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

Abe Dunn, Adam Hale Shapiro, and Eli Liebman

January 17, 2012

Abstract

This study introduces a new framework for measuring and analyzing medical-care expenditures applied to the study of commercial medical-care markets. The framework focuses on expenditures at the disease level that are decomposed between price and utilization. These measures show that a particular MSA may have high overall prices, but may actually have low medical-care spending per episode due to low utilization. Prices within an MSA appear to be quite homogeneous, implying that regional factors explain a large degree of price variation. However, within an MSA there is a large degree of heterogeneity in utilization patterns between disease categories. This implies that most MSAs do not have systematically high or low utilization for *all* disease categories. We find evidence of a negative correlation between price and utilization across MSAs for many diseases, so it appears that the greater expenditures from higher prices are partly offset by lower utilization.

*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

There has been a considerable amount of research assessing the geographic variation of Medicare expenditures, but relatively little is known about geographic variation in commercial-market expenditures. This is a large hole in our understanding of the overall healthcare market. The private market includes around 174 million enrollees, compared to Medicare which has 44 million. In addition, medical-care expenditures from private insurers account for 60 percent more spending than Medicare.[1] As private markets are influenced by different economic forces than regulated markets, there is likely distinct variation in medical-care expenditures. Unlike the Medicare markets where payments to providers are fixed by the Center for Medicare and Medicaid Services, prices in the private sector are set through negotiations between insurers and providers. Understanding how medical-care expenditures differ across these two markets may ultimately provide important insights into how government regulation affects medical-care spending and outcomes. A sound identification of geographic variation in total medical expenditures may be one step in comprehending how to achieve lower healthcare costs. Indeed, there is most likely waste in medical-care expenditures in the United States, since the country spends 50 percent more than most other developed nations on healthcare in terms of a fraction of its GDP, but does not necessarily produce measurably better health outcomes (See Garber and Skinner (2008) and Anderson and Hussey (2001)).

In this study, we provide a new framework for researchers and policy makers to analyze expenditures in the commercial medical-care market. Similar to Aizcorbe and Nestoriak (2011), Aizcorbe et al. (2010), and Dunn et al. (2010), we construct a Medical-Care Expenditure Index (MCE) that tracks the overall medical-care expenditure of treating an episode of a disease. Our study differs from this previous work on 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

[1] Specifically, private market accounts for more than 35 percent of total medical-care expenditures, while Medicare's share is 22 percent. The other sources of funding include Medicaid accounting for 17 percent, out-of-pocket costs accounting for 14 percent, and other sources accounting for the remaining 14 percent of spending. These figures are from the personal health expenditures reported in the National Health Expenditure Accounts for the year 2009. Also note that around 57 percent of the out-of-pocket costs are from individuals with private insurance (MEPS 2007 Data).

The 174 million in the private sector includes only those under 65 years of age (Health United States (2009)).

variation in underlying service prices (for example, the price of a visit to a doctor to manage a pregnancy), but holding service utilization constant (for example, fixing the number of 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. Our methodology for decomposing service prices and service utilization starts from the most granular level (that is, the particular procedure provided by the physician treating a certain disease) to more precisely capture price and utilization differences. We find that these three measures produce vastly different pictures of healthcare variation across the country. For example, of the 85 MSAs analyzed in this study, Wichita ranks 6th in terms of its SPI. However, in terms of its MCE it ranks 59th. The relatively low level of medical-care spending per episode in Wichita is because it has relatively low utilization of services per episode—its SUI is ranked 72nd.

The framework presented here may be valuable for future researchers trying to understand differences in medical-care expenditures across geographic markets. For instance, these measures are important for understanding expenditure differences and potential savings across markets. If significant variation in utilization is identified then market participants can seek out ways to control potentially wasteful spending. On the other hand, if variation in spending is primarily driven by price differences, then measures to curb utilization may have a limited impact on high-pending areas, and policy makers should focus on factors that affect the price of services, such as the bargaining power of providers and factors affecting the cost of the underlying services (for example, regulatory barriers or wasteful administrative costs).

Using our medical-care expenditure decomposition, we present a descriptive picture of how disease spending and its components, service prices and service utilization, vary across markets. It is important to emphasize that these measures are meant to provide a measure of the utilization and price of services for *similar* patients across regions. For this reason, our indexes are constructed at the disease level, which categorizes patients based on the particular disease and severity of illness. This is in contrast to a population-based index which ignores the patient's disease and therefore includes variation in the incidence of diseases across regions. We find that categorizing patients according to the disease (including severity of illness) produces different measures of regional spending patterns than a simple population-based measure.

Looking at the variation observed at the disease level, we find a large degree of variation across the 85 MSAs that we study in this paper—the weighted average coefficients of variation for MCE, SPI, and SUI are 0.22, 0.16 and 0.19, respectively. The relative magnitude of service price or service utilization variation depends on the disease. For example, it appears that for some diseases the variation in service price is relatively large (for example, pregnancy), while for other diseases the variation in utilization appears large (for example, depression). Interestingly, regional factors explain only 15 percent of the observed variation in utilization but 39 percent of the variation in prices for the different diseases.

Examining the aggregate measures, where we average over the disease-specific indexes, we find relatively limited variation in medical-care expenditure across geographic markets (that is, MSAs). The coefficient of variation for the aggregate MCE index is 0.10 and appears to be similar in magnitude to spending variation in non-health goods and services in the economy (for example, spending on food has a coefficient of variation of 0.12). Thus, it appears that averaging over diseases masks the underlying geographic variation in spending across specific diseases, especially the utilization component. This finding leads us to believe that there are not necessarily high- and low-utilization areas for all disease categories. Rather, certain geographic areas may be over-utilizing services for some diseases and under-utilizing for others. Therefore, there may be greater efficiency gains by comparing specific disease categories rather than focusing on the aggregate level.

Unlike Medicare markets where variation is primarily driven by differences in utilization across markets, it appears that variation in price is particularly important in commercial markets. The variation in service prices suggests that even if utilization is controlled for, differences in spending are likely to persist across areas. We find that most of the price variation is from hospital services followed by variation in prices for physician services and then pharmacy prices. A likely explanation for limited variation in pharmacy prices is that the competition in pharmacy markets is very similar across the nation, since the same drugs are generally available across all markets; while the competitive environment and the products themselves may vary greatly across markets for physician and hospital services.

To assess the economic importance of geographic variation in medical-care expenditures, we calculate changes in the episode expenditures from conducting hypothetical shifts in service utilization and service prices at the disease level. First, we find that if disease expenditures in the high-spending areas (that is, the top quartile) had similar utilization to those in low-spending areas (that is, the bottom quartile), there would be a 23 percent reduction in per episode expenditures. Similarly large differences are shown by shifting the service prices from the low-spending areas to high-spending areas, where we find a reduction in the average episode expenditures of 16 percent. The potential savings based on observed variation in service prices and service utilization is sizable, but these simple exercises do not capture equilibrium responses to changes in price and utilization. In fact, we find evidence that low utilization areas tend to be higher priced areas, so that the benefits of lower utilization are partly offset by higher prices.[2]

An additional check on the efficiency of medical care spending is to assess how well these three measures (that is MCE, SPI, and SUI) are correlated with quality of treatment.[3] Although evidence from Medicare data suggests that more spending is not necessarily associated with a greater quality of service, less research has studied the relationship between spending and quality in the commercial sector. Looking at disease-specific quality measures and expenditures, we observe no clear pattern— some relationships are significant and positive, some are significant and negative, and many are insignificant. However, we find some positive and significant correlations between aggregate MSA-level SUI and SPI measures and composite measures of quality that combine multiple quality measures. The strongest positive relationship appears to be between aggregate measures of utilization and preventative care. The underlying causal link between these correlations is left for future research.

2 Literature Review

Geographic variation in healthcare utilization and expenditure is a growing area of study. The work by Dartmouth researchers has shown considerable variation in medical-care expenditures and service utilization across areas of the United States. Results from studies using Medicare data show large geographic variation in spending and utilization that does not appear to be associated with better patient outcomes or quality. For example, for a national sample of Medicare enrollees, Baicker and Chandra (2004) find that spending and quality of care are actually inversely related.

While there is a large body of work studying variation across geographic markets for samples of Medicare beneficiaries, fewer studies have analyzed variation in medical-care

[2] This negative relationship between service prices and service utilization is also confirmed by examining correlations between the SPI and SUI indexes, which show a negative and statistically significant correlation (both in the aggregate and at the disease level). Regression estimates that control for MSA fixed effects and apply instrumental variable techniques confirm this negative relationship.

