2.2 percentage points found using the Pharmetrics data. Our estimate of yearly inflation bias over the 2003 : 1–2005 : 4 period of 0.9 percent is closer to the estimates from Aizcorbe et al. (2010) that measure the difference in the SPI and MCE for the 2001–2005 time period using the MEPS data. They find a difference in the MCE and SPI growth rates of 1 percent per year. Although we obtain a bias of 0.9 percent for this time period, recall that the bias for the entire sample period from 2003 : 1 to 2007 : 4 is a bit lower, around 0.7 percent per year.

DECOMPOSITION: SHIFTS IN UTILIZATION

Next, we turn to the decomposition to examine what factors are causing the differences between the MCE and SPI indexes. To keep the analysis comparable to that of AN, who perform unweighted analysis, Table 4 reports the unweighted decomposition of the sample for the period 2003 : 1–2005 : 4. The growth rates for the MCE and SPI indexes for each of the MPC disease categories are reported. These growth rates indicate that expenditures per patient, as well as the average price of services, increased for many of the major categories. The third column shows the difference between the two indexes. As indicated by the negative numbers, for most conditions expenditures per patient did not grow as fast as they would have if patients had received the same bundle of services in 2005 that patients received in 2003. There are three exceptions in the MarketScan data: (1) ophthalmology, (2) obstetrics, and (3) orthopedics and rheumatology. Combined, the three exceptions we found in the MarketScan data comprise

Table 4: MarketScan Decomposition Results Unweighted, 2003: 1–2005: 4

<i>Major Practice Category</i>		<i>MCE</i>	<i>SPI</i>	<i>MCE-SPI</i>	<i>Hospital Inpatient Outpatient</i>		<i>Office Visits</i>	<i>Pharmacy</i>	<i>Other</i>
1	Infectious diseases	1.15	1.17	−0.02	0.01	0.00	−0.01	−0.02	0.00
2	Endocrinology	1.10	1.16	−0.07	−0.04	0.00	−0.01	−0.02	−0.01
3	Hematology	1.16	1.19	−0.04	−0.03	0.00	0.00	0.00	0.00
4	Psychiatry	1.04	1.11	−0.08	0.00	0.00	−0.03	−0.02	−0.01
5	Chemical dependency	1.03	1.06	−0.03	−0.05	0.06	−0.02	0.01	−0.03
6	Neurology	1.10	1.17	−0.06	−0.03	−0.01	0.00	−0.01	−0.01
7	Ophthalmology	1.09	1.08	0.01	0.00	0.00	0.01	−0.01	0.02
8	Cardiology	1.01	1.12	−0.11	−0.08	−0.01	0.00	−0.02	−0.01
9	Otolaryngology	1.12	1.13	−0.01	0.00	0.00	0.00	−0.02	0.02
10	Pulmonology	1.06	1.09	−0.03	−0.01	0.00	0.00	−0.01	0.00
11	Gastroenterology	1.15	1.14	0.00	−0.01	0.00	0.00	−0.01	0.02
12	Hepatology	1.04	1.07	−0.04	−0.02	−0.01	0.00	−0.01	0.00
13	Nephrology	1.06	1.14	−0.08	−0.02	−0.05	0.00	−0.01	0.00
14	Urology	1.08	1.10	−0.02	−0.02	0.00	0.00	0.00	0.00
15	Obstetrics	1.16	1.09	0.07	0.06	0.01	0.00	0.00	0.00
16	Gynecology	1.13	1.13	0.00	0.00	0.01	0.00	−0.01	0.00
17	Dermatology	1.15	1.17	−0.02	−0.01	0.00	−0.01	−0.01	0.00
18	Orthopedics and rheumatology	1.16	1.14	0.02	0.01	−0.01	0.02	−0.01	0.01
19	Neonatology	1.16	1.19	−0.03	−0.03	0.00	0.00	0.00	−0.01

Note. Standard errors are reported in Table A4 of the appendix.

about 19 percent of total spending. Although the exceptions represent a relatively small share of spending, they are still examples where the cost of treating entire diseases rose faster than the cost of the individual treatments, owing to increases in utilization.

