Geographic Variation in Commercial Medical-Care Expenditures: A Framework for Decomposing Price and Utilization*

Abe Dunn, Adam Hale Shapiro, and Eli Liebman

September 27, 2013

Abstract

This study introduces a new framework for measuring and analyzing medical-care expenditures. The framework focuses on expenditures at the disease level that are decomposed between price and utilization. We find that both price and utilization differences are important contributors to expenditure differences across commercial markets. Further examination shows that for some diseases utilization drives variation while for others price is more important. Finally, when disease-specific measures are aggregated across diseases, much of the important disease-specific variation is masked, leading to much smaller measures of aggregate variation.

*We would like to thank seminar participants at the National Bureau of Economic Research Summer Institute CRIW workshop and participants at the Total Expenditures Meeting held by the National Center for Quality Assurance. 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.

1 Introduction

Relatively little is known about geographic variation in commercial-market expenditures in the United States. Although there has been a considerable amount of research assessing Medicare expenditures, the commercial market (that is, the privately insured medical-care market) accounts for 60 percent more spending.[1] This dearth of research in this area is a large hole in our understanding of the overall health care market. In this study, we provide a new framework for analyzing medical-care expenditure variation applied to the commercial market. As a central part of our framework, we construct a medical-care expenditure index (MCE) designed to track the expenditure of treating an episode of a disease. Focusing on an MCE recognizes that both differences in service prices and the utilization of different services may affect the total cost of treatment across geographic markets.

Constructing disease-based price indexes has been widely advocated by health economists (see Berndt et al. (2001)) and has led to proposals to incorporate MCEs in official statistics (see National Research Council (2010)). In response, there have been several studies examining the cost of treating disease episodes over time (e.g., Aizcorbe and Nestoriak (2011), Aizcorbe et al. (2010), Dunn et al. (2010), Dunn et al. (2012), and Bradley (2013)).[2] Our study differs from these papers along two dimensions: (1) we track geographic variation in the MCE as opposed to time-series variation; and (2) we introduce a methodology for decomposing the MCE between its two key components, a service price index (SPI) and a service utilization index (SUI). The SPI isolates the variation in underlying service prices (for example, the price of a 15-minute office visit to a doctor to manage a pregnancy), but holding service utilization constant (for example, fixing the number of 15-minute visits to the doctor across markets for each pregnancy). By contrast, the SUI isolates the variation in medical-care expenditures attributable to the quantity of services provided per episode of care. Specifically, the SUI holds the prices of the underlying services constant but allows the number of services to vary.

Using our medical-care expenditure decomposition, we present a descriptive analysis of how spending for a specific disease and its components, service price and service utilization,

[1] Private-insurer and Medicare markets are fundamentally different. Unlike the Medicare markets where payments to providers are fixed, prices in the private sector are set through negotiations between insurers and providers. Moreover, empricial evidence suggests that commercial and Medicare markets may be quite distinct, even within the same geographic area (see Chernew et al. (2010)).

[2] Disease price changes over time have also been documented in specific case studies such as heart attacks (Cutler et al. (1998)), cataracts (Shapiro et al. (2001)), and depression (Berndt et al. (2002)).

vary across the 85 MSAs that we study in this paper. There are measurable differences across aggregate MSA indexes. We find that the MSA with the 90th percentile MCE has an MCE that is 28 percent larger than the 10th percentile. We also find that the MCE, SPI and SUI produce vastly different pictures of variation across the country. For example, of the 85 MSAs analyzed in this study, Memphis, TN, ranks 37th in terms of its SPI. However, in terms of its MCE it ranks 76th. The relatively low level of medical-care spending per episode in Memphis is due to its relatively low utilization of services per episode—its SUI is ranked 82nd.

Looking at variation at the disease-level, we find that there exists a large degree of variation in spending across markets. The "typical" disease, as measured by a weighted average across diseases, has coefficients of variation for MCE, SPI, and SUI of 0.22, 0.16 and 0.17, respectively. Importantly, the source of the variation in spending depends on the particular disease being assessed. For example, the variation in service price is relatively large for pregnancy, while the variation in utilization is relatively large for depression.

Using the constructed indexes at the disease-level, we explore whether there are common price and utilization components that affect all diseases in an MSA. That is, can some MSAs be characterized as high utilization or low price areas across all diseases? Interestingly, MSA-specific factors explain only a small portion of the variation in spending patterns across disease categories—16 percent of the observed variation in utilization and 37 percent of the variation in prices. We find that differences in disease-specific variation within an MSA appear to cancel out when aggregating across diseases, leading to considerably smaller variation statistics for aggregate indexes. Specifically, we find that the coefficient of variation for the aggregate MCE index is 0.10 while the utilization index is just 0.06. Thus, it appears that averaging over diseases masks the underlying geographic variation in spending for specific diseases. This suggests that focusing on more aggregate measures may understate the actual variation in medical care practices across markets.

There are a few additional findings in this paper. First, whether one looks at service prices in the aggregate or at the disease level, the estimates reveal that variation in service price is particularly important in commercial markets, which contrasts with Medicare markets where researchers have concluded that variation is primarily driven by differences in utilization (see Gottlieb et al. (2010)). Second, we find a negative correlation between price and utilization, so that low utilization areas tend to be higher priced areas. Third, we demonstrate that the variation across markets is considerably larger when examining disease expenditures per capita without controlling for the treated prevalence of patients across areas.

2 Literature Review

There are many approaches to analyzing geographic differences in spending and utilization across markets.[3] Some research focuses on differences in treatment for certain diseases[4], while other studies examine aggregate differences in overall medical-care expenditures.[5] This paper combines aspects of both these approaches because it focuses on aggregate medical-care expenditures in a geographic area, but we break these aggregate expenditures into disease-level components. The key advantage is that this allows us to analyze the sources of expenditure differences in greater detail. For instance, the treatment of heart disease in an area may involve intense utilization, but lower prices; while treating lower back pain in the same location may involve low levels of intensity and high prices.

Previous studies also differ in the unit at which expenditures are measured. Some studies track expenditures on a per capita basis[6], while other studies look at expenditures per episode-of-care.[7] The decision to assess one or the other depends on the goal of the researchers. For instance, an aggregate measure of per-capita spending can shed light on the general health of the population—as this measure will be lower if a smaller proportion of the population needs medical treatment—while an aggregate measure of per-episode spending may be more informative about the efficiency of the providers in an area. Unlike the per-capita spending measure, the per-episode measure accounts for the health status or the likelihood of treatment and diagnosis across areas. Table 1 lists the MSAs with the five highest and lowest medical-care spending per person in our data set. This table shows that there can indeed be considerable differences between per-capita and per-episode spending measures. For instance, out of the 85 cities that we study, Birmingham has the 4th highest spending per capita, but ranks 17th in terms of spending per episode-of-care. Given that Alabama has one of the highest obesity rates in the country, it is quite possible that Birmingham's population-based measure is large relative to the episode-based measure

[3] Recent reviews of the literature on geographic variation in health care spending are in Congressional Budget Office (CBO) (2008) and Skinner (2012).

[4] For example, Chandra and Staiger (2007) look at different types of treatments for heart attack patients across markets.

[5] For example, Cutler and Sheiner (1999), MedPac (2003), Fuchs, McClellan and Skinner (2004), and Sheiner (2012).

[6] These studies include Cutler and Sheiner (1999), Gage, Moon and Chi (1999), Zuckerman et al. (2010), and Sheiner (2012).

[7] Many of the studies that look at expenditures per episode of care have focused on growth rates (e.g., Thorpe et al. (2004), Aizcorbe and Nestoriak (2011), Aizcorbe et al. (2010), Roehrig and Rousseau (2011), Dunn et al. (2012), and Bradley (2013)).

because the population is less healthy.[8] As can be seen by the standard errors reported in Table 1, the variation in expenditures within each MSA is substantial; however, as with most other papers in this literature, this paper compares average expenditures across markets in order to focus on systematic differences in spending.[9]

Table 1. Medical-Care Expenditures Per Person and Medical-Care Expenditures Per Episode

MSA	Rank	Medical-Care Expenditures Per Person	s.e.	Rank	Medical-Care Expenditures Per Episode	s.e.
Milwaukee-Waukesha-West Allis, WI	1	$2,937	($8,867)	1	$1,036	($4,411)
Salinas, CA	2	$2,904	($10,850)	4	$979	($5,517)
Fort Worth-Arlington, TX	3	$2,891	($10,051)	9	$948	($4,996)
Birmingham-Hoover, AL	4	$2,867	($8,357)	17	$871	($3,842)
Oakland-Fremont-Hayward, CA	5	$2,842	($10,234)	6	$963	($5,207)
Memphis, TN-MS-AR	81	$2,164	($6,077)	71	$742	($2,993)
Riverside-San Bernardino-Ontario, CA	82	$2,160	($8,975)	30	$837	($4,934)
MSA in the South	83	$2,046	($6,873)	27	$834	($3,611)
Las Vegas-Paradise, NV	84	$2,035	($8,028)	44	$792	($4,338)
Pittsburgh, PA	85	$2,031	($6,413)	75	$743	($3,190)

Notes. Medical-Care Expenditures per Person and Medical-Care Expenditures per Episode are based on a subset of the claims sample from the MarketScan® data base. Both the selected sample of claims and the MarketScan® data are described in greater detail in the following sections of this paper. The standard errors capture the level of expenditure variation within each MSA.