[3] We use guidelines from the National Committee for Quality Assurance (NCQA) to construct quality measures in our study.

expenditures in the commercial sector.[4] Although it may be tempting to draw conclusions from the significant analysis conducted in the Medicare market, evidence suggests that commercial and Medicare markets may be quite distinct, even within the same geographic area. In particular, Chernew et al. (2010) compare spending across Medicare and commercial markets and found only a small negative correlation. Although their study finds a significant positive correlation in utilization of 0.59 for inpatient days per capita, the link between the two markets remains unclear. Moreover, given the regulated structure of Medicare pricing, it appears that price variation seems to be relatively unimportant in Medicare markets (See Gottlieb et al. (2010)). However, this may not be the case in commercial markets where insurers and providers are free to negotiate over rates. Therefore, additional research may be necessary to understand the unique features of spending, pricing and utilization in commercial markets.

There are many approaches for analyzing geographic differences in spending and utilization across markets. Some research focuses on differences at the micro level, examining the use of specific procedures for certain diseases (for example, Chandra and Staiger (2007) look at different types of treatments for heart attack patients across markets), while other studies examine aggregate differences in overall medical-care expenditures (for example, Cutler and Sheiner (1999), Fuchs, McClellan and Skinner (2004), and MedPac (2003)). This paper combines aspects of both these approaches because it focuses on aggregate medical-care expenditures in a geographic area, but these measures are constructed at the most micro level possible. This approach may be useful for understanding aggregate differences in medical-care expenditures, since providers differ by specialty (for example, cardiology or orthopedic doctors), type (for example, hospitals or physicians), and factors unique to a local market (for example, cultural or information differences (Chandra and Staiger (2007))).

Previous studies also differ in the unit at which expenditures are measured. Studies by Cutler and Sheiner (1999) and Gage, Moon and Chi (1999) track expenditures on a per capita basis, while studies by Aizcorbe and Nestoriak (2011), Aizcorbe et al. (2010), and Dunn et al. (2010) track expenditures on a per episode-of-care basis. The two types of measures provide different information about medical-care spending. The decision to assess one or the other depends on the policy question as well as the goal of the researchers. For

[4] The seminal work in the study of geographic differences in medical-care expenditures was pioneered by John Wennberg, with a description of his earlier work and implications of geographic variation discussed in Wennberg (1984). A more recent review of the literature on geographic variation in health care spending is in Congressional Budget Office (CBO) (2008).

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 provider. Table 1, which lists the MSAs with the five highest and lowest medical-care spending per person in our dataset, shows that there can indeed be considerable differences between a per-capita and per-episode type measure. 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. While there are a multitude of potential explanations for this discrepancy, a plausible reason is that there is likely large variation in the incidence of diseases across MSAs. It is quite possible that Birmingham's population-based measure is large relative to the episode-based measure because the MSA has a relatively large number of people with expensive diseases. Overall, for both types of measures, differences in the incidence of disease likely generate a large amount of variation in spending across MSAs.

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

MSA	Rank	Medical-Care Expenditures Per Person	Rank	Medical-Care Expenditures Per Episode
Milwaukee-Waukesha-West Allis, WI	1	$2,915	1	$1,032
Salinas, CA	2	$2,882	4	$974
Fort Worth-Arlington, TX	3	$2,854	9	$940
Birmingham-Hoover, AL	4	$2,851	17	$870
Oakland-Fremont-Hayward, CA	5	$2,809	6	$956
Memphis, TN-MS-AR	81	$2,154	71	$741
Riverside-San Bernardino-Ontario, CA	82	$2,122	30	$825
MSA in Mississippi	83	$2,036	27	$831
Las Vegas-Paradise, NV	84	$2,022	44	$791
Pittsburgh, PA	85	$1,997	75	$734

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.

Following the recommendations of Berndt et al. (2001) we measure the medical-care expenditure of an episode of care at the disease level. Measuring expenditures at the level of the episode and the disease provides a measure of the cost of treatment for similar types of individuals and is therefore constructed to wipe out factors that affect the health of the population in a given geographic area. Tracking expenditures at this level may be particularly important given the potential for shifts in service between provider types. In fact, there are several documented shifts in treatment over time that have been studied in the literature, such as heart attacks (Cutler et al. (1998)), cataracts (Shapiro et al. (2001)), and depression (Berndt et al. (2001)). More broadly, Aizcorbe and Nestoriak (2011) track the"price" of treating a disease, or what we call the expenditure for an episode of care, over time for a broad range of conditions. 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). They document several important shifts in utilization across provider types that drive a wedge between the two indexes; finding that the SPI generally grows faster than the MCE, suggesting that medical care inflation based on an SPI measure may be overstated.[5] The existence of these observed shifts over time suggest that different allocations of services across geographic markets may also be important. Indeed, in our paper we show 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 shifts in utilization are also important in determining the level of expenditures across geographic markets.

Aizcorbe and Nestoriak (2011) provide an innovative approach for studying medical-care expenditures and differences in expenditures across a broad range of conditions. In this study, we adopt some of their basic methodology. However, we introduce two key methodological innovations. First, we decompose medical-care expenditures per episode into a service-price as well as a service-utilization component. This additional decomposition allows us to look at differences in medical-care expenditures caused by differences in the quantity of services provided in addition to differences in service prices. Second, we analyze services at a more micro level—the level of the specific procedure. This allows us to capture greater heterogeneity in the types of services performed across markets. For example, rather than pricing a visit to a doctor, we focus on the price for a particular procedure and modifier code.[6] For completeness, we also present an additional decomposition that demonstrates how the episode-based expenditure measure, which is the central focus of our analysis, relates to a population-based expenditure measure that looks at expenditures per person. This methodology is discussed in greater detail in the following sections.

[5] This finding is supported by Aizcorbe et al. (2010) and Dunn et al. (2010).

[6] Another difference is that we focus on completed episodes, rather than the cost of the disease per year. For example, we look at the expenditures on a completed episode of a pregnancy, rather than the amount spent on pregancy for a calendar year. Although this is an important distinction, we find that these two ways of analyzing disease expenditures are quite similar in practice when episodes are aggregated over a large population.

3 Methodology of Index Construction

The MCE index is a measure of the medical-care expenditures for the treatment of an episode of care for a certain disease, and is defined as the dollar amount of medical care used until treatment is completed.[7] Formally, we denote the average expenditure per episode of treating disease d in area r as $c_{d,r}$. Denoting $c_{d,B}$ as the average expenditure per episode across *all* areas, the MCE index for disease d is the ratio of the two measures:

$$MCE_{d,r} = \frac{c_{d,r}}{c_{d,B}} \tag{1}$$

Thus, 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.

Our decomposition rests on the fact that the average expenditure, $c_{d,r}$, can be divided between 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 condition and 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.

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.[8] 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). Therefore, we would like service utilization to reflect the amount of procedures done. Since not all procedures are equivalent, we weight each procedure by the average dollar amount paid for that procedure. This is a similar concept to a "Relative Value Unit" or "RVU", which measures the approximate cost of each procedure and is used by Medicare to reimburse physicians for each procedure that is performed.[9] For prescription drugs, we define the unit of service as a prescription filled,

[7] For example, for an individual with a broken foot, the episode of treatment will be defined by the dollar of medical services used to treat that condition from the first visit to a provider until the foot is healed. For medical conditions that are chronic, we interpret an episode as expenditure for services used to treat the chronic condition over a one year period.

[8] The key service types are inpatient hospital, outpatient hospital, general physician, physician specialist, and prescription drugs.

[9] This framework has also been adopted by the commercial market. In a survey of 20 health plans

albeit this is a bit of a misnomer since a prescription is really a "good," not a service. Since prescriptions vary depending on the active ingredient, the manufacturer, and strength, we weight each unique drug purchase by the average dollar amount we observe for that particular prescription across geographic areas. For hospital facility charges for inpatient stays, the prices paid to facilities are often set based on the disease and the number of nights in the facility. Therefore, for inpatient stays we define the unit of service as a night-of-stay. For outpatient facility services we define the service as the visit itself. The exact construction of these measures is explained in more detail later in this paper.

Given the definition of service and expenditure, the price for a particular service type and disease can be calculated by dividing its expenditure by the 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 expenditure on disease d for service type s in area r. For example, the price of an inpatient stay for treating heart disease is the total expenditure of inpatient treatment for heart disease in an area, divided by the quantity of inpatient services for heart disease in that area.