The evidence suggests that shifts in the bundle of services are pervasive. The last five columns of Table 4 indicate where the shifts in treatment are occurring, where a negative number indicates a shift away from a particular service category. In line with AN's findings, our results suggest that many of the shifts in services seem to be related to declining utilization at inpatient hospitals, possibly due to surgeries being performed elsewhere (e.g., outpatient hospital, office visits, or ambulatory surgical centers). Alternatively, the decline in inpatient hospital services may also represent a drop in the total amount of services (i.e., number of encounters) used to treat a particular condition, with the share of expenditures going to many of the other services

Table 5: Comparison of Price Indices and Sources of Differences

Major Practice Category	Estimate 1			Estimate 2			Estimate 3			Estimate 4		
	Unweighted 2003q1–2005q4			Weighted 2003q1–2005q4			Weighted 2003q1–2007q4			Weighted 2003q1–2007q1		
	MCE	SPI	MCE-SPI	MCE	SPI	MCE-SPI	MCE	SPI	MCE-SPI	MCE	SPI	MCE-SPI
1 Infectious diseases	1.15	1.17	–0.02	1.19	1.21	–0.02	1.34	1.39	–0.05	1.25	1.33	–0.07
2 Endocrinology	1.10	1.16	–0.07	1.09	1.16	–0.07	1.09	1.18	–0.09	1.06	1.15	–0.09
3 Hematology	1.16	1.19	–0.04	1.14	1.19	–0.05	1.24	1.28	–0.04	1.24	1.27	–0.03
4 Psychiatry	1.04	1.11	–0.08	1.04	1.11	–0.08	1.05	1.15	–0.11	1.07	1.14	–0.07
5 Chemical dependency	1.03	1.06	–0.03	1.03	1.07	–0.03	1.05	1.17	–0.12	1.04	1.08	–0.04
6 Neurology	1.10	1.17	–0.06	1.11	1.18	–0.07	1.22	1.31	–0.08	1.13	1.20	–0.07
7 Ophthalmology	1.09	1.08	0.01	1.08	1.07	0.01	1.13	1.18	–0.05	1.07	1.13	–0.06
8 Cardiology	1.01	1.12	–0.11	1.01	1.12	–0.11	1.03	1.17	–0.14	1.02	1.12	–0.10
9 Otolaryngology	1.12	1.13	–0.01	1.12	1.13	–0.01	1.18	1.19	–0.01	1.11	1.13	–0.02
10 Pulmonology	1.06	1.09	–0.03	1.08	1.10	–0.02	1.15	1.22	–0.06	1.10	1.17	–0.07
11 Gastroenterology	1.15	1.14	0.00	1.15	1.16	0.00	1.25	1.25	0.01	1.17	1.16	0.01
12 Hepatology	1.04	1.07	–0.04	1.04	1.08	–0.04	1.11	1.16	–0.05	1.03	1.07	–0.04
13 Nephrology	1.06	1.14	–0.08	1.05	1.10	–0.05	0.94	0.99	–0.05	0.97	1.10	–0.13
14 Urology	1.08	1.10	–0.02	1.07	1.10	–0.03	1.17	1.21	–0.04	1.12	1.15	–0.03
15 Obstetrics	1.16	1.09	0.07	1.17	1.10	0.07	1.28	1.20	0.08	1.14	1.13	0.01
16 Gynecology	1.13	1.13	0.00	1.13	1.13	0.00	1.23	1.23	0.00	1.16	1.17	0.00
17 Dermatology	1.15	1.17	–0.02	1.15	1.17	–0.02	1.23	1.27	–0.04	1.21	1.24	–0.03
18 Orthopedics and rheumatology	1.16	1.14	0.02	1.15	1.15	0.01	1.22	1.21	0.01	1.15	1.15	0.00
19 Neonatology	1.16	1.19	–0.03	1.34	1.39	–0.05	1.49	1.50	0.00	1.18	1.20	–0.02

Note. Standard errors are reported in Table A5 of the appendix.

remaining unchanged.¹⁴ In addition to changes in inpatient services, we also observe shifts in other service categories, such as a decline in office visits for the Psychiatry condition category.

Next, Table 5 examines these decompositions using alternative time periods and also applying population weights. Estimate 1 results shown in the first three columns are for the unweighted sample from 2003 : 1 to 2005 : 4. The results for Estimate 2, shown in the next three columns, are for the same time period but have population weights applied to the sample. Looking at these results, we find that they are quite similar across all disease categories, except for neonatology, which reports a price increase that is 15 percent faster relative to the unweighted sample.¹⁵

The third set of results, Estimate 3, shows the decomposition of the weighted results for the full time period, 2003 : 1 through 2007 : 4. The results are quite similar to those for the weighted period, although the difference between the MCE and SPI indexes tends to be slightly more negative in many instances, reflecting that more shifting occurs over a longer period of time. In fact, ophthalmology moves from a price increase due to shifts in utilization to a price decline. The final set of estimates, Estimate 4, removes seasonality effects by reporting estimates on the first quarter of 2007, rather than the fourth quarter, to correspond to the 2003 : 1 base period. Comparing Estimates 3 and 4, it appears that seasonality may be a factor causing positive shifts in price from changes in utilization. After removing seasonality effects, only obstetrics and gastroenterology show slightly positive shifts in price due to service shifts (a 1 percent increase in both disease categories). Here, we see that the shifts that cause the MCE index to be higher than the SPI are less pronounced for both Obstetrics and Orthopedics and Rheumatology. A likely cause of the seasonality effect is that certain services have seasonal variation, such as births, and the high cost of the birth occurs during delivery. For future analysis, this seasonality effect may be accounted for in a number of additional ways, including focusing on annual expenditures or changing the time frame to focus on completed episodes.