Similar to our work, Aizcorbe and Nestoriak (2011) track the "price" of treating a disease, or what we call the expenditure for an episode of care. Specifically, they compare an MCE index that allows expenditures to shift across providers to an index that holds the basket of services fixed (an SPI). Looking over time, they document several important shifts in utilization across provider types that drive a wedge between the two indexes. The existence of these observed shifts over time suggest that there may also be different allocations of services across geographic markets. Indeed, this paper shows how a simple service-price index that holds utilization fixed may be a misleading indicator for the "price" of treating a disease. This is because differences in utilization are also important in determining the level of expenditures across geographic markets.

We expand the basic methodology of Aizcorbe and Nestoriak (2011) in a few ways. Most importantly, we focus on differences across geographic markets, rather than changes over time. Second, our methodology for decomposing service price and service utilization starts from the most granular level to more precisely capture price and utilization differences. For example, rather than tracking the price per encounter with a doctor, we focus on the prices

[8] See statistics reported by the Center for Disease Control and Prevention (http://www.cdc.gov/obesity/data/adult.html).

[9] The per capita and per episode estimates reported in Table 1 are averages over thousands of individuals in each MSA and are precisely estimated.

of the particular procedures performed during the office visit.[10] This allows us to capture greater heterogeneity in the types of services performed across markets.[11] Furthermore, we extend our decomposition to relate our expenditure per episode measure to a measure of expenditure per capita.

3 Methodology of Index Construction

3.1 A Motivating Example

To help motivate our methodology, we start with a simple example. Consider a single MSA, r, where people are treated for hypertension (h) (i.e., high blood pressure) where there exists only one type of treatment available— a 15-minute office visit. Let

$$\begin{aligned} N_{h,r} &= \text{Number of treated hypertension episodes in the MSA.} \\ c_{h,r} &= \text{Average expenditure for hypertension per episode.}^{12} \\ q_{h,r} &= \text{Number of 15-minute office visits per episode.} \\ p_{h,r} &= \text{Price per 15-minute office visit (i.e., } \frac{c_{h,i}}{q_{h,i}}\text{).} \end{aligned}$$

Also suppose there is a comparison or base region, B, where the price for a 15-minute office visit for hypertension is $p_{h,B}$. In this simple case, the relative price level of r to B is simply $\frac{p_{h,r}}{p_{h,B}}$. Clearly, this ratio reflects only differences in the contracted prices, not the number of 15-minute office visits. Similarly, the relative utilization level is $\frac{q_{h,r}}{q_{h,B}}$ which depends only on the number of 15-minute office visits performed per episode. It follows that the relative expenditure per episode between r and B may be expressed as:

$$\frac{c_{h,r}}{c_{h,B}} = \left(\frac{p_{h,r} \cdot q_{h,B}}{p_{h,B} \cdot q_{h,B}}\right) \cdot \left(\frac{p_{h,r} \cdot q_{h,r}}{p_{h,r} \cdot q_{h,B}}\right). \tag{1}$$

[10] Bundorf et al. (2009) have a similar methodology for decomposing service-prices at the procedure level, however they do not analyze expenditures at the disease level. Dunn, Liebman, and Shapiro (2012) highlight the practical importance of using more granular services for measuring inflation.

[11] As is common in the literature, we apply a unit value index because of data constraints (e.g., Bundorf et al. (2009), Roehrig and Rousseau (2011), Thorpe et al. (2004), and Aizcorbe and Nestoriak (2011)). The application of unit value indexes in health care is widely used and this is discussed in greater detail in National Research Council (2010). Because we drill down to the procedure level (rather than the more-coarse encounter level), our resulting price indexes (based on unit values) are arguably more likely free of bias than indexes that are based on encounters.

The first term in (1) is a price index, and the second term is a utilization index. Expanding on this example, now suppose that hypertension 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, $q_{h,r,O}$ and $p_{h,r,O}$ are the utilization and price for the physician office visits, and $q_{h,r,D}$ and $p_{h,r,D}$ are the utilization and price for prescription drugs. Continuing with the index decomposition that is parallel to (1), but with two services, the decomposition becomes:

$$\frac{c_{h,r}}{c_{h,B}} = \frac{p_{h,r,O} \cdot q_{h,r,O} + p_{h,r,D} \cdot q_{h,r,D}}{p_{h,B,O} \cdot q_{h,B,O} + p_{h,B,D} \cdot q_{h,B,D}} \tag{2}$$

$$= \left(\frac{p_{h,r,O} \cdot q_{h,B,O} + p_{h,r,D} \cdot q_{h,B,D}}{p_{h,B,O} \cdot q_{h,B,O} + p_{h,B,D} \cdot q_{h,B,D}}\right) \cdot \left(\frac{p_{h,r,O} \cdot q_{h,r,O} + p_{h,r,D} \cdot q_{h,r,D}}{p_{h,r,O} \cdot q_{h,B,O} + p_{h,r,D} \cdot q_{h,B,D}}\right) \tag{3}$$

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

3.2 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.[13] Formally, denote the average expenditure paid to medical providers for an episode of treating disease d for MSA r as $c_{d,r}$. 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:

$$MCE_{d,r} = \frac{c_{d,r}}{c_{d,B}}. \tag{4}$$

Thus, similar to the example above, if the $MCE_{d,r}$ is larger than one, it signifies that the expenditure for treating disease d is larger than average (or what we call the "base" area) and if the index is less than one it signifies that the expenditure is less than the average.

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

[13] 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.

dividing those dollars by the number of episodes: $c_{d,r} = \sum_s p_{d,r,s} Q_{d,r,s}/N_{d,r}$, where $Q_{d,r,s}$ is the quantity of services for service type, s; $p_{d,r,s}$, is the service price for service type s; and $N_{d,r}$ is the number of episodes treated. To simplify, let $q_{d,r}$ be a vector of services utilized for the typical treatment of diseases in an MSA, $q_{d,r} = Q_{d,r}/N_{d,r}$, where the component of the utilization vector for service type s is, $q_{d,r,s} = Q_{d,r,s}/N_{d,r}$. Similarly, let $p_{d,r}$ 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: $p_{d,r,s} = \frac{c_{d,r,s}}{q_{d,r,s}}$ where $c_{d,r,s}$ is the average episode expenditure for disease d for service type s in MSA r. This decomposition allows us to create a service price and service utilization index. The service price index (SPI) is then calculated as:

$$SPI_{d,r} = \frac{p_{d,r} \cdot q_{d,B}}{c_{d,B}}, \tag{5}$$

which holds the utilization of services fixed at a base level. The SPI measures the compensation necessary to purchase a fixed utilization of medical goods when moving from the average national base to a particular MSA. The service utilization index (SUI) may be defined as:

$$SUI_{d,r} = \frac{p_{d,B} \cdot q_{d,r}}{c_{d,B}}, \tag{6}$$

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 in an MSA at fixed national base prices when moving from the average national base to a particular MSA. 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 different than 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.[14] The relationship between these three indexes is described by the following decomposition:

[14]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.

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

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

4 Data

We use retrospective claims data for a convenience sample of commercially-insured patients from the MarketScan® Research Database from Truven Health. The specific claims data used is the "Commercial Claims and Encounters Database," which contains data contributed by employers and health insurance plans. The data includes medical and drug claims 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" 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 2006 and 2007. We also limit our sample to enrollees with drug benefits because drug purchases will not be observed for individuals without drug coverage.[15] The MarketScan database tracks claims from all providers using a nationwide convenience sample of enrollees. Each enrollee has a unique identifier and can be linked to a particular MSA. All claims have been paid and adjudicated.[16] However, it should be noted that the drug payment information excludes potential rebates from drug manufacturers, so price levels are likely overstated relative to the actual prices that include rebates.

The claims data has been processed using the Symmetry grouper 7.6 from Optum. The grouper assigns each claim to a particular Episode Treatment Group (ETG) disease

[15] In our selected sample, we find about 62 percent of the data is from employer plans while the remaining 38 percent is from health insurance plans.

[16] Additional details about the data and the grouper used in this paper are in Dunn et al. (2010).

category.[17] The grouper uses a proprietary algorithm, based on clinical knowledge, that is applied to the claims data to assign each record to a clinically homogenous episode of care. The episode grouper allocates all spending from individual claim records to distinct diseases.[18] An advantage of using the grouper is that it uses the 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 expertise of those that developed the grouper software. To ensure that we could properly identify all the claims for each individual's episode, we require full enrollment for the entire year, plus 6 months prior enrollment (e.g., enrollment until July 2005 for enrollees in 2006) and 6 months post enrollment (e.g., enrollment until June 2008 for enrollees in 2007).[19]

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.

4.1 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.[20] 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

[17] The ETG grouper allocates each record into one of over 500 disease groups.

[18] All episodes are initiated using only diagnostic information, so information on services or procedures performed are not used to initiate episodes. In cases where the spending could potentially be allocated to multiple diseases, the grouper uses additional information on the claim, such as the information from the patient's history or the types of procedures performed to allocate spending across disease episodes.