This decomposition allows us to create a service price and service utilization index. To simplify, let $q_{d,r}$ be a vector of services utilized for the typical treatment of diseases in an area, $q_{d,r} = Q_{d,r}/N_{d,r}$, where the component of the utilization vector for service type s is , $Q_{d,r,s}/N_{d,r}$. Also, let $p_{d,r}$ be a vector of service prices, where the component of the vector for service type s is, $p_{d,r,s}$. The service price index (SPI) is then calculated as:

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

which holds the utilization of services fixed at a base period level. Similarly, the service utilization index (SUI) may be defined as:

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

which holds the price of services fixed while allowing the utilization of services to vary. Note that there is a precise relationship between these three indexes that is described by the following decomposition:

$$MCE_{d,r} = SPI_{d,r} + SUI_{d,r} + (p_{d,B} \cdot q_{d,r} - c_{d,B})(p_{d,r} \cdot q_{d,B} - c_{d,B})/((c_{d,B})^2) - 1$$

conducted by Dyckman & Associates, all 20 health plan fee schedules were influenced by a resource-based relative value scale (RBRVS). There are deviations from the basic RBRVS methodology, so taking the average of observed prices in the market for each procedure is one measure used for capturing the typical "resources" used for a procedure.

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, $(p_{d,B} \cdot q_{d,r} - c_{d,B})(p_{d,r} \cdot q_{d,B} - c_{d,B})/((c_{d,B})^2)$, and subtracting 1. The cross term accounts for joint changes in both price vectors and utilization vectors and, in practice, the term is near zero. In the case where there are very few differences in utilization across markets, $SUI_{d,r}$ is fixed near 1, then the $MCE_{d,r}$ will entirely be determined by service prices. Similarly, if there are very few differences in service prices across markets, $SPI_{d,r}$, is near 1, and the $MCE_{d,r}$ will entirely be determined by utilization.

4 Data

We use retrospective claims data for a sample of commercially-insured patients from the MarketScan® Research Database from Thomson Reuters. The specific claims data used is the "Commercial Claims and Encounters Database" which contains data from the employer and health plan sources concerning medical and drug data for several million commercially-insured individuals, including employees, their spouses, and dependents. Each observation in the data corresponds to a line item in an "explanation of benefits" 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.[10] 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.[11]

The claims data has been processed using the Symmetry grouper 7.6 from Ingenix. The grouper assigns each claim to a particular Episode Treatment Group (ETG) disease category.[12] 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 a distinct condition; the grouper also uses other information on the claim (for example, 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

[10] In our selected sample that removes individuals without drug benefit information, we find about 62 percent of the data is from employer plans while the remaining 38 percent is from health insurance plans.

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

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

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 episodes, 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).[13]

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 condition of diabetes will be given a severity level of 4 while the least severe diabetes condition 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.[14] Total episode expenditures are measured as the total dollar amount received by all providers for the services used to treat an episode of a specific disease (including both out-of-pocket payments and amounts paid by insurance firms).

Service utilization measures were 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. Using the definitions of the unit of service for each service type, the price of the service is calculated as the total expenditures for a particular disease and service category, divided by the quantity of services performed for that disease and service category. Furthermore,

[13] 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.

[14] 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.

service utilization for a particular category is defined as the quantity of services divided by the total number of episodes for a particular disease. Below is a listing of the service types and how the quantity of services is measured.

Physician Office Visits - The physician visits are based on procedures performed in a physician's office. We assign a measure comparable to an RVU for each procedure performed by the physician for that office visit. Specifically, for each CPT and modifier, we calculate a relative value unit by computing the average fee for that procedure performed in an office setting. The total amount of services performed in an office is calculated by summing over these calculated "RVU" units.[15] For instance, if a 15 minute office visit has an average price of $100 across the data, its value will be 100 RVUs.

Hospital Inpatient - Inpatient hospital stays consist of both facility fees paid to the hospital, but also fees paid to the physician. 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 per night multiplied by the number of nights. 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. However, we average over procedure prices in an inpatient setting. The total amount of services performed in an inpatient setting is calculated by adding the physician and facility amounts.[16]

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. To capture these differences, we calculate the average price for each NDC code, since each prescription is given a unique NDC code. The average price for each NDC code represents the amount of the service used. If the expenditure on a prescription is greater than this amount, it suggests that prices are above average in an area.[17]

[15] Although procedure codes are observed for 98 percent of physician office claim lines, in those cases where we don't observe a procedure code we calculate the average price for a missing procedure code for patients with a particular disease. The results of the paper do not change substantially if those claim lines missing procedure codes are dropped from the analysis.

[16] As an alternative, we have also examined changing this definition to consider the facility price per inpatient day. The results do not change significantly based on this alternative measure of utilization.

[17] An 11-digit National Drug Code (NDC) uniquely identifies the manufacturer, the strength, dosage, formulation, package size, and type of package. In this geographic analysis, we treat branded and generic

All Other - The other category primarily includes ambulatory care, independent labs, and emergency room visits. For these services, the amount of each category is measured as the average cost for a visit to that particular place of service, for example, the average cost of an ambulatory care visit to treat ischemic heart disease. For cases where procedure codes are available, we use the average cost of that procedure code for that place of service.

These measures of service quantity subsequently allow us to create service prices that correspond well with how fees are negotiated in the marketplace. In practice it appears that physicians and hospitals often negotiate on a percentage amount from some pre-determined base, such as, 10 percent above Medicare rates.[18] As our measure of service price can be intuited as expenditure divided by a proxy for "RVUs", 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.

There are also a few methodological points that are important to consider. First, a small fraction of the procedures (less than 5% of the claims observations for non-facility claim lines) are missing procedure codes. For these procedures we take the average price of the missing procedure codes for that service and disease type. The results presented here do not change when alternative methods for calculating utilization are used. For instance, we obtain similar results when we drop claim lines that are missing procedure codes. Also, some claim lines have negative billing amounts that represent corrections to bills. These negative amounts are left in the analysis so that the corrected billing amounts are used in our analysis.[19] Another potential concern is that some codes that appear relatively infrequently. To address this concern we use the maximum amount of potential data to construct the quantity of service measure.[20]

products that contain the same active molecule as a distinct drugs.

[18]As mentioned previously, of the 20 plans surveyed by Dyckman & Associates (2003), all of the plans use some variation of the Medicare resource-based relative value scale (RBRVSD) methodology to set prices.

[19]For example, if a patient was billed $100 for an office visit, but the bill should have been $125, the correction to the bill is often made by adding a claim line of -$100 and an additional claim line of $125. These negative amounts will cancel with the positive amounts that are being corrected over the same episode, so the mistaken amounts should not appear in our analysis.

[20]We exclude claims that contain capitated payments, but include all other claims that are available from 2003 through 2008, prior to sample selection. This provides a maximum amount of data to construct a base "amount". The results do not change when the amounts are constructed from a smaller set of data, such as only using the years 2006 and 2007. The most likely reason that this change has no effect, is that most of the expenditures are concentrated on more frequent cpt codes.

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, which is a national survey of health expenditures meant to be representative of the entire U.S. non-institutionalized population.[21] This first selection rule leaves a sample of 85 MSAs.[22] When analyzing diseases within this population, we wish to remove those diseases that appear relatively infrequently. The concern is that it may be challenging to obtain precise estimates of episode expenditures for infrequently observed diseases. We are also concerned that the variation in severity for a particular disease may be large. To account for these issues, we define each disease as an ETG-severity combination, so that each ETG-severity combination will be examined separately. We next select those diseases for which we observe 10,000 episodes in the data for the selected MSAs.[23] Those diseases with 10,000 or more episodes account for 78 percent of overall expenditures and 96 percent of the episodes.

Population weights are applied to each MSA to adjust for differences in age and sex across populations, so the expenditure estimates may be comparable across markets. Specifically, enrollees in each MSA are assigned weights so the weighted population has an age and sex distribution that is identical to that of the US commercially-insured population.[24]

[21] The commercially insured sample in the MEPS data is around 14,799 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.

[22] 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 used 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.

[23] 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 appendix to this paper.

[24] Using the enrollment data in each MSA, weights are applied to different age and sex categories so that the total enrollment files match the population for commercially-insured individuals in the U.S. for 2007. Information on the population is obtained from the Medical Expenditure Panel Survey.

MSA observations in 2006 and 2007 are each weighted to the national level population in 2007. That

Table 3 provides some basic descriptive statistics for the selected population and the overall disease expenditures for the selected diseases. It shows the average disease expenditures for the two-year period of 2006 and 2007 and is based on the weighted sample of enrollees, so these figures are representative of the 2007 U.S. population.[25] Based on the ETG groupings, the top 5 ETG categories based on overall expenditure include 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 $641 per episode, while Severity 3 Ischemic Heart Disease costs $19,592.

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 very similar results in each year.

[25] 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 MSAs 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.