Although the MarketScan data are selected from a particular subsegment of the commercially insured population,¹⁶ the analysis of this distinct sample provides insight into the potential importance of shifts in treatment. In 2009, health care spending comprised 16 percent of GDP; therefore, if a difference of a similar magnitude was true across patient types (i.e., including the publicly insured and the uninsured), a transition to the MCE price deflator in BEA's National Accounts would result in a sizeable increase in measured real GDP growth by about 0.15 percentage points per year.¹⁷ While this

indicates a potential large bias, it is less than that found in the AN study that indicates a potential bias as high as 0.25 percent a year. The potential bias is much closer in magnitude to that of Aizcorbe et al. (2010) that find a real GDP bias of around 0.16 percent using MEPS data.

CONCLUSION

This study builds on the work of AN by investigating whether their findings hold more generally. To do this, we apply the methodology of AN using an alternative claims database, MarketScan, which is a distinct dataset that covers a large segment of the commercially insured population, and we use population weights to make the data representative at the national level.

In analyzing differences between MCE and SPI indexes, we find qualitatively similar results to those of AN for the 2003 : 1–2005 : 4 period. Namely, the growth in the MCE index is lower than that of the SPI index. It appears that an important reason for the gap between the MCE and SPI indexes is a shift away from inpatient hospital services. In addition, this study finds a difference in the MCE and SPI index for the 2005–2007 period, which is outside the AN study. Overall, the results in this study support the growing evidence that not accounting for shifts in utilization leads to an overstatement of inflation in health care markets and an understatement of real GDP by a similar amount. Using either our weighted or unweighted analysis and focusing on the full period of study from 2003 : 1 to 2007 : 4, we estimate that the magnitude of the potential inflation bias is 0.7 percentage points per year. One important caveat to the analysis presented here and in previous work is that the SPI studied here is distinct from the SPI reported by the BLS. A key difference is that the SPI reported here is a price per encounter, whereas the BLS price index is a price per service (e.g., price per 15 minute office visit or price per X-ray). More work is necessary to determine whether these potential differences may matter.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Matthew Lebowitz provided excellent research assistance. The authors would also like to thank Thomson Reuters for providing the MarketScan data. 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.

Disclosures: None.

Disclaimers: None.

NOTES

1. See Berndt et al. (2000) and Schultze and Mackie (2002) for a full discussion of the issues.
2. Heart attacks (Cutler et al. 1998), cataracts (Shapiro, Shapiro, and Wilcox 2001), and depression (Berndt, Busch, and Frank 2001).
3. Although the SPI index we study is similar in spirit to the BLS price measure, it should be noted that the methodologies used to construct these indexes are quite distinct.
4. This study also contributes to two goals mentioned by a panel of health experts from the National Academy of Sciences (NAS) in *Accounting for Health and Health Care* (2010). First, it follows the advice of the NAS that has determined that the Bureau of Economic Analysis should “investigate the impact of different expenditure allocation approaches... on price index construction and performance.” Second, we follow the suggestion of the NAS by using a subgroup of the national population as a first step in “demonstrat[ing] that dollars spent in the economy on medical care can be allocated into disease categories in a fashion that yields meaningful information.”
5. The commercially insured are about 61.6 percent of the population (Health, United States, 2009 With Special Feature on Medical Technology 2010).
6. Plans that pay providers based on capitation are not used here because actual payments are not observed.
7. For inpatient stays, we define the service as the medical confinement.
8. Another potential contributing factor is that a greater proportion of individuals are from the South in the MarketScan data, since we observe that this region tends to spend a bit larger fraction of their overall expenditures on cardiology-related conditions.
9. The 70 percent versus 30 percent split was calculated based on patient counts on our selected sample. The number of data contributors of each type was reported in Adamson, Chang, and Hansen (2008).
10. For example, it is unclear whether some of the health insurer contributors are the same across the two datasets.
11. The weights applied are constructed based on the age, gender, and region of the population.
12. Standard errors are calculated using the delta method. The standard errors reported here are relatively large and conservative compared to standard errors based on yearly, rather than quarterly, price indexes. See authors for additional details.