[19] About 13.8 percent of expenditures are not assigned to any ETG disease category (that is screening for diseases and other records that cannot be assigned a category). Those claims that are not assigned disease categories are removed from our analysis.

The six month "cushion" ensures that episodes occurring at the beginning or the end of a year are not truncated. The results do not appear sensitive to this six month cushion. We obtain similar results when there is no cushion or when the cushion is for an entire year.

[20] For an episode to fall into the sample, the episode must end in the 2006 or 2007 year of the data. Episodes records that begin in 2005 and end in 2006 or 2007 are included in this study, while episodes that begin in 2007 and end in 2008 are not included.

firms).

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" 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.

4.1.1 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 visit, where a visit is defined by the enrollee and the date of service or admission:

Physician office - Physician visits 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 from a physician office visit is computed as $q_{office} = \sum_{cpt \in Visit} \overline{p}_{cpt,office}$, where $cpt \in Visit$ is a complete list of CPT procedures performed during the visit 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 in an MSA, r, as the base price multiplied by a scalar price, α_r, where $p^r_{cpt} = \alpha_r \overline{p}_{cpt}$. For instance, if a CPT code that equals 99213 indicating a 15-minute established patient office visit 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 a 15-minute office visit is \$120 in MSA r ($p^r_{99213} = \$120$), then the price of the service will be calculated as \$120/100RVU=1.2 \$/RVU (i.e., $\alpha_r = \frac{p^r_{cpt}}{\overline{p}_{cpt}}$).[21]

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, $q_{inpatient} = \overline{p}_{d,inpatient} + \sum_{cpt \in Visit} \overline{p}_{cpt,inpatient}$ 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 a visit to an inpatient facility for treating disease d. The term $\sum_{cpt \in Visit} \overline{p}_{cpt,inpatient}$ 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 visits are calculated in an identical fashion to the inpatient hospital visits. That is, the facility amount is calculated based on the average outpatient visit 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 $q_{drug} = \sum_{NDC \in Visit} \overline{p}_{NDC}$, where $NDC \in Visit$ is a complete list of NDC codes purchased from a visit to 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 a visit 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

[21] This methodology for calculating utilization for physician services is identical to that conducted by Dunn and Shapiro (2012).

average cost of that procedure code for that place of service.

Our decomposition relies on the institutional feature that insurers and providers typical negotiate from a percentage of a base fee schedule (for example, 10 percent above Medicare rates).[22] As our measure of service price can be intuited as the expenditures from a visit 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.[23] 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.[24] See Section 1 of the Online Appendix for additional details.

[22] 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. Gowrisankaran, Nevo, and Town (2013) incorporate this assumption in their bargaining study of hospital prices: "We assume that the price paid for treatment is ... the base price multiplied by the disease weight. This is essentially how most hospitals are reimbursed by Medicare, and many [Insurers] incorporate this payment structure into their hospital contracts."

[23] Note that our approach differs from Bundorf et al. (2009) which also studied individual service prices over time, though not at the disease level. Since they do not calculate disease-level prices, they are able to separately price each individual CPT code and NDC code. In contrast, this is not possible in a disease-based framework because not all CPT codes and NDC codes are observed for every disease across all MSAs.

Alternatively, we could have aggregated to the state-level and broad disease categories, to ensure that more of these cells are populated. However, this would also result in a significant loss in information about geographic differences. As a partial check on our assumption in practice, Dunn, Liebman and Shapiro (2012) look at service price inflation over time using the pricing methodology proposed here compared to the methodology applied in Bundorf et al. (2009). They find similar rates of inflation for each service category.

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

$\mathrm{SPI}_{Lasp} = \frac{P_{1,r} \cdot Q_{1,B} + P_{2,r} \cdot Q_{2,B} ... + P_{N,r} \cdot Q_{N,B}}{P_{1,B} \cdot Q_{1,B} + P_{2,B} \cdot Q_{2,B} ... + P_{N,B} \cdot Q_{N,B}}$. Assuming that physicians set prices from a base fee schedule, then the pricesin MSA r can be computed as α_r times the base fee schedule. That is, $P_{1,r} = \alpha_r \ P_{1,B}$, $P_{2,r} = \alpha_r \ P_{2,B}, ...,$ and $P_{N,r} = \alpha_r \ P_{N,B}$, so

$\mathrm{SPI}_{Lasp} = \frac{P_{1,r} \cdot Q_{1,B} + P_{2,r} \cdot Q_{2,B} ... + P_{N,r} \cdot Q_{N,B}}{P_{1,B} \cdot Q_{1,B} + P_{2,B} \cdot Q_{2,B} ... + P_{N,B} \cdot Q_{N,B}}$
$= \frac{\alpha_r (P_{1,B} \cdot Q_{1,B} + P_{2,B} \cdot Q_{2,B} ... + P_{B,r} \cdot Q_{N,B})}{P_{1,B} \cdot Q_{1,B} + P_{2,B} \cdot Q_{2,B} ... + P_{N,B} \cdot Q_{N,B}} = \alpha_r$. 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.

4.2 Selected Sample and Descriptive Statistics

When studying variation across MSAs, there is some concern that we have a large enough sample within each MSA so that an average over the population will be meaningful. To ensure that each MSA has a sufficient number of individuals, we select only those MSAs in the data that have an average of 20,000 enrollees per year over the 2006-2007 time period (that is 40,000 enrollee year observations). The minimum sample size in each city is more than double the sample size of the commercially-insured sample from the Medical Expenditure Panel Survey (MEPS), a national survey of health expenditures meant to be representative of the entire U.S. non-institutionalized population.[25] This first selection rule leaves a sample of 85 MSAs.[26]

The variation in severity may be large for some disease categories. To account for differences in disease severity, we define each disease as an ETG- severity combination, so that each ETG-severity combination will be examined separately. Since we may not have precise estimates for infrequently observed diseases, we select those diseases for which we observe 10,000 episodes in the data for the selected MSAs. Those diseases with 10,000 or more episodes account for 78 percent of overall expenditures and 96 percent of the episodes.[27]

Enrollees are assigned weights so the weighted population in each MSA has an age and sex distribution that is identical to that of the U.S. commercially-insured population.[28]

[25] The commercially-insured sample in the MEPS data is around 15,000 individual observations in each year. In this study we are using two years of data which includes more than 40,000 individual-year observations per MSA.

[26] These 85 MSAs account for 70 percent of the enrollment population available in the MarketScan data that are located in an identifiable MSA for these years of study. For each MSA reported in this analysis, at least three unique employers contribute to the data in each MSA and at least three unique carriers also contribute to this data.

[27] The results in this paper are not sensitive to either the selection rule for the diseases or the MSAs. The results also look very similar when we do not control for the severity of the disease. These robustness checks are outlined in greater detail in the online appendix to this paper.

[28] 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 commercial population is obtained from the Medical Expenditure Panel Survey.

MSA observations in 2006 and 2007 are each weighted to the national level population in 2007. That is, the sample in 2006 is weighted to the 2007 national population *and* the sample in 2007 is also weighted to the 2007 national population. After weighting the populations to the national level, the data is aggregated over the two years. This ensures that 2006 and 2007 receive equal weights in the price index, even if the enrollment within an MSA changes over these years.

We have conducted similar analysis looking at only 2006 and only 2007 year data. We obtain similar

Table 2 provides some basic descriptive statistics for the selected population and the overall disease expenditures for the selected diseases. The top 5 ETG categories, based on overall expenditure, are ischemic heart disease, pregnancy, joint degeneration of the back, hypertension and diabetes. Although there are 310 diseases in the sample, these first 5 ETG categories (16 diseases) account for 21 percent of the expenditures. In general, most of the expenditures are accounted for by a limited number of diseases with the top 15 ETG categories listed here accounting for 42 percent of total expenditures from the selected diseases, so the aggregate price index will be heavily influenced by the high-spending diseases. There is a wide range in the expenditure per episode across diseases. Severity 1 Hypertension costs just $642 per episode, while Severity 3 Ischemic Heart Disease costs $20,220.