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

	Description	Severity	Total Dollars (Billions)	Episodes	Dollars Per Episode	Fraction of Spending	Fraction of Spending Category
1	Pregnancy, with delivery	1	$13.65	1,483,299	$9,200.09	3.1%	4.7%
	Pregnancy, with delivery	2	$6.94	509,232	$13,623.34	1.6%	
2	Joint degeneration, localized - back	1	$10.90	6,246,567	$1,745.54	2.5%	4.4%
	Joint degeneration, localized - back	2	$4.81	1,156,768	$4,161.57	1.1%	
	Joint degeneration, localized - back	3	$3.69	302,136	$12,205.88	0.8%	
3	Ischemic heart disease	1	$8.48	2,439,447	$3,475.98	1.9%	4.4%
	Ischemic heart disease	2	$6.62	1,223,808	$5,407.47	1.5%	
	Ischemic heart disease	3	$3.90	198,831	$19,592.37	0.9%	
4	Hypertension	1	$11.27	17,572,123	$641.48	2.6%	4.0%
	Hypertension	2	$3.33	3,844,359	$865.42	0.8%	
	Hypertension	3	$1.72	1,589,968	$1,080.60	0.4%	
	Hypertension	4	$1.35	595,232	$2,271.75	0.3%	
5	Diabetes	1	$9.81	6,412,962	$1,529.86	2.2%	3.7%
	Diabetes	2	$1.79	730,306	$2,455.40	0.4%	
	Diabetes	3	$1.79	540,763	$3,307.72	0.4%	
	Diabetes	4	$2.76	478,053	$5,774.17	0.6%	
6	Routine exam	1	$13.47	62,625,868	$215.12	3.1%	3.1%
7	Mood disorder, depressed	1	$8.67	7,199,731	$1,204.64	2.0%	2.7%
	Mood disorder, depressed	2	$1.88	1,134,679	$1,659.46	0.4%	
	Mood disorder, depressed	3	$1.10	351,858	$3,124.52	0.3%	
8	Malignant neoplasm of breast	1	$6.12	819,649	$7,465.18	1.4%	2.4%
	Malignant neoplasm of breast	2	$4.28	225,091	$19,030.50	1.0%	
9	Hyperlipidemia, other	1	$10.24	15,881,523	$644.60	2.3%	2.3%
10	Joint degeneration, localized - neck	1	$6.17	4,111,996	$1,499.47	1.4%	2.2%
	Joint degeneration, localized - neck	2	$0.87	367,287	$2,363.83	0.2%	
	Joint degeneration, localized - neck	3	$2.51	287,407	$8,718.57	0.6%	
11	Chronic sinusitis	1	$5.09	9,974,944	$510.40	1.2%	1.9%
	Chronic sinusitis	2	$1.16	1,318,903	$879.14	0.3%	
	Chronic sinusitis	3	$1.99	866,030	$2,303.05	0.5%	
12	Joint degeneration, localized - knee & low er leg	1	$5.07	2,313,867	$2,190.78	1.2%	1.7%
	Joint degeneration, localized - knee & low er leg	2	$1.15	312,879	$3,663.28	0.3%	
	Joint degeneration, localized - knee & low er leg	3	$1.34	201,427	$6,635.43	0.3%	
13	Asthma	1	$2.46	4,260,258	$576.86	0.6%	1.7%
	Asthma	2	$3.37	3,454,175	$975.67	0.8%	
	Asthma	3	$0.66	345,752	$1,909.96	0.2%	
	Asthma	4	$1.06	293,256	$3,605.37	0.2%	
14	Joint derangement - knee & lower leg	1	$1.20	694,114	$1,725.10	0.3%	1.7%
	Joint derangement - knee & lower leg	2	$6.01	1,165,905	$5,156.78	1.4%	
15	Inflammation of esophagus	1	$4.73	3,573,545	$1,324.61	1.1%	1.6%
	Inflammation of esophagus	2	$1.33	688,744	$1,924.77	0.3%	
	Inflammation of esophagus	3	$0.86	258,916	$3,338.16	0.2%	
	Other		$250.70	371,234,263	$675.31	57.5%	57.5%
	Total		$436.28	539,285,923	$809.00	100.0%	

For the analysis that follows, it is important to remind the reader that the MCE, SUI, and SPI measures are constructed at the MSA-disease level, so the discussion of the variation in the MCE, for instance, is the variation in spending for a particular disease across MSAs. This is also important because it averages over individual episode expenditures which are likely to be idiosyncratic. Thus, this paper focuses on an average expenditures for treating a disease within an MSA, often using hundreds of observed episodes of a disease per MSA.

5 Results

5.1 MSA-Disease Indexes

Table 3 reports the standard deviation 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 coefficient of variation (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 0.19 for the full sample.[26]

The underlying cause for the variation may be attributed to either utilization or price, which are shown using the standard deviation of the $SPI_{d,r}$ and $SUI_{d,r}$. For some conditions it appears that price variation primarily affects the variation across areas, while for other conditions, the utilization variation appears to be more important. For example, the Severity 1 Mood Disorder, Depression has a relatively low price variation, but the utilization variation is greater for all severity levels. This could potentially be explained by talk therapy being more popular in some areas, while treatment with depression drugs may be more common in other areas; although the prices for each of these services may not vary by a large amount across areas. In contrast, a condition 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.

[26] 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 64 percent of the expenditure from the selected sample.

Table 3. Sources of Price Variation Across MSAs by Disease - $MCE_{d,r}$, $SPI_{d,r}$ and $SUI_{d,r}$

	Description	Severity	COV of $MCE_{d,r}$	COV of $SPI_{d,r}$	COV of $SUI_{d,r}$
1	Pregnancy, with delivery	1	0.18	0.16	0.07
	Pregnancy, with delivery	2	0.19	0.19	0.10
2	Joint degeneration, localized - back	1	0.18	0.12	0.14
	Joint degeneration, localized - back	2	0.28	0.16	0.20
	Joint degeneration, localized - back	3	0.29	0.19	0.22
3	Ischemic heart disease	1	0.22	0.17	0.18
	Ischemic heart disease	2	0.22	0.22	0.16
	Ischemic heart disease	3	0.33	0.25	0.21
4	Hypertension	1	0.10	0.09	0.11
	Hypertension	2	0.11	0.11	0.12
	Hypertension	3	0.14	0.11	0.12
	Hypertension	4	0.35	0.26	0.29
5	Diabetes	1	0.12	0.06	0.10
	Diabetes	2	0.21	0.15	0.28
	Diabetes	3	0.21	0.12	0.17
	Diabetes	4	0.22	0.20	0.19
6	Routine exam	1	0.15	0.12	0.06
7	Mood disorder, depressed	1	0.20	0.07	0.17
	Mood disorder, depressed	2	0.26	0.12	0.20
	Mood disorder, depressed	3	0.25	0.17	0.23
8	Malignant neoplasm of breast	1	0.20	0.14	0.16
	Malignant neoplasm of breast	2	0.27	0.18	0.20
9	Hyperlipidemia, other	1	0.10	0.09	0.10
10	Joint degeneration, localized - neck	1	0.19	0.11	0.18
	Joint degeneration, localized - neck	2	0.30	0.17	0.41
	Joint degeneration, localized - neck	3	0.30	0.25	0.24
11	Chronic sinusitis	1	0.17	0.07	0.13
	Chronic sinusitis	2	0.20	0.08	0.17
	Chronic sinusitis	3	0.25	0.17	0.18
12	Joint degeneration, localized - knee & lower leg	1	0.25	0.18	0.18
	Joint degeneration, localized - knee & lower leg	2	0.31	0.21	0.27
	Joint degeneration, localized - knee & lower leg	3	0.29	0.31	0.29
13	Asthma	1	0.11	0.06	0.10
	Asthma	2	0.11	0.07	0.10
	Asthma	3	0.32	0.13	0.29
	Asthma	4	0.25	0.16	0.24
14	Joint derangement - knee & lower leg	1	0.26	0.17	0.20
	Joint derangement - knee & lower leg	2	0.21	0.20	0.17
15	Inflammation of esophagus	1	0.12	0.09	0.09
	Inflammation of esophagus	2	0.14	0.11	0.11
	Inflammation of esophagus	3	0.32	0.15	0.25
	Weighted Average (Full Sample - 10,000 Episodes in the Data)		0.223	0.159	0.185
	Weighted Average (Only Diseases with 50,000 Episodes in the Data)		0.178	0.128	0.148

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 several 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; (3) Include MSA-ETG Category fixed effects.[27]

The results of these regressions are shown in Table 4. MSA fixed effects explain about one fourth of the variation for the MCE. Interestingly, the MSA fixed effects explain a

[27] 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.

relatively larger portion of the variation in the $SPI_{d,r}$ than they explain the variation in the $SUI_{d,r}$. Specifically, the R^2 is 0.15 for the $SUI_{d,r}$, compared to 0.39 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 50 percent. This suggests that there are important disease-specific factors within each market that cause variation in utilization and price.[28] Therefore, within an MSA there is a large degree of heterogeneity in utilization patterns among disease groups.