13. One potential concern is that the growing sample size in MarketScan may impact our results. To check this, we have redone the analysis using a fixed panel of families. Specifically, we use only those subscribers and their dependents that appear in the sample for all 5 years of the data. We find qualitatively similar results to those reported in the study. For the unweighted analysis, we find a 21.9 percent growth in the MCE and a 24.7 percent growth in the SPI for the 2003 : 1–2007 : 4 period. We chose not to report this result in Table 3 because of the potential selection issues that arise with using only those families that are in the data for all 5 years (e.g., This sample is aging and for 2003 the sample excludes any individual that may die in the years 2004–2007).
14. It is also possible for both a shift in utilization away from inpatient and a drop in overall service utilization to occur at the same time.
15. It appears likely that high neonatology spending is greater in some areas than in others, so that weighing the estimates properly appears to be important.
16. Our data do not include the publicly insured or the uninsured.
17. This is calculated based on the 2003 : 1–2005 : 4 MCE and SPI price growth differences using the weighted sample. The estimate is a bit lower, 0.12 percent, if one applies the weighted values for the entire sample.

REFERENCES

- Accounting for Health and Health Care. 2010. *Approaches to Measuring the Sources and Costs of their Improvement*. Washington, DC: The National Academies Press.
- Adamson, D. M., S. Chang, and L. G. Hansen. 2008. *Health Research Data for the Real World: The MarketScan Databases*. Thomson Reuters: In White Paper.
- Aizcorbe, A., and N. Nestoriak. 2011. *Changing Mix of Medical Care Services: Stylized Facts and Implications for Price Indexes*. *Journal of Health Economics* 30 (3): 568–74.
- Aizcorbe, A., E. Liebman, S. Pack, D. M. Cutler, M. E. Chernew, and A. B. Rosen. 2010. *Measuring Health Care Costs of Individuals with Employer-Sponsored Health Insurance in the U.S.: A Comparison of Survey and Claims Data*. In BEA Working Papers. Washington, DC: Bureau of Economic Analysis.
- Berndt, E. R., S. H. Busch, and R. G. Frank. 2001. “Treatment Price Indexes for Acute Phase Major Depression.” In *Medical Care Output and Productivity, Studies in Income and Wealth*, edited by D. M. Cutler, and E. R. Berndt, pp. 463–505. Chicago: University of Chicago Press.
- Berndt, E. R., D. M. Cutler, R. G. Frank, Z. Griliches, J. P. Newhouse, and J. E. Triplett. 2000. “Medical Care Prices and Output.” In *Handbook of Health Economics*, edited by J. C. Anthony, and P. N. Joseph, pp. 119–180. Amsterdam: Elsevier.
- . 2001. “Price Indexes for Medical Care Goods and Services: An Overview of Measurement Issues.” In *Medical Care Output and Productivity*, edited by D. M. Cutler and E. R. Berndt, pp. 141–98. Chicago: University of Chicago Press.
- Bundorf, M. K., A. Royalty, and L. C. Baker. 2009. “Health Care Cost Growth among the Privately Insured.” *Health Affairs* 28 (5): 1294–304.

- Chernew, M., and M. Fendrick. 2009. "Clinically Sensitive Cost-Sharing for Prescription Drugs: Thinking beyond the Silos." *Medical Care* 50: 5–507.
- Cutler, D. M., M. McClellan, J. P. Newhouse, and D. Remler. 1998. "Are Medical Prices Declining? Evidence from Heart Attack Treatments." *Quarterly Journal of Economics* 4 (113): 991–1024.
- Hartman, M., A. Martin, O. Nuccio, and A. Catlin. 2010. "Health Spending Growth at a Historic Low in 2008." *Health Affairs* 29 (1): 147.
- Health, United States, 2009 With Special Feature on Medical Technology. 2010. Edited by National Council for Health Statistics. Hyattsville, MD: U.S. Department of Health and Human Services.
- Schultze, C., and C. Mackie, eds. 2002. *At What Price?: Conceptualizing and Measuring Cost-of-Living and Price Indexes*. Washington, DC: National Academies Press.
- Shapiro, I. L., M. Shapiro, and D. Wilcox. 2001. "Measuring the Value of Cataract Surgery." In *Medical Output and Productivity*, edited by D. M. Cutler, and E. R. Berndt, pp. 411–37. Chicago: University of Chicago Press.
- Song, X., W. D. Marder, R. Houchens, J. E. Conklin, and R. Bradley. 2009. "Can a Disease-Based Price Index Improve the Estimation of the Medical Consumer Price Index." In *Price Index Concepts and Measurement*, edited by W. E. Diewert, J. Greenlees, and C. R. Hulten, pp. 329–68. Chicago: University of Chicago Press.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2: MarketScan Decomposition Results Unweighted, 2003: 1–2005: 4.

Appendix SA3: Comparison of Price Indices and Sources of Differences.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.