Table 2. Average Annual Expenditures and Average Number of Episodes - Weighted to U.S. Totals for Commercial Insurance

	Description	Severity	Total Dollars (Billions)	Episodes (Thousands)	Dollars Per Episode	s.e. Dollars Per Episode	Fraction of Spending	Fraction of Spending Category
1	Pregnancy, with delivery	1	$14.09	1,497	$9,414	($5,015)	3.2%	4.8%
	Pregnancy, with delivery	2	$7.11	513	$13,845	($9,725)	1.6%	
2	Joint degeneration, localized - back	1	$10.96	6,249	$1,755	($5,225)	2.5%	4.4%
	Joint degeneration, localized - back	2	$4.84	1,157	$4,182	($11,518)	1.1%	
	Joint degeneration, localized - back	3	$3.74	302	$12,370	($23,294)	0.8%	
3	Ischemic heart disease	1	$8.61	2,440	$3,530	($8,334)	2.0%	4.4%
	Ischemic heart disease	2	$6.71	1,224	$5,484	($13,816)	1.5%	
	Ischemic heart disease	3	$4.02	199	$20,220	($28,955)	0.9%	
4	Hypertension	1	$11.29	17,575	$642	($1,128)	2.6%	4.0%
	Hypertension	2	$3.34	3,845	$868	($1,832)	0.8%	
	Hypertension	3	$1.72	1,591	$1,082	($2,741)	0.4%	
	Hypertension	4	$1.36	595	$2,288	($8,678)	0.3%	
5	Diabetes	1	$9.83	6,414	$1,532	($5,537)	2.2%	3.7%
	Diabetes	2	$1.78	730	$2,431	($4,507)	0.4%	
	Diabetes	3	$1.77	541	$3,281	($5,182)	0.4%	
	Diabetes	4	$2.76	478	$5,764	($12,274)	0.6%	
6	Routine exam	1	$13.62	63,192	$216	($224)	3.1%	3.1%
7	Mood disorder, depressed	1	$8.62	7,202	$1,197	($1,999)	2.0%	2.6%
	Mood disorder, depressed	2	$1.81	1,135	$1,591	($3,337)	0.4%	
	Mood disorder, depressed	3	$1.09	352	$3,092	($6,981)	0.2%	
8	Malignant neoplasm of breast	1	$6.13	820	$7,483	($19,001)	1.4%	2.4%
	Malignant neoplasm of breast	2	$4.30	225	$19,104	($34,591)	1.0%	
9	Hyperlipidemia, other	1	$10.24	15,886	$645	($976)	2.3%	2.3%
10	Joint degeneration, localized - neck	1	$6.21	4,114	$1,509	($4,084)	1.4%	2.2%
	Joint degeneration, localized - neck	2	$0.88	367	$2,387	($6,483)	0.2%	
	Joint degeneration, localized - neck	3	$2.55	287	$8,880	($15,414)	0.6%	
11	Chronic sinusitis	1	$5.09	9,978	$510	($1,217)	1.2%	1.9%
	Chronic sinusitis	2	$1.16	1,319	$879	($1,932)	0.3%	
	Chronic sinusitis	3	$2.00	866	$2,304	($4,214)	0.5%	
12	Joint degeneration, localized - knee & lower leg	1	$5.17	2,314	$2,234	($6,777)	1.2%	1.7%
	Joint degeneration, localized - knee & lower leg	2	$1.17	313	$3,731	($9,362)	0.3%	
	Joint degeneration, localized - knee & lower leg	3	$1.36	201	$6,735	($13,738)	0.3%	
13	Asthma	1	$2.46	4,264	$576	($1,121)	0.6%	1.7%
	Asthma	2	$3.38	3,455	$977	($1,876)	0.8%	
	Asthma	3	$0.66	345	$1,927	($3,382)	0.2%	
	Asthma	4	$1.08	293	$3,672	($8,054)	0.2%	
14	Joint derangement - knee & lower leg	1	$1.23	701	$1,755	($3,609)	0.3%	1.7%
	Joint derangement - knee & lower leg	2	$6.14	1,180	$5,202	($5,740)	1.4%	
15	Inflammation of esophagus	1	$4.88	3,706	$1,317	($2,988)	1.1%	1.6%
	Inflammation of esophagus	2	$1.37	702	$1,951	($3,608)	0.3%	
	Inflammation of esophagus	3	$0.91	264	$3,447	($9,454)	0.2%	
	Other		$253.52	371,975	$682		57.5%	57.5%
	Total		$441.06	541,387	$809		100.0%	

Notes. The national weights are applied to each city and the total expenditures and episodes are divided by the number of cities in our sample, 85, times the number of years of data, 2 (Thus we divide by 170 (=85*2)). Therefore, these figures actually overcount smaller MSAs included in the sample, relative to their share of the U.S. population. We equally count MSAs in this table because our base expenditure is constructed to measure the cost of a specific disease for a typical person in an MSA, not the cost of specific disease for a person in the U.S. population. Recalculating this table weighting by each MSA's population, we find that the fraction of spending for each disease category changes only slightly and the expenditures per episode increases by a very small amount, from $809 per episode to $813.

The standard error in dollars per episode reported in Table 2 is quite high, often

results in each year.

implying a coefficient of variation above 1. However, it is important to emphasize that the MCE, SUI, and SPI measures are constructed as averages at the MSA-disease level, so we are measuring the expenditure for a *typical* patient, which is much more precisely estimated. As stated previously, this approach is quite common in the geographic variation literature, which focuses on *systematic* differences across areas.

Table 3 shows the key service types that are studied in the paper along with the amount and expenditure share associated with each type.

Table 3. Average Annual Spending Share Across Services - Weighted to U.S. Totals for Commercial Insurance

Place of Service	Total Spending (Billions)	Share of Spending
Inpatient Hospital	$81.59	18.5%
Outpatient Hospital	$101.71	23.1%
Office - General MD	$39.00	8.8%
Office - Specialist MD	$69.82	15.8%
Other (Emergency, Ambulatory Centers etc)	$52.96	12.0%
Pharmacy	$95.98	21.8%
Total	$441.06	100.0%

Notes. Similar to Table 2, the national weights are applied to each city and the total expenditures and episodes are divided by the number of cities in our sample, 85, times the number of years of data, 2 (Thus we divide by 170 (=85*2)).

5 Results

5.1 MSA-disease Indexes

Table 4 reports the coefficient of variation (COV) of the $MCE_{d,r}$, $SPI_{d,r}$ and $SUI_{d,r}$ for the 15 largest ETG categories in the data ranked by expenditures. Severity 4 Hypertension has the largest variation in expenditure per episode across areas with a COV of 0.35, and Hyperlipidemia (high cholesterol) and Severity 1 Hypertension have the lowest COV of 0.10. The weighted average COV for the $MCE_{d,r}$, $SPI_{d,r}$, and $SUI_{d,r}$ are 0.22, 0.16, and 0.17 for the full sample.[29]

The underlying cause for the variation may be attributed to either utilization or price, which are shown using the COV of the $SPI_{d,r}$ and $SUI_{d,r}$. For some diseases, expenditure differences are primarily affected by price, while for other diseases utilization variation is more important. For example, the Severity 1 Mood Disorder, Depression, has a relatively low price variation. This could potentially be explained by a preference for talk therapy in some areas and drug treatment in others, although the prices for each of these ser-

[29] The COV remains large even when focusing on the most frequently observed diseases. The bottom of the table reports the weighted average COV for those diseases with more than 50,000 episodes in the data. These diseases account for around 65 percent of the expenditure from the selected sample.

vices may not vary by a large degree across areas. In contrast, a disease like Severity 1 Pregnancy or the cost associated with a Routine Exam have relatively little variation in utilization compared to variation in price across areas. One possibility is that treatments for these diseases are relatively clear, although the prices for the underlying services vary substantially across markets.

Table 4. Sources of Price Variation Across MSAs by Disease - $MCE_{d,r}$, $SPI_{d,r}$ and $SUI_{d,r}$