Table 4. Decomposing the Sources of Service Utilization and Service Price Variation

Log($MCE_{d,r}$)

	R^2	Adj R^2	MSE	N
MSA-FE	0.217	0.2017	0.19243	11305
MSA-MPC-FE	0.4573	0.3758	0.17016	11305
MSA-Disease-FE	0.7178	0.5077	0.16087	11305

Log($SPI_{d,r}$)

	R^2	Adj R^2	MSE	N
MSA-FE	0.3944	0.3826	0.12578	11305
MSA-MPC-FE	0.584	0.5214	0.11075	11305
MSA-Disease-FE	0.7947	0.6451	0.09536	11305

Log($SUI_{d,r}$)

	R^2	Adj R^2	MSE	N
MSA-FE	0.1502	0.1336	0.16739	11305
MSA-MPC-FE	0.3891	0.2974	0.15075	11305
MSA-Disease-FE	0.6752	0.4397	0.13462	11305

Notes. Based on a regressions on log($SUI_{d,r}$) and log($SPI_{d,r}$) for those diseases that have more than one severity and includes 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. Here, 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, 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.

[28] 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 shows the MCE_r for each MSA in the data, although some MSAs are not shown due to confidentiality concerns.[29] The MCE_r ranges from a high of 1.29 in Milwaukee to a low of 0.80 in Youngstown, PA. 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 1.27, 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 Fort-Worth-Arlington, TX, have higher expenditures due to higher than average prices and utilization.

The variation in these MSA indexes gives some measure of the overall spending variation per disease and the source of the variation. The COV for the aggregate MCE_r is 0.10. This 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.[30] More generally, the variation in these indexes is low relative to price and spending differences across regions for other goods and services. Aten and D'souza (2008) find a COV of 0.15 based on a price index for all goods and services across MSAs.[31] 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. The variation in the two components of the MCE_r, the SUI_r and the SPI_r, are below this value. The variation in the SUI_r (0.071) is actually lower than the variation in the SPI_r (0.10).[32] Therefore, based on these aggregate figures, average price differences across areas appear to be larger than utilization differences. As a check on these figures, we examine the correlation between our SPI_r and the regional price indexes for all goods and services from Aten and D'souza (2008) that may be viewed as a regional cost of living index. We find a positive correlation coefficient of 0.45 between the log of SPI_r and the log of their regional price index that is significant at the 1 percent level, which is consistent with service prices varying with the cost of living across areas. We find no correlation between

[29] In cases where an MSA cannot be reported, we provide the most disaggregate level of information that we are allowed (e.g. "MSA in the South" or "MSA in Michigan").

[30] The CBO also reports the COV for state-level Medicare spending per capita of 0.11 in 2005.

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

[32] The lower variation in the utilization index is not driven by selecting the coefficient of variation as the measure of dispersion. We also show differences in the 90th and 10th percentile, which also show that the utilization variation tends to be less.

the log of SUI_r and the log of the regional price index.[33] While it is likely that medical care service prices reflect the cost of living in an area, many idiosyncratic factors may impact utilization.

Table 5. MSA Medical-Care Price Indexes and Variation in Indexes - MCE_r, SPI_r and SUI_r

MSA Name	Rank MCE_r	MCE_r	Rank SPI_r	SPI_r	Rank SUI_r	SUI_r
Milwaukee-Waukesha-West Allis, WI	1	1.293	2	1.273	39	1.004
Salinas, CA	2	1.247	1	1.406	84	0.887
MSA in the Midwest	3	1.238	7	1.149	8	1.073
Oakland-Fremont-Hayward, CA	4	1.219	4	1.265	53	0.972
MSA in the Midwest	5	1.169	16	1.095	12	1.062
Minneapolis-St. Paul-Bloomington, MN-WI	6	1.160	8	1.143	26	1.018
Gary, IN	7	1.143	44	1.003	3	1.127
Fort Worth-Arlington, TX	8	1.142	11	1.108	22	1.034
Indianapolis, IN	9	1.142	13	1.106	27	1.018
MSA in California	10	1.132	3	1.271	81	0.902
Peoria, IL	11	1.129	10	1.120	16	1.047
Dallas-Plano-Irving, TX	12	1.129	9	1.124	31	1.016
Houston-Sugar Land-Baytown, TX	13	1.125	20	1.084	21	1.035
Miami-Miami Beach-Kendall, FL	14	1.106	22	1.055	18	1.044
Denver-Aurora, CO	15	1.100	33	1.030	11	1.063
Pittsburgh, PA	71	0.907	84	0.855	7	1.090
Louisville, KY-IN	72	0.903	70	0.936	62	0.956
MSA in the South	73	0.901	58	0.972	76	0.920
MSA in the South	74	0.899	23	1.053	85	0.875
Nassau-Suffolk, NY	75	0.898	37	1.018	83	0.896
Memphis, TN-MS-AR	76	0.897	46	1.001	74	0.926
Kingsport-Bristol-Bristol, TN-VA	77	0.887	83	0.887	32	1.013
Providence-New Bedford-Fall River, RI-MA	78	0.883	79	0.897	30	1.016
Warren-Farmington Hills-Troy, MI	79	0.876	78	0.912	51	0.979
MSA in the South	80	0.876	59	0.970	75	0.923
Detroit-Livonia-Dearborn, MI	81	0.867	80	0.896	44	0.993
MSA in Michigan	82	0.867	72	0.926	67	0.939
Augusta-Richmond County, GA-SC	83	0.865	73	0.924	71	0.930
MSA in the South	84	0.848	75	0.923	77	0.920
Youngstown-Warren-Boardman, OH-PA	85	0.797	85	0.834	55	0.969
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

Although the variation appears to be similar 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 MCE_r (0.10) is less than half of the size of the COV of the $MCE_{d,r}$ (0.22).[34] This finding is of particular interest because it corresponds with the fixed-effects

[33] The correlation coefficient is -0.178 and is not significant at the 10 percent level.

[34] 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 3. 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

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. 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. 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. For example, Gary, IN, is ranked as the highest utilization city based on the aggregated SUI, but Gary ranks below the average based on utilization for the disease category "Severity 1 Mood Disorder, Depressed".

5.3 MSA-Service-Type Indexes

Both Tables 3 and 5 in the previous sections suggest that price variation in the underlying service types may play an important role in explaining differences in expenditures across markets. Table 6 below shows the key service types that are studied in the data along with the amount of expenditures associated with each type.

Table 6. Total 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	$77	17.7%
Outpatient Hospital	$101	23.2%
Office - General MD	$39	8.9%
Office - Specialist MD	$69	15.9%
Other (Emergency, Ambulatory Centers etc)	$54	12.3%
Pharmacy	$96	22.0%
Total	$436	100.0%

Given the importance of service prices in explaining differences in expenditures across areas, we examine the underlying service types that might explain these differences. To do so, we create MSA-service-type indexes which are meant to capture variation in spending,

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 alteranative disease categories, the variation we observe at this level remains considerably larger than the aggregate SUI.

prices and utilization across MSAs for the same type of service. 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 $SUI_{r,s}$.[35] Table 7 shows the variation in the indexes for each of the main service types. Overall, it appears that both Outpatient Hospital and Office-General MD spending vary by the most, with pharmacy spending varying the least. Interestingly, most of the variation in the expenditures at this level of aggregation stem from variation in the price of services. Prescription drug prices vary the least with a COV of 0.07 with inpatient and outpatient hospital service prices varying the most with a COV of 0.21. 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. 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	0.215	0.206	0.073
Outpatient Hospital	0.247	0.206	0.078
Office - General MD	0.246	0.119	0.059
Office MD - Speciality	0.185	0.117	0.066
Other (Emergency, Ambulatory Centers, etc)	0.212	0.155	0.071
Pharmacy	0.077	0.066	0.059
Weighted Average	0.190	0.148	0.069

Although the price variation appears to be relatively large for many of the services, there is relatively little variation in the utilization index across these service types. Similar to the SUI_r, the likely reason for the limited variation is that the MSA service-type index averages over disease types. By contrast, the price levels for the different services may be common across diseases for a specific MSA due to reasons related to bargaining leverage of

[35]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.

physicians or hospitals (See Dunn and Shapiro (2011)). For example, a general MD who negotiates a higher price with an insurer (for example, 10 percent above Medicare rates) will receive a higher service price regardless of the disease of his patients. A more detailed listing of these specific indexes at the service-type level is shown in the appendix of this paper.

6 The Economic Importance of Price and Utilization Variation

6.1 MSA Variation

This section presents analysis demonstrating the economic importance of the variation in service price and utilization across markets. Specifically, we focus on the potential expenditure reduction if either utilization or prices were shifted from the levels observed in high expenditure MSAs to the levels observed in the low expenditure MSAs, identified by our measure MCE_r. It is important to note that these exercises are solely meant to highlight the importance of observed variation across markets because they ignore the behavioral response to market changes.