	Description	Severity	COV of $MCE_{d,r}$	s.e.	COV of $SPI_{d,r}$	s.e.	COV of $SUI_{d,r}$	s.e.
1	Pregnancy, with delivery	1	0.18	(0.009)	0.17	(0.004)	0.04	(0.005)
	Pregnancy, with delivery	2	0.20	(0.015)	0.19	(0.008)	0.05	(0.006)
2	Joint degeneration, localized - back	1	0.18	(0.007)	0.13	(0.004)	0.15	(0.006)
	Joint degeneration, localized - back	2	0.28	(0.024)	0.18	(0.017)	0.18	(0.012)
	Joint degeneration, localized - back	3	0.29	(0.022)	0.20	(0.014)	0.17	(0.015)
3	Ischemic heart disease	1	0.22	(0.010)	0.17	(0.009)	0.17	(0.007)
	Ischemic heart disease	2	0.22	(0.016)	0.22	(0.021)	0.15	(0.013)
	Ischemic heart disease	3	0.33	(0.055)	0.24	(0.030)	0.17	(0.018)
4	Hypertension	1	0.10	(0.002)	0.09	(0.002)	0.11	(0.002)
	Hypertension	2	0.11	(0.007)	0.10	(0.008)	0.12	(0.004)
	Hypertension	3	0.14	(0.029)	0.13	(0.035)	0.12	(0.012)
	Hypertension	4	0.35	(0.099)	0.33	(0.056)	0.24	(0.038)
5	Diabetes	1	0.11	(0.006)	0.06	(0.006)	0.10	(0.003)
	Diabetes	2	0.18	(0.026)	0.13	(0.043)	0.15	(0.018)
	Diabetes	3	0.20	(0.015)	0.11	(0.015)	0.16	(0.009)
	Diabetes	4	0.21	(0.023)	0.17	(0.021)	0.14	(0.015)
6	Routine exam	1	0.15	(0.001)	0.12	(0.001)	0.06	(0.001)
7	Mood disorder, depressed	1	0.20	(0.004)	0.06	(0.004)	0.17	(0.003)
	Mood disorder, depressed	2	0.23	(0.015)	0.11	(0.022)	0.18	(0.009)
	Mood disorder, depressed	3	0.24	(0.024)	0.20	(0.042)	0.18	(0.012)
8	Malignant neoplasm of breast	1	0.20	(0.027)	0.14	(0.009)	0.16	(0.022)
	Malignant neoplasm of breast	2	0.27	(0.026)	0.18	(0.013)	0.20	(0.014)
9	Hyperlipidemia, other	1	0.10	(0.002)	0.08	(0.003)	0.10	(0.002)
10	Joint degeneration, localized - neck	1	0.19	(0.007)	0.11	(0.005)	0.18	(0.007)
	Joint degeneration, localized - neck	2	0.31	(0.031)	0.19	(0.035)	0.27	(0.025)
	Joint degeneration, localized - neck	3	0.30	(0.029)	0.23	(0.016)	0.22	(0.015)
11	Chronic sinusitis	1	0.17	(0.003)	0.08	(0.002)	0.13	(0.003)
	Chronic sinusitis	2	0.20	(0.027)	0.09	(0.005)	0.16	(0.017)
	Chronic sinusitis	3	0.25	(0.015)	0.17	(0.010)	0.18	(0.009)
12	Joint degeneration, localized - knee & lower leg	1	0.25	(0.020)	0.16	(0.008)	0.17	(0.009)
	Joint degeneration, localized - knee & lower leg	2	0.31	(0.028)	0.18	(0.017)	0.23	(0.021)
	Joint degeneration, localized - knee & lower leg	3	0.29	(0.023)	0.24	(0.057)	0.28	(0.019)
13	Asthma	1	0.11	(0.006)	0.06	(0.004)	0.10	(0.007)
	Asthma	2	0.11	(0.006)	0.07	(0.003)	0.10	(0.006)
	Asthma	3	0.31	(0.189)	0.17	(0.035)	0.16	(0.037)
	Asthma	4	0.24	(0.054)	0.28	(0.154)	0.25	(0.036)
14	Joint derangement - knee & lower leg	1	0.26	(0.017)	0.17	(0.009)	0.21	(0.020)
	Joint derangement - knee & lower leg	2	0.21	(0.008)	0.21	(0.005)	0.18	(0.011)
15	Inflammation of esophagus	1	0.12	(0.004)	0.10	(0.008)	0.09	(0.003)
	Inflammation of esophagus	2	0.14	(0.015)	0.12	(0.017)	0.11	(0.013)
	Inflammation of esophagus	3	0.33	(0.077)	0.20	(0.042)	0.22	(0.046)
	Weighted Average (Full Sample - 10,000 Episodes in the Data)		0.223		0.161		0.167	
	Weighted Average (Only Diseases with 50,000 Episodes in the Data)		0.178		0.131		0.140	

Notes. Standard errors are calculated using a bootstrap with 200 random draws of the sample with replacement.

The measures of variation are large, but appear smaller than many studies of variation, such as those reported by the Dartmouth Atlas group. The most likely reason for the smaller variation is that our estimates are conservative in many dimensions. First, as is common in the literature, we control for age and sex distribution of each MSA by applying population weights, making the population distribution in each area more homogeneous. Second, in contrast to much of the Dartmouth research, we control for the disease of the patient, which limits the amount of variation that we observe. That is, we look at variation, conditional on a patient being treated for a disease, while the Dartmouth research often looks at variation in the use of particular procedures. The last section of this paper shows that controlling for the disease reduces the average expenditure variation by over 25 percent. Third, to ensure that variation is not arising from small sample sizes, we focus on MSAs with a large number of enrollees. In Section 4 of the Online Appendix, we show using both the MarketScan data and data from the Dartmouth Atlas, that including only the larger 85 geographic areas (based on enrollment or population), rather than all areas, reduces disease-specific variation by at least 30 percent. We take these steps so that we are better able to isolate the systematic differences in treatment across regions for patients with similarly diagnosed diseases.

5.1.1 Sources of Variation

To get a better sense of the source of variation for service utilization and service prices, we estimate several regression models to determine how much of the $MCE_{d,r}$ variation, as well as $SUI_{d,r}$ and $SPI_{d,r}$ variation, may be explained by common MSA factors and MSA disease-category factors. More specifically, we run regressions of $\log(MCE_{d,r})$, $\log(SUI_{d,r})$ and $\log(SPI_{d,r})$ on three distinct fixed-effect models: (1) include only MSA fixed effects; (2) include MSA-Major Practice Category (MPC) fixed effects, where MPCs are 21 broadly-categorized disease groups; and (3) include MSA-ETG Category fixed effects.[30]

The results of these regressions are shown in Table 5. MSA fixed effects explain about one fifth of the variation for the MCE. Interestingly, the MSA fixed effects explain a larger portion of the variation in the $SPI_{d,r}$ regression, relative to the $SUI_{d,r}$ regression. Specifically, the R^2 is 0.16 for the $SUI_{d,r}$, compared to 0.37 for the $SPI_{d,r}$. A likely reason for this difference is that many prices are determined by contracts that are set by insurers and providers, regardless of the illness of the patient, while factors that impact utilization may be more idiosyncratic and reflect the norms among physicians for treating a particular disease in an area. Including the MSA-MPC fixed effects more than doubles the R^2 for the $SUI_{d,r}$ and also increases the R^2 for the $SPI_{d,r}$ by about 60 percent. This suggests that there are important disease-specific factors within each market that cause variation in utilization and price.[31] Therefore, within an MSA there is a large degree of heterogeneity in utilization patterns among disease groups.

[30] For each regression, we include only those diseases with multiple severities. Similar results are found if we focus on all diseases and compare regressions with MSA fixed effects to regressions with MSA-MPC fixed effects.

[31] Similar results are found if the threshold for the number of diseases observed in the data is increased from 10,000 to 50,000.

Table 5. Decomposing the Sources of Service Utilization and Service Price Variation

Log($MCE_{d,r}$)

	R^2	Adj R^2	MSE	N
MSA-FE	0.2185	0.2033	0.19258	11305
MSA-MPC-FE	0.4642	0.3837	0.16938	11305
MSA-ETG Category-FE	0.7201	0.5171	0.14993	11305

Log($SPI_{d,r}$)

	R^2	Adj R^2	MSE	N
MSA-FE	0.3672	0.3549	0.1289	11305
MSA-MPC-FE	0.6026	0.5429	0.1085	11305
MSA-ETG Category-FE	0.8077	0.6682	0.09244	11305

Log($SUI_{d,r}$)

	R^2	Adj R^2	MSE	N
MSA-FE	0.1635	0.1472	0.15384	11305
MSA-MPC-FE	0.4176	0.3301	0.13634	11305
MSA-ETG Category-FE	0.7112	0.5018	0.11758	11305

Notes. Based on regressions of log($MCE_{d,r}$), log($SUI_{d,r}$), and log($SPI_{d,r}$) for those diseases that have more than one severity. All regressions include disease-severity fixed effects. Similar results are found when one includes all diseases and compares the fit of the model with MSA fixed effects to the fit with MPC-MSA fixed effects.

5.2 MSA Indexes

To examine differences in spending, utilization, and prices across MSAs we average the $MCE_{d,r}$, $SUI_{d,r}$, $SPI_{d,r}$ over diseases for each MSA. We create MSA indexes by weighting each MSA-disease-specific index by the expenditure share of that disease for the entire U.S. to create MCE_r, SUI_r, and SPI_r. This weighting keeps the proportion of diseases fixed for each MSA and allows us to compare MSAs by looking at a fixed basket of diseases.

Table 6 below shows the MCE_r for each MSA in the data, although some MSAs are not shown due to confidentiality concerns. The MCE_r ranges from a high of 1.29 in Milwaukee to a low of 0.79 in Youngstown, PA. These indexes are estimated precisely, with many of the MSA indexes showing statistically significant differences. For instance, applying a t-test, the MCE in Milwaukee is significantly different than the MCE in Salinas at the 10 percent level. The table also reports SPI_r, which reflects differences in service prices, and SUI_r, which reflects differences in service utilization. A glance at this table shows that the underlying cause for a high MCE_r may be due to either higher service prices, higher service utilization, or a combination of the two. For example, it appears that Milwaukee, WI, has a high MCE_r primarily because it has a high SPI_r of 1.25, although the SUI_r is close to 1, the national average. In contrast, Gary, IN, has higher than average expenditures primarily because of service utilization, while the SPI_r is close to the national average. Other MSAs, such as Chicago, have higher expenditures due to higher than average prices and utilization.[32]

The variation in these aggregate MSA indexes give some measure of the overall spending variation. The COV for the aggregate MCE_r and SPI_r is about 0.10 and the COV for the SUI_r is just 0.064. These are relatively low measure of across-market variation compared to other goods and services.[33] Data from Aten and D'Souza (2008) imply a COV of 0.15 based on a price index for all goods and services across a subset of MSAs used in our study.[34] The CBO uses statistics from the BLS for the years 2004-2005 for a select sample of 24 cities and finds coefficients of variation for spending of 0.12 for food, 0.143 for housing, and 0.143 for transportation.

Table 6. MSA Medical-Care Price Indexes and Variation in Indexes - MCE_r, SPI_r and SUI_r

MSA Name	Rank MCE_r	MCE_r	s.e.	Rank SPI_r	SPI_r	s.e.	Rank SUI_r	SUI_r	s.e.
Milwaukee-Waukesha-West Allis, WI	1	1.288	(0.010)	3	1.254	(0.006)	33	1.011	(0.006)
Salinas, CA	2	1.245	(0.020)	1	1.391	(0.017)	84	0.888	(0.008)
MSA in the Midwest	3	1.240	(0.012)	6	1.149	(0.008)	10	1.069	(0.008)
Oakland-Fremont-Hayward, CA	4	1.219	(0.014)	4	1.251	(0.011)	57	0.971	(0.007)
Minneapolis-St. Paul-Bloomington, MN-WI	5	1.170	(0.010)	8	1.122	(0.005)	16	1.047	(0.007)
MSA in the Midwest	6	1.165	(0.005)	15	1.093	(0.003)	13	1.058	(0.003)
Fort Worth-Arlington, TX	7	1.143	(0.007)	10	1.114	(0.004)	28	1.022	(0.004)
Indianapolis, IN	8	1.138	(0.006)	12	1.105	(0.005)	34	1.010	(0.004)
Gary, IN	9	1.136	(0.009)	44	0.997	(0.005)	4	1.124	(0.007)
MSA in the West	10	1.133	(0.013)	2	1.262	(0.010)	80	0.902	(0.006)
Dallas-Plano-Irving, TX	11	1.127	(0.005)	7	1.129	(0.004)	36	1.004	(0.003)
Peoria, IL	12	1.121	(0.007)	9	1.120	(0.004)	25	1.024	(0.005)
Houston-Sugar Land-Baytown, TX	13	1.120	(0.005)	20	1.068	(0.003)	17	1.040	(0.003)
Miami-Miami Beach-Kendall, FL	14	1.102	(0.012)	22	1.063	(0.008)	20	1.036	(0.009)
Denver-Aurora, CO	15	1.102	(0.010)	33	1.026	(0.006)	11	1.065	(0.007)
Las Vegas-Paradise, NV	71	0.902	(0.006)	34	1.026	(0.005)	77	0.924	(0.004)
Nassau-Suffolk, NY	72	0.902	(0.008)	30	1.031	(0.007)	83	0.890	(0.006)
Louisville, KY-IN	73	0.901	(0.006)	65	0.947	(0.004)	68	0.936	(0.005)
MSA in the South	74	0.899	(0.007)	24	1.051	(0.006)	85	0.870	(0.005)
MSA in the South	75	0.898	(0.008)	64	0.948	(0.005)	65	0.949	(0.006)
Memphis, TN-MS-AR	76	0.891	(0.004)	37	1.018	(0.004)	82	0.895	(0.003)
Providence-New Bedford-Fall River, RI-MA	77	0.884	(0.008)	82	0.878	(0.004)	27	1.023	(0.007)
Kingsport-Bristol-Bristol, TN-VA	78	0.881	(0.007)	81	0.878	(0.004)	31	1.014	(0.007)
Warren-Farmington Hills-Troy, MI	79	0.872	(0.002)	78	0.896	(0.003)	39	0.998	(0.007)
MSA in the South	80	0.871	(0.006)	60	0.953	(0.005)	73	0.927	(0.005)
MSA in the Midwest	81	0.865	(0.005)	70	0.925	(0.004)	67	0.937	(0.005)
Augusta-Richmond County, GA-SC	82	0.862	(0.008)	71	0.925	(0.006)	76	0.924	(0.006)
Detroit-Livonia-Dearborn, MI	83	0.861	(0.004)	79	0.893	(0.002)	48	0.985	(0.005)
MSA in the South	84	0.844	(0.004)	77	0.906	(0.003)	70	0.933	(0.003)
Youngstown-Warren-Boardman, OH-PA	85	0.793	(0.006)	85	0.821	(0.003)	54	0.977	(0.005)
mean		1.000			1.018			0.996	
sd		0.098			0.097			0.064	
COV		0.098			0.095			0.064	
p10		0.887			0.915			0.920	
p90		1.142			1.124			1.073	
N		85			85			85	

Notes. Standard errors are calculated using a bootstrap with 200 random draws of the sample with replacement.

[32] Recall that the MCE is only approximately equal to $SPI + SUI - 1$, since there is also a cross term that explains the remaining difference.

[33] The estimate is consistent with other measures of medical care spending variation. The level of variation is similar to the state-level per capita spending variation computed by the Congressional Budget Office (CBO) (2008) for 2004 of around 0.125.

[34] We calculate this COV using their data for a sample of 70 cities that match to our sample of MSAs.

Although the variation appears to be similar to, or lower than, other aggregate measures of spending variation, much of the variation across markets appears to be smoothed out through the aggregation of the disease-specific indexes up to the MSA level. For instance, the COV of the MSA utilization index, SUI_r (0.064), is less than half of the size of the average COV of the disease-specific index $SUI_{d,r}$ (0.17).[35] This finding corresponds with the fixed-effects regressions in the previous section, which suggest that there is a large degree of heterogeneity in utilization patterns among disease groups within an MSA. Thus, certain MSAs are not systematically "high" utilization and "low" utilization areas for all diseases, so high and low spending diseases tend to cancel out in aggregate measures. For example, Gary, IN, is ranked as one of the highest utilization cities based on the aggregated SUI, but Gary ranks below the average based on utilization for the disease category "Severity 1 Mood Disorder, Depressed." Prior research suggests that much of the variation in utilization across medical-care markets may be attributed to variation in practice styles and how information disseminates among physicians. For example, Wennberg (1984) reports huge variation in the probability of having tonsils removed across geographic markets.[36] If factors influencing practice patterns are unique for each disease within an MSA, then averaging over the diseases may smooth the variation in utilization in the aggregate indexes.

In contrast to utilization, which may be influenced by idiosyncratic regional practice patterns, MSA fixed effects explain a larger share of the difference in $SPI_{d,r}$, suggesting that there may be more systematic factors affecting service prices across diseases. In particular,

[35] Although one may be concerned that this result may be driven by a small number of episodes at the disease level, similarly large variation is observed when restricted to those diseases with more than 50,000 episodes, shown at the bottom of Table 4. In all cases, the coefficient of variation in spending at the disease level is greater than the aggregate measures, especially for utilization where the COV remains more than double the overall SUI. Another concern is that the analysis may be affected by outliers or small samples; we check for both of these. Specifically, we obtain similar results if we remove outliers for each disease. We also obtain a similarly larger COV at the disease level relative to the aggregate if we define the disease at the level of the Major Practice Category, which aggregates over many ETG disease categories or if we examine only the most frequently observed diseases. Although the COV shrinks when we look at these alternative disease categories, the variation we observe at this level remains considerably larger than the aggregate SUI.

[36] We cite the Wennberg (1984) study because it is a well-known example of a case where information is specific to the treatment of a certain kind of disease. However, the type of variation analyzed in Wennberg (1984) is distinct from our estimates, since it is a population-based measure. In fact, physician diagnostic patterns may also influence utilization patterns (e.g., in the extreme case, it may be that conditional on diagnosis, the rate of tonsil removal does not vary). As we will discuss later and in more detail in an appendix, this topic has been a concern in the literature.

common factors such as the cost of living in an area, may be correlated with the cost of providing services and affect physician and hospital negotiated prices in a systematic fashion. To see this, we examine the correlation between our SPI_r and the regional price indexes for all goods and services from Aten and D'Souza (2008). As expected, we find a positive correlation coefficient of 0.47 between the log of SPI_r and the log of the Aten and D'Souza regional price index that is significant at the 1 percent level. In contrast, we find no significant correlation between the log of SUI_r and the log of the regional price index.[37]

In addition to looking at the variation in the MCE and its components, we also briefly examine the correlation between utilization and price. Simple correlations between the $\log(SUI_r)$ and $\log(SPI_r)$ or the $\log(SUI_{d,r})$ and $\log(SPI_{d,r})$ confirm a negative and statistically significant correlation between utilization and price measures. Specifically, the correlation between the log of the SUI_r and SPI_r is -0.29 which is significant at the 1 percent level. In addition, the correlation between the log of the $SUI_{d,r}$ and the $SPI_{d,r}$ is -0.11 and also statistically significant at the 1 percent level. Therefore, it appears that the greater expenditures from higher prices are partly offset by lower utilization. The relationship between the service price index and the service utilization index is analyzed in greater detail in Section 4.2 of the Online Appendix.

Alternative indexes: It is well known that applying alternative methodologies to construct price indexes will affect the estimated prices. In this paper we focus on a Laspeyres methodology to construct the SPI and SUI measures. However, in the Online Appendix we present three alternative index measures, repeating Tables 4 and 6 for each of alternative index measure. Two of these approaches offer alternative ways for decomposing the ETG disease episodes: (1) a Paasche index; (2) a non-index based approach that provides an exact decomposition. The third approach checks the impact of applying the ETG grouping methodology. Specifically, instead of applying the ETG, we apply an alternative commercial claims grouper, the Medical Episode Grouper (MEG) from Truven Health. Again, the results are quite similar to those found using the ETG grouper, and these results are reported in the Online Appendix. Although there are some differences in the ranking for the MSA level indexes, they are highly correlated with those indexes reported in Table 6, with correlations above 0.90 for the MCE_r, SUI_r, and SPI_r.