To perform these exercises, we first rank the MSAs based on MCE_r and place each MSA into one of four "quartile-bins." Next, we calculate the average price for each service type and disease in each quartile-bin, creating the vector $p_d^{quart}(MCE_r)$, where the MCE_r in parentheses indicates that we are ranking MSAs by this measure. We then take the average price in the low-spending quartiles, $p_d^{25}(MCE_r)$, and apply these price levels to actual utilization levels:

$$\text{Price Change:} \qquad MCE_{d,r}^{p_d^{25}(MCE_r)} = \frac{p_d^{25}(MCE_r) \cdot q_{d,r}}{c_{d,B}}.$$

One way to view $MCE_{d,r}^{p_d^{25}(MCE_r)}$ is that it is a measure of the expenditure savings from lower prices.

Similarly, to look at the reduction in expenditure from shifting service utilization, we take the average utilization of each service for treating each disease in the low-spending areas, $q_d^{25}(MCE_r)$, and apply these prices to actual utilization levels in high-spending areas:

$$\text{Utilization Change:} \qquad MCE_{d,r}^{q_d^{25}(MCE_r)} = \frac{p_{d,r} \cdot q_d^{25}(MCE_r)}{c_{d,B}}.$$

The results of this analysis applied to those MSAs in the top quartile are shown in Table 8.1. The table shows an average 4.5 percent decrease in episode expenditures from the utilization reduction, and a 14.1 percent change from the price reduction. So it appears that price is a more important factor than utilization when looking at aggregate healthcare expenditures in high-spending areas compared to low-spending areas. This is consistent with the observation that the variation in the SPI_r was greater than the variation in the SUI_r. However, the precise effect of these shifts depends on the specific MSA. In some MSAs the savings are much greater than in others (for example, the higher MCE areas tend to have larger savings). In addition, in some cases the price shift may actually lead to an increase in overall expenditures. For example, in markets such as Boston spending is higher due to utilization, not higher service prices. In this case, substituting prices from the low-spending areas can actually increase the MCE index because the prices in the low-spending area are actually higher than the prices in Boston.

Table 8.1 MCE_r Changes from Shifting Prices and Utilization from the Highest Spending Quartile to Average Levels in the Lowest Spending Quartile

		Price Change		Utilization Change	
MSA	MCE_r	New MCE_r	Change	New MCE_r	Change
Milwaukee-Waukesha-West Allis, WI	1.293	0.950	-0.265	1.235	-0.045
Salinas, CA	1.247	0.839	-0.327	1.359	0.090
MSA in the Midwest	1.238	1.017	-0.179	1.113	-0.101
Oakland-Fremont-Hayward, CA	1.219	0.919	-0.246	1.225	0.005
MSA in the Midwest	1.169	1.004	-0.141	1.061	-0.092
Minneapolis-St. Paul-Bloomington, MN-WI	1.160	0.969	-0.165	1.109	-0.044
Gary, IN	1.143	1.069	-0.065	0.971	-0.150
Fort Worth-Arlington, TX	1.142	0.979	-0.143	1.076	-0.058
Indianapolis, IN	1.142	0.967	-0.153	1.071	-0.062
MSA in California	1.132	0.855	-0.244	1.232	0.088
Peoria, IL	1.129	0.993	-0.121	1.085	-0.040
Dallas-Plano-Irving, TX	1.129	0.953	-0.156	1.089	-0.035
Houston-Sugar Land-Baytown, TX	1.125	0.980	-0.129	1.048	-0.068
Miami-Miami Beach-Kendall, FL	1.106	0.988	-0.107	1.024	-0.075
Denver-Aurora, CO	1.100	1.004	-0.087	0.999	-0.092
Portland-Vancouver-Beaverton, OR-WA	1.094	0.995	-0.091	1.059	-0.032
San Diego-Carlsbad-San Marcos, CA	1.080	0.880	-0.185	1.129	0.045
Tacoma, WA	1.072	0.992	-0.074	0.998	-0.069
Boston-Quincy, MA	1.069	1.126	0.053	0.962	-0.100
MSA in Mississippi	1.064	1.065	0.001	0.921	-0.134
Bridgeport-Stamford-Norwalk, CT	1.049	0.906	-0.137	1.074	0.023
Average Change in MCE_r			-0.141		-0.045
Median Change in MCE_r			-0.141		-0.058

Notes. This table looks at those cities that are ranked in the highest quartile based on the MCE_r. The first column reports the MCE_r indexes for these cities. Next the table reports two alternative MCE_r indexes. First, we examine how the MCE_r would change when prices in the low-spending quartile regions (based on the MCE_r) are applied to the high quartile regions. Second, we examine how the MCE_r would change when quantities in the low-spending quartile regions (based on the MCE_r) are applied to the high quartile regions. Both the new MCE_r and the change in the MCE_r are reported for these two excercises.

Alternatively we conduct a similar analysis as above, but instead we sort the MSAs by the SUI_r instead of the MCE_r. The analysis reported in Table 8.2 is identical to the analysis in Table 8.1, but the quartiles of the data are divided based on the SUI_r as opposed to the MCE_r. That is, we examine the effects of taking the utilization and prices observed in the low utilization MSAs, and applying those amounts to the high utilization areas. Using the notation above, we would be calculating the price changes ($p_d^{25}(SUI_r)$) and utilization changes ($q_d^{25}(SUI_r)$) based on the SUI_r quartiles, to calculate new MCE_r values, $MCE_{d,r}^{p_d^{25}(SUI_r)}$ and $MCE_{d,r}^{q_d^{25}(SUI_r)}$. As one would expect, we find larger expenditure savings of around 12 percent from a shift in utilization. However, if the high utilization areas take on the price levels in low utilization areas, the expenditures actually *increase* by 12 percent on average. Therefore, it appears that low utilization areas also tend to be higher price areas on average.

Table 8.2 MCE_r Changes from Shifting Prices and Utilization from the Highest Utilization Quartile to Average Levels in the Lowest Utilization Quartile

		Price Change		Utilization Change	
MSA	MCE_r	MCE_r Adj.	Change	MCE_r Adj.	Change
MSA in the Midwest	1.238	1.150	-0.071	1.056	-0.147
MSA in the Midwest	1.169	1.145	-0.020	1.014	-0.133
Gary, IN	1.143	1.227	0.074	0.921	-0.194
Peoria, IL	1.129	1.127	-0.002	1.028	-0.090
Houston-Sugar Land-Baytown, TX	1.125	1.104	-0.018	0.994	-0.116
Miami-Miami Beach-Kendall, FL	1.106	1.108	0.002	0.958	-0.134
Denver-Aurora, CO	1.100	1.131	0.028	0.950	-0.136
Portland-Vancouver-Beaverton, OR-WA	1.094	1.138	0.040	1.031	-0.058
Tacoma, WA	1.072	1.123	0.048	0.964	-0.100
Boston-Quincy, MA	1.069	1.324	0.239	0.940	-0.120
MSA in the South	1.064	1.221	0.148	0.887	-0.166
Cambridge-Newton-Framingham, MA	1.049	1.337	0.275	0.926	-0.117
Kalamazoo-Portage, MI	1.018	1.172	0.150	0.928	-0.089
Birmingham-Hoover, AL	1.016	1.116	0.098	0.876	-0.138
Philadelphia, PA	1.000	1.279	0.280	0.902	-0.097
MSA in the Northeast	0.997	1.214	0.218	0.857	-0.140
MSA in the Midwest	0.990	1.223	0.235	0.833	-0.159
St. Louis, MO-IL	0.988	1.143	0.157	0.862	-0.128
Jacksonville, FL	0.980	1.115	0.137	0.868	-0.114
Akron, OH	0.917	1.137	0.239	0.824	-0.101
Pittsburgh, PA	0.907	1.192	0.314	0.803	-0.115
			0.122		-0.123
			0.137		-0.120

Notes. This table looks at those cities that are ranked in the highest quartile based on the SUI_r. The first column reports the MCE_r indexes for these cities. First, we examine how the MCE_r would change when prices in the low-utilization quartile regions (based on the SUI_r) are applied to the high quartile regions. Second, we examine how the MCE_r would change when quantities in the low-utilization quartile regions (based on the SUI_r) are applied to the high quartile regions. Both the new MCE_r and the change in the MCE_r are reported for these two excercises.

6.2 Disease Variation

The calculation of the reduction in expenditures from adjusting utilization and prices may be much different than the amounts reported in Tables 8.1 and 8.2 if one looks at the expenditures from treating specific diseases across cities. For example, one location may have very low utilization for diabetes, but high utilization for pregnancy, which may be ignored by the overall utilization index. As an alternative measure of the potential savings, we focus on high expenditure and low expenditure areas on a disease-by-disease basis. That is, for each disease, we sort the data into low expenditure and high expenditure MSA quartile-bins. This exercise is identical in fashion to that shown above, but the MSA ranking into different quartiles is done by $MCE_{d,r}$ instead of MCE_r, so that an MSA may belong in a high-quartile bin for one disease and a low-quartile bin for another. That is, instead of substituting prices and utilization measures based on MSA rankings, $p_d^{25}(MCE_r)$ and $q_d^{25}(MCE_r)$, we will be substituting based on disease rankings, $p_d^{25}(MCE_{d,r})$ and $q_d^{25}(MCE_{d,r})$. These steps are taken for every disease in the data and we calculate the weighted average savings using national expenditure shares of each disease.