[37] The correlation coefficient is -0.191 and is not significant at the 10 percent level.

5.3 MSA-Service-Type Indexes

Table 6 implies differences in the source of expenditures across MSAs. This subsection examines the variation in expenditures for the underlying service types. To do so, we create MSA-service-type indexes which are meant to capture variation in spending, prices and utilization across MSAs for different service-type categories (e.g., physician office or inpatient hospital). These indexes are constructed in a similar manner to the aggregated MSA indexes, except we focus on a single service category (that is, ignoring all other categories) within an MSA. Here we average over diseases within a certain service-type category for a particular MSA and create the service indexes, $MCE_{r,s}$, $SPI_{r,s}$, and $SUI_{r,s}$.[38] Table 7 shows the variation in the indexes for each of the main service types. Overall, it appears that outpatient hospital and office-general MD spending vary most, with pharmacy spending varying the least. In addition, the service price variation for prescription drugs varies the least with a COV of 0.07 with inpatient and outpatient hospital service prices varying the most with a COV of around 0.20. One potential reason for the lower variance in price levels for pharmaceutical products is that competition among prescription drugs is likely to be very similar across markets, since the same drugs are typically available in each market.[39] In contrast, the hospital and physician providers are offering services that are unique to each local market.

Table 7. Coefficient of Variation of Service Indexes Across Service Types

Service Category	COV $MCE_{r,s}$	COV $SPI_{r,s}$	COV $SUI_{r,s}$
Inpatient Hospital	0.203	0.196	0.090
Outpatient Hospital	0.247	0.205	0.335
Office - General MD	0.246	0.118	0.216
Office MD - Speciality	0.185	0.116	0.188
Other (Emergency, Ambulatory Centers, etc)	0.211	0.157	0.162
Pharmacy	0.077	0.068	0.084
Weighted Average	0.188	0.146	0.181

[38]For instance, to construct $SPI_{r,s}$ the price of each service type s for treating disease d, $p_{d,r,s}$, is weighted by the expenditure share of that service type across diseases. For example, let the inpatient hospital expenditure share for disease d be denoted $\theta_{d,Inpatient}$ where $\sum \theta_{d,Inpatient} = 1$. Then the price index for the service category would be: $SPI_{r,s} = \sum_d p_{d,r,s} \cdot \theta_{d,Inpatient}$. In contrast to the overall index that is weighted by the total expenditure share for each disease, this index is weighted by the expenditure share of a service. To normalize the prices we divide by the average price index for that service type across all MSAs.

[39]Although this explanation is likely, one should also recall that the drug prices do not reflect rebates, which could produce mismeasurement problems for drug price variation.

The utilization variation appears to be relatively large for many of the services compared to the aggregate variation observed across markets (i.e., variation in SUI_r) reported in Table 6. A likely reason for the larger variation across service categories is that substitution across service categories increases variation. For example, suppose treatment is shifted from outpatient hospital visits to physician specialty offices. If the quantity of services is the same for both service categories, it would lead to no change in the SUI_r; however, it would increase the variation in the $SUI_{r,s}$. The low COV on the $SUI_{r,s}$ for inpatient and pharmacy services may suggest that there are few substitutes for these categories. A more detailed listing of these specific indexes at the service-type level is shown in the Online Appendix in Tables A7.1, A7.2 and A7.3.

6 A Comparison with a Population-Based Measure

The episode-based measures discussed above provide insight into regional differences in the efficiency of care because they measure spending, price, and utilization only for those individuals being treated. As we discussed earlier, they do not provide much insight about differences in the health status of the population across geographic regions because it ignores the proportion of the population being treated. An alternative measure, which would take into account the health status of the MSA, may be constructed with the population as the denominator, rather than the episode. In this section, we compare the episode-based measure to a population-based measure.

To construct a population-based measure, we let $C_{d,r}$ represent demographically-adjusted expenditure per capita (i.e., per enrollee). This is simply the total expenditure per capita on disease d in MSA r after age and gender weights are applied to each MSA, so that the total age and sex distribution is identical across all MSAs. It follows that the demographically-adjusted expenditure per capita index ($DECI$) is:

$$DECI_{d,r} = \frac{C_{d,r}}{C_{d,B}}. \tag{8}$$

The $DECI_{d,r}$, may be decomposed into two main components and a cross-term:

$$DECI_{d,r} = MCE_{d,r} + PREV_{d,r} - 1 + \frac{(prev_{d,r} - prev_{d,B})(c_{d,r} - c_{d,B})}{prev_{d,B}c_{d,B}}. \tag{9}$$

One of the main components is the episode-based index, $MCE_{d,r}$, and the other component is an index representing the degree of treated prevalence for disease d in MSA r. Specifically, the treated-prevalence index is:

$$PREV_{d,r} = \frac{prev_{d,r}}{prev_{d,B}} \tag{10}$$

where $prev_{d,r} = \frac{N_{d,r}}{population_r}$, is the number of episodes divided by the number of commercially-insured individuals. Note that $prev_{d,r}$ includes only those individuals that are aware of their disease and seek medical attention, and excludes those individuals who are unaware of their disease or are aware and choose not to be treated. Equation (9) makes it clear that a population-based measure of expenditure for a particular disease will rise if there is either an increase in the treated prevalence of the disease or an increase in the expenditure per episode.[40]

Table 8 shows the COV of the $DECI_{d,r}$ for the top 3 diseases[41] and weighted averages of these measures for a large sample of diseases. The table demonstrates how including variation in treated prevalence produces a higher measure of variation. The measures of the average COV for $PREV_{d,r}$ are above 0.20 and COV for $DECI_{d,r}$ is around 0.30. The level of variation in $DECI_{d,r}$ is similar to the magnitude observed in many of the Dartmouth Atlas research studies such as Wennberg (1990), which often focus on the use of specific procedures and do not control for the illness of the patient. In addition, Section 4 of the Online Appendix shows variation statistics from Dartmouth Atlas data, which are similar in magnitude to the $DECI_{d,r}$ variation statistics reported in Table 8 below. Therefore, one may view the episode-based measures that remove variation in treated prevalence (that is, MCE, SPI, and SUI) as more conservative measures of variation.

[40] Note that an aggregate index of expenditure per capita may be constructed in very tractable fashion using the $DECI_{d,r}$. When $DECI_{d,r}$ is weighted by the national expenditure share for each disease, this simply becomes a measure of medical-care expenditures per capita relative to the base region's medical-care expenditures per capita: $DECI_r = \sum_d DECI_{d,r} \cdot (\text{Expenditure Share}_d)$
$= \sum_d \frac{C_{d,r}}{C_{d,B}} \cdot \left(\frac{C_{d,B}}{\sum_d C_{d,B}}\right) = \frac{\sum_d C_{d,r}}{\sum_d C_{d,B}} = \frac{\text{Medical-Care Expenditures Per Person}_r}{\text{Medical-Care Expenditures Per Person}_B}$. The expenditure per capita estimate reported in Table 1 is the numerator of this aggregate index.

[41] Measures of the top 15 diseases are shown in the online appendix

Table 8. Sources of Price Variation Across MSAs by Disease - $MCE_{d,r}$, $SPI_{d,r}$, $SUI_{d,r}$, $PREV_{d,r}$, and $DECI_{d,r}$

	Description	Severity	COV of $MCE_{d,r}$	s.e.	COV of $SPI_{d,r}$	s.e.	COV of $SUI_{d,r}$	s.e.	COV of $PREV_{d,r}$	s.e.	COV of $DECI_{d,r}$	s.e.
1	Pregnancy, with delivery	1	0.18	(0.009)	0.17	(0.004)	0.04	(0.005)	0.17	(0.001)	0.17	(0.009)
	Pregnancy, with delivery	2	0.20	(0.015)	0.19	(0.008)	0.05	(0.006)	0.16	(0.001)	0.18	(0.011)
2	Joint degeneration, localized - back	1	0.18	(0.007)	0.13	(0.004)	0.15	(0.006)	0.18	(0.001)	0.18	(0.006)
	Joint degeneration, localized - back	2	0.28	(0.024)	0.18	(0.017)	0.18	(0.012)	0.22	(0.002)	0.26	(0.018)
	Joint degeneration, localized - back	3	0.29	(0.022)	0.20	(0.014)	0.17	(0.015)	0.31	(0.002)	0.39	(0.021)
3	Ischemic heart disease	1	0.22	(0.010)	0.17	(0.009)	0.17	(0.007)	0.30	(0.001)	0.29	(0.009)
	Ischemic heart disease	2	0.22	(0.016)	0.22	(0.021)	0.15	(0.013)	0.31	(0.002)	0.35	(0.016)
	Ischemic heart disease	3	0.33	(0.055)	0.24	(0.030)	0.17	(0.018)	0.31	(0.003)	0.33	(0.033)
	Weighted Average (Full Sample - 10,000 Episodes in the Data)		0.223		0.161		0.167		0.246		0.304	
	Weighted Average (Only Diseases with 50,000 Episodes in the Data)		0.178		0.131		0.140		0.234		0.270	

Notes. Standard errors are calculated using a bootstrap with 200 random draws of the sample with replacement.

Whether a researcher controls for treated prevalence or not will greatly depend on the focus of the study. There are a number of trade-offs that should be considered. For instance, research by Song et al. (2010) suggests that the probability that providers assign a diagnosis may be endogenous. Specifically, they find that Medicare patients that move to high-utilization areas tend to receive a greater number of diagnoses. See Section 4 of the Online Appendix for a more complete discussion of population-based estimates.

7 Conclusion

Focusing on variation on a disease-by-disease basis, we find the coefficient of variation for the typical disease to be around 0.22. Similarly large measures of variation are observed for both service utilization and service price measures, with a coefficient of variation of around 0.17 for each. The actual measure of variation depends greatly on the disease, with most of the variation attributable to service prices for some diseases and utilization for others. In addition, the variation in service utilization appears to be disease-specific, with MSA-specific factors explaining very little of the across market differences in disease-specific measures. Interestingly, unlike in Medicare markets, it appears that service prices are an important contributor to expenditure differences across commercial health insurance markets.

More generally, this paper presents a framework for exploring the different components of expenditure variation across markets that may be applied to many different research questions. For example, it is possible that greater utilization or service prices may be indicative of higher quality in commercial markets. Additional work is necessary to document how specific spending patterns are related to quality and productivity in the health sector. Another possible avenue for future research is to examine whether the negative relationship between the SUI and the SPI mentioned in this study is spurious or whether it signifies a demand relationship (see Dunn (2013)). Third, the approach applied here to analyze commercial markets may also be applied to the Medicare markets. Applying this similar methodology to Medicare may help facilitate comparison of these two markets and may provide a better understanding of spending differences across markets. Finally, this paper describes differences in the MCE, SPI, SUI, and PREV across markets, but does not attempt to explain underlying reasons for these differences. More work should be done to understand factors that affect these basic components of medical care spending, both across markets and over time.

References

[1] Aten, Bettina and Roger D'Souza, (2008), "Research Spotlight: Regional Price Parities Comparing Price Level Differences Across Geographic Areas", *Survey of Current Business*, 88(11) pgs 64-74.

[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] Anderson, Gerard and Peter Sotir Hussey, (2001), "Comparing Health System Performance in OECD Countries", *Health Affairs*, 20(3) pgs 219-232.

[4] Baicker, Katherine, and Amitabh Chandra, (2004), "Medicare Spending, the Physician Workforce, and the Quality of Health Care Received by Medicare Beneficiaries", *Health Affairs*, Web Exclusive, April pgs 184-197.

[5] Berndt, Ernst R., David M. Cutler, Richard G. Frank, Joseph P. Newhouse,and Jack E. Triplett, (2000), "Medical Care Prices and Output", in: A. Culyer and J. P. Newhouse (Eds.), Handbook of Health Economics, Elservier, Amsterdam. pgs 119-180.

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

[7] Bradley, Ralph, (2013), "Feasible Methods to Estimate Disease Based Price Indexes", *Journal of Health Economics*, 32 pgs 1294-1304.

[8] Bundorf, Kate, Anne Royalty, and Laurence Baker, (2009), "Health Care Cost Growth Among the Privately Insured", *Health Affairs*, 28(5) pgs 1294-1304.

[9] Chandra, Amitabh and Douglas Staiger, (2007), "Testing a Roy Model with Productivity Spillovers: Evidence from the Treatment of Heart Attacks", *Journal of Political Economy*, 115(1) pgs 103-140.

[10] Congressional Budget Office (CBO), (2008), "Geographic Variation in Health Care Spending", Congress of the United States, February.

[11] Chernew, Michael, Lindsay Sabik, Amitabh Chandra, Teresa Gibson, and Joseph Newhouse, (2010), "Geographic Correlation Between Large-Firm Commercial Spending and Medicare Spending", *American Journal of Managed Care*, 16(2) pgs 131-138.

[12] 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.

[13] Cutler, David, and Louise Sheiner, (1999), "The Geography of Medicare", *American Economic Review: Papers and Proceedings*, 89(2) pgs 228-233.

[14] Dunn, Abe, Eli Liebman, Sarah Pack, and Adam Shapiro, (2012), "Medical Care Price Indexes for Patients with Employer-Provided Insurance: Nationally-Representative Estimates from MarketScan Data", *Health Services Research*, 48(3) pgs 1173–1190.

[15] Dunn, Abe, (2013), "Health Insurance and the Demand for Medical Care: Instrumental Variable Estimates using Health Insurer Claims Data", Bureau of Economic Analysis, *Working Paper.*

[16] Dunn, Abe, Eli Liebman, and Adam Shapiro, (2012), "Implications of Utilization Shifts on Medical-Care Price Measurement", Bureau of Economic Analysis, *Working Paper.*

[17] Dunn, Abe and Adam Shapiro, (2011), "Physician Market Power and Medical-Care Expenditures", Bureau of Economic Analysis, *Working Paper*.

[18] Fisher, Elliot, David Wennberg, Therese Stukel, Daniel Gottlieb, F. Lucas, Etoile Pinder, (2003), "The Implications of Regional Variation in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care", *Annals of Internal Medicine*, 138(4) pgs 273-287.

[19] Fisher, Elliot, David Wennberg, Therese Stukel, Daniel Gottlieb, F. Lucas, Etoile Pinder, (2003), "The Implications of Regional Variation in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care", *Annals of Internal Medicine*, 138(4) pgs 288-298.

[20] Fuchs, Victor, Mark McClellan, and Jonathan Skinner, (2004), "Area Differences in Utilization of Medical Care and Mortality Among U.S. Elderly", Perspectives on the Economics of Aging, University of Chicago Press, Chapter 10.

[21] Gage, Barbara, Marilyn Moon, and Sang Chi, (1999), "State-level Variation in Medicare Spending", *Health Care Financing Review*, 21(2) pgs 85-98.

[22] Gowrisankaran, Gautam, Aviv Nevo, and Robert Town, (2013), "Mergers When Prices are Negotiated: Evidence from the Hospital Industry", *NBER Working Paper No. 18875*.

[23] Gottlieb, Daniel, Weiping Zhou, Yunjie Song, Kathryn Gilman Andrews, Jonathan S. Skinner, and Jason M. Sutherland, (2010), "Prices Don't Drive Regional Medicare Spending Variations", *Health Affairs*, 29(3) pgs 537–543.

[24] Medicare Payment Advisory Commission (MedPac), (2003), "Report to the Congress: Variation and Innovation in Medicare, Chapter 1".

[25] National Research Council of the National Academies (U.S.) Panel to Advance a Research Program on the Design of National Health Accounts, (2010), "Accounting for Health and Health Care: Approaches to Measuring the Sources and Costs of Their Improvement". Washington (DC): National Academies Press (U.S.). 2, Medical Care Accounts and Health Accounts: Structure and Data. Available from: http://www.ncbi.nlm.nih.gov/books/NBK53339/.

[26] Roehrig, Charles and David Rousseau, (2011), "The Growth in Cost Per Case Explains Far More of U.S. Health Spending Increases Than Rising Disease Prevalence", *Health Affairs*, 30(9) pgs 1657-1663.

[27] Rosen Allison, Ana Aizcorbe, Alexander Ryu, Nicole Nesoriak, David Cutler, and Michael Chernew, (2013), "Policy Makers Will Need a Way to Update Bundled Payments That Reflects Highly Skewed Spending Growth of Various Care Episodes", Health Affairs. 32(5) pgs 944-951.

[28] Shapiro, Irving L., Matthew D. Shapiro, and David Wilcox, (2001), "Measuring the Value of Cataract Surgery", in: D. M. Cutler and E. R. Berndt (Eds.), Medical Output and Productivity, University of Chicago Press Chicago. pgs 411-438.

[29] Sheiner, Louise, (2012), "Why the Geographic Variation in Health Care Spending Can't Tell Us Much about the Efficiency or Quality of our Health Care System", Federal Reserve Board of Governors, Working Paper.

[30] Song, Yunjie, Jonathan Skinner, Julie Bynum, Jason Sutherland, John E. Wennberg, and Elliott S. Fisher, (2010), "Regional Variations in Diagnostic Practices", *New England Journal of Medicine*, 363 pgs 45-53.

[31] Skinner, Jonathan, (2012), "Causes and Consequences of Regional Variations in Health Care", *Handbook of Health Economics*, pgs 45-93.

[32] Wennberg, John, (1984), "Dealing with Medical Practice Variations: A Proposal for Action", *Health Affairs*, 3, pgs 6-32.

[33] Wennberg, John, (1990), "Small Area Analysis and the Medical Care Outcome Problem", in L. Sechrest, E. Perrin, and J. Bunder, eds. *Research Methodology: Strengthening Causal Interpretation of Non-Experimental Data*, Rockville, MD: Department of Health and Human Services, PHS90-3454 pgs 177-213.

[34] Zhang, Yuting, Katherine Baicker, and Joseph Newhouse, (2010), "Geographic Variation in Medicare Drug Spending", *New England Journal of Medicine*, 363 pgs 405-409.

[35] Zuckerman, Stephen, Timothy Waidmann, Robert Berenson, and Jack Hadley, (2010), "Clarifying Sources of Geographic Differences in Medicare Spending", *New England Journal of Medicine*, 363 pgs 54-62.