Table 9.1 reports the results from the analysis where savings is computed for each disease by comparing prices and utilization from high-spending MSAs for each disease to the low-spending areas for each diseases. As one might expect, those diseases with relatively large variation in price (as measured by the SPI), such as "Pregnancy, with Delivery", observe a greater reduction in expenditure from shifts in prices, relative to shifts in utilization. On the other hand, those diseases with greater relative variation in utilization, such as "Mood Disorder, Depressed", show a greater reduction in expenditure from shifts in utilization, relative to shifts in prices. The weighted average savings for the "typical" disease is shown at the bottom of Table 9.1. The typical reduction in expenditures from the shift in price is 16 percent and the reduction in expenditures is 23 percent based on utilization changes. That is, it appears that savings are quite large based on both price and utilization when one examines disease-by-disease expenditure differences. Even when the sample is limited to the most frequently observed diseases (i.e. those with more than 50,000 observations in the data) we observe savings of similar magnitude, as shown in the last row of the table.[36]

[36] The reduction in expenditures reported here is large, and prior research looking at Medicare data suggests that large reductions in utilization may be attained without a reduction in quality. In particular, research examining the Medicare population suggests that savings may be as great as 20 to 30 percent if services of questionable value are no longer performed (See Skinner et al. (2005) and Fisher et al. (2003)).

Table 9.1. Percentage Change in Expenditures from Shifts in Utilization and Shifts in Price by Disease - From High-Spending Quartiles to Low-Spending Quartiles

	Description	Severity	%Chg in $MCE_{d,r}$ from Shift in Price	%Chg in $MCE_{d,r}$ from Shift in Utilization
1	Pregnancy, with delivery	1	-23.9%	-9.1%
	Pregnancy, with delivery	2	-25.6%	-9.0%
2	Joint degeneration, localized - back	1	-15.6%	-17.6%
	Joint degeneration, localized - back	2	-22.3%	-27.9%
	Joint degeneration, localized - back	3	-25.0%	-28.2%
3	Ischemic heart disease	1	-17.4%	-24.8%
	Ischemic heart disease	2	-22.6%	-15.7%
	Ischemic heart disease	3	-21.4%	-23.5%
4	Hypertension	1	-3.4%	-10.3%
	Hypertension	2	-5.9%	-13.2%
	Hypertension	3	-9.5%	-18.1%
	Hypertension	4	-15.8%	-33.8%
5	Diabetes	1	-3.6%	-19.0%
	Diabetes	2	-4.5%	-25.9%
	Diabetes	3	-10.0%	-27.8%
	Diabetes	4	-15.2%	-25.3%
6	Routine exam	1	-21.0%	-8.2%
7	Mood disorder, depressed	1	-7.5%	-29.4%
	Mood disorder, depressed	2	-15.4%	-35.0%
	Mood disorder, depressed	3	-13.4%	-35.9%
8	Malignant neoplasm of breast	1	-19.2%	-23.5%
	Malignant neoplasm of breast	2	-17.3%	-32.3%
9	Hyperlipidemia, other	1	-5.8%	-12.5%
10	Joint degeneration, localized - neck	1	-6.2%	-28.9%
	Joint degeneration, localized - neck	2	-14.9%	-34.5%
	Joint degeneration, localized - neck	3	-23.3%	-29.9%
11	Chronic sinusitis	1	-11.6%	-24.5%
	Chronic sinusitis	2	-8.9%	-31.6%
	Chronic sinusitis	3	-22.9%	-27.5%
12	Joint degeneration, localized - knee & lower leg	1	-19.5%	-27.1%
	Joint degeneration, localized - knee & lower leg	2	-20.5%	-32.3%
	Joint degeneration, localized - knee & lower leg	3	-12.1%	-36.9%
13	Asthma	1	-3.8%	-19.4%
	Asthma	2	-5.6%	-15.6%
	Asthma	3	-9.6%	-25.9%
	Asthma	4	-8.7%	-31.3%
14	Joint derangement - knee & lower leg	1	-22.4%	-25.9%
	Joint derangement - knee & lower leg	2	-28.4%	-7.8%
15	Inflammation of esophagus	1	-11.5%	-13.3%
	Inflammation of esophagus	2	-14.6%	-13.4%
	Inflammation of esophagus	3	-20.6%	-29.9%
	Weighted Average Savings (Full Sample - 10,000 Episodes in Data)		-15.5%	-23.0%
	Weighted Average Savings (Only Diseases with 50,000 Episodes in Data)		-13.4%	-19.5%

Notes. For each disease, we rank each MSA based on the disease-expenditure index, $MCE_{d,r}$. We then calculate potential savings for that disease in the highest quartile spending area. First, we calculate potential savings by shifting prices from the lowest spending areas for that disease (based on the $MCE_{d,r}$) to the highest spending areas. Second, we calcualte potential savings by shifting utilization from the lowest spending quartiles for that disease (based on the $MCE_{d,r}$) to the highest spending quartiles. For both calculations we report the average change in the MCE for the highest spending quartile MSAs for that disease.

In contrast to Table 9.1 that ranks MSAs into high and low expenditure areas for each disease, Table 9.2 categorizes MSAs into high and low utilization areas for each disease (i.e. ranking each area by $SUI_{d,r}$). Looking at the second column, one can see potentially large reductions in expenditures from shifting utilization levels from the low utilization areas to the high utilization areas. The weighted average savings in expenditures for the typical disease is 31 percent. However, looking at the first column, when the prices from the low utilization areas are substituted into the high utilization areas, we see an *increase* in expenditures for most diseases. This suggests a negative correlation between price and utilization at the disease level. Simple correlations between the SUI_r and SPI_r or the $SUI_{d,r}$ and $SPI_{d,r}$ 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.

Table 9.2. Percentage Change in Expenditures from Shifts in Utilization and Shifts in Price by Disease - From High Utilization Quartiles to Low Utilization Quartiles

	Description	Severity	%Chg in $MCE_{d,r}$ from Shift in Price	%Chg in $MCE_{d,r}$ from Shift in Utilization
1	Pregnancy, with delivery	1	4.8%	-16.1%
	Pregnancy, with delivery	2	16.8%	-19.0%
2	Joint degeneration, localized - back	1	10.4%	-25.7%
	Joint degeneration, localized - back	2	5.6%	-35.0%
	Joint degeneration, localized - back	3	10.8%	-39.5%
3	Ischemic heart disease	1	18.1%	-36.4%
	Ischemic heart disease	2	22.2%	-30.5%
	Ischemic heart disease	3	16.7%	-39.2%
4	Hypertension	1	16.5%	-20.9%
	Hypertension	2	17.8%	-22.9%
	Hypertension	3	9.3%	-24.0%
	Hypertension	4	27.8%	-43.2%
5	Diabetes	1	4.6%	-20.7%
	Diabetes	2	11.0%	-31.1%
	Diabetes	3	3.4%	-32.4%
	Diabetes	4	7.3%	-34.4%
6	Routine exam	1	-5.2%	-12.4%
7	Mood disorder, depressed	1	-2.1%	-31.3%
	Mood disorder, depressed	2	-5.4%	-38.4%
	Mood disorder, depressed	3	7.1%	-40.3%
8	Malignant neoplasm of breast	1	10.5%	-29.8%
	Malignant neoplasm of breast	2	7.4%	-39.8%
9	Hyperlipidemia, other	1	10.1%	-17.4%
10	Joint degeneration, localized - neck	1	12.9%	-31.1%
	Joint degeneration, localized - neck	2	32.4%	-42.1%
	Joint degeneration, localized - neck	3	33.9%	-42.6%
11	Chronic sinusitis	1	-3.3%	-27.9%
	Chronic sinusitis	2	0.0%	-34.5%
	Chronic sinusitis	3	7.6%	-34.6%
12	Joint degeneration, localized - knee & lower leg	1	11.3%	-34.4%
	Joint degeneration, localized - knee & lower leg	2	31.4%	-44.5%
	Joint degeneration, localized - knee & lower leg	3	36.3%	-50.0%
13	Asthma	1	3.5%	-21.1%
	Asthma	2	6.2%	-20.3%
	Asthma	3	9.2%	-32.2%
	Asthma	4	14.5%	-39.0%
14	Joint derangement - knee & lower leg	1	18.2%	-34.5%
	Joint derangement - knee & lower leg	2	35.6%	-30.1%
15	Inflammation of esophagus	1	8.2%	-18.4%
	Inflammation of esophagus	2	10.1%	-22.7%
	Inflammation of esophagus	3	8.5%	-36.2%
	Weighted Average Savings (Full Sample - 10,000 Episodes in Data)		14.2%	-31.2%
	Weighted Average Savings (Only Diseases with 50,000 Episodes in Data)		11.9%	-26.6%

Notes. For each disease, we rank each MSA based on the disease-service utilization index, $SUI_{d,r}$. We then calculate potential savings for that disease in the highest quartile utilization area. First, we calculate potential savings by shifting prices from the lowest utilization areas for that disease (based on the $SUI_{d,r}$) to the highest utilization areas. Second, we calcualte potential savings by shifting utilization from the lowest utilization quartiles for that disease (based on the $SUI_{d,r}$) to the highest utilization quartiles. For both calculations we report the average change in the MCE for the highest utilization quartile MSAs for that disease.

Overall, Tables 8.1, 8.2, 9.1 and 9.2 demonstrate that potential savings from either price or utilization shifts may be much greater if one focuses on changing pricing or utilization at the disease level. In addition, it appears that both differences in price and utilization are important determinants of spending differences across markets, which contrasts with the Medicare markets where utilization differences appear to be more important. Finally, another key finding in this section is that price and utilization appear to be negatively correlated, both at the disease level and when aggregate measures are applied. Therefore, areas that may appear to be low cost based on utilization measures often have higher prices.

7 A Comparison With a Population-Based Measure

The episode-based measure assessed in this study provides insight on regional differences in the efficiency of care because it measures spending, price, and utilization only for those individuals being treated. As we discussed earlier, this measure does 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 region, 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 first calculate the treated prevalence of each disease in the population as the number of episodes being treated divided by the total population in each MSA, $prev_{d,r} = \frac{N_{d,r}}{Population_r}$. Age and gender weights are applied to each MSA, so that the total age and sex distribution is identical across all MSAs. Using the measure of treated prevalence, the disease cost per population may be calculated by multiplying the expenditure per episode by the treated prevalence: $c^{pop}_{d,r} = c_{d,r} \cdot (prev_{d,r}) = \sum_s p_{d,r,s} Q_{d,r,s} / Population_r$. Similarly, the utilization per population is derived as $q^{pop}_{d,r} = q_{d,r} \cdot (prev_{d,r}) = Q_{d,r}/Population_r$. The corresponding population-based MCE and SUI follow:[37]

$$MCE^{pop}_{d,r} = \frac{c^{pop}_{d,r}}{c^{pop}_{d,B}} \tag{2}$$

$$SUI^{pop}_{d,r} = \frac{p_{d,B} \cdot q^{pop}_{d,r}}{c^{pop}_{d,B}}$$

A prevalence index may also be constructed by dividing the disease prevalence relative to the base region's prevalence.

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

[37] The value of $c^{pop}_{d,B}$ is just an average over disease expenditures per population across the MSAs in the sample. The SPI measure does not change.

Using the above equations, the population-based MCE, $MCE_{d,r}^{pop}$, may be decomposed into its two components which include the episode-based index, $MCE_{d,r}$ and the treated prevalence index, $PREV_{d,r}$:

$$\log(MCE_{d,r}^{pop}) = \log(MCE_{d,r}) + \log(PREV_{d,r}). \tag{3}$$

This equation makes it clear that a population-based measure of expenditure for a particular disease will rise if there is either an increase in the prevalence of the disease or an increase in the expenditures per episode.

We can create aggregate MSA indexes for the population-based measure in a similar fashion to the episode-based measure. When $MCE_{d,r}^{pop}$ is weighted by the national expenditure share for each disease, this simply becomes a measure of medical-care expenditures per person relative to the base region's medical-care expenditures per person:

$$\begin{aligned} MCE_r^{pop} &= \sum_D MCE_{d,r}^{pop} \cdot (\text{Expenditure Share}) \\ &= \sum_D \frac{C_{d,r}^{pop}}{C_{d,B}^{pop}} \cdot \left(\frac{C_{d,B}^{pop}}{\sum_D C_{d,B}^{pop}} \right) = \frac{\sum_D C_{d,r}^{pop}}{\sum_D C_{d,B}^{pop}} \\ &= \frac{\text{Medical-Care Expenditures Per Person}_r}{\text{Medical-Care Expenditures Per Person}_B}. \end{aligned}$$

This is the same measure depicted in the first column of Table 1 and, as we discussed previously, *includes* any variation attributable to the prevalence of certain diseases. Thus this measure will be high in areas in which there is a large prevalence of disease (poor health areas), but especially high in areas with a large prevalence of expensive diseases.

Table 10 reports the variation observed using the episode-based approach, MCE_r, compared to the variation observed using a population-based approach, MCE_r^{pop}. The first three columns repeat the results shown at the bottom of Table 5. The middle column shows the amount of variation in treated prevalence, which is the prevalence measure weighted by the national expenditure share of the different diseases. It appears that differences in the number of episodes across regions varies about the same amount as the other aggregate measures, with a coefficient of variation of 0.08. The following two columns report the MCE_r^{pop} and SUI_r^{pop} measures, which are similar in magnitude to the episode-based MCE_r and SUI_r measures. The variation in MCE_r^{pop} is a bit smaller than the MCE_r and the coefficient of variation for SUI_r^{pop} is slightly larger than the corresponding episode-based measure.

The correlation between the episode-based measures and population-based measures is positive and statistically significant, but the magnitudes of the correlations indicate that these are quite distinct measures. The correlation between the two types of MCE measures is positive and highly significant, with a correlation coefficient of 0.65, while the correlation is lower for the SUI with a correlation of 0.51.[38] The higher correlation in the MCE indexes relative to the SUI should be expected because the two MCE indexes share a common price component through the SPI. Table A10.1 and A10.2 in the appendix reports these measures for specific cities and shows that the measures may be very different in many instances. For example, in Warren-Farmington Hills-Troy, MI, the episode-based SUI measure is 0.979 (and ranked 52nd out of 85 cities), but the population-based measure is 1.16 (and ranked 4th).

Table 10. MSA Variation in Medical-Care Indexes - Population-Based and Episode-Based Indexes

	Episode-Based			Prevalence	Population-Based	
	MCE_r	SPI_r	SUI_r	$PREV_r$	MCE^{pop}_r	SUI^{pop}_r
mean	1.000	1.018	0.996	1.000	1.000	1.000
sd	0.098	0.097	0.064	0.078	0.079	0.085
COV	0.098	0.095	0.064	0.078	0.079	0.085
p10	0.887	0.915	0.920	0.902	0.912	0.888
p90	1.142	1.124	1.073	1.126	1.132	1.119
N	85	85	85	85	85	85

8 Regional Diagnostic Practices

This paper focuses on analyzing the variation in spending by disease-episode, which controls for health status by categorizing patients into disease bins. In this respect, our work is similar to other work that analyzes variation in expenditures by controlling for disease, such as, Zhang et al. (2010) and MedPac (2003, 2009). One advantage of categorizing patients by disease when analyzing variation in medical-care expenditures is that one may obtain more precise estimates of spending, since this methodology accounts for the unique health condition of the individual.

A disadvantage of this approach is that the diagnosis decisions may vary geographically, which is suggested by recent research of Song et al. (2010). This study finds that an individual moving to a high-spending area from a low-spending area receives a greater number of diagnosis than she would have received if she had not moved. In this case, it is possible that a patient given a diagnosis code indicating a minor illness in one MSA may be given a diagnosis code indicating a more severe illness in another MSA. This could cause

[38] Similar correlations between the episode-based and population-based SUI and MCE measures are found at the disease level. These correlations are all statistically significant at the 1 percent level.

an episode-based measure to understate the cost of treatment in high-spending areas. For instance, utilization may appear low for patients with a severe disease in a certain MSA because these patients are not relatively sick patients. This possibility is consistent with the findings of Welch et al. (2011) which shows substantial variation in geographic frequency of diagnosis, but little variation in population-based mortality rates across high and low diagnosis areas. In more general terms, if less severe cases are more likely to be treated in some regions than others, then areas with higher treated prevalence may have lower average spending for the same disease category.

As a simple check on the potential bias of the episode-based measure, we estimate the regression:

$$\log(SUI_{d,r}) = \alpha \log(PREV_{d,r}) + \varepsilon_{d,r} \tag{4}$$

If $\alpha = 0$ this would indicate that treated prevalence is not associated with utilization levels, which would provide evidence that there is no systematic difference in treated individuals for the same diseases between areas with different degrees of prevalence.

Table 11 presents estimates of the regression of log service utilization on log prevalence. The relationship is negative and statistically significant, indicating some bias in the episode-based measure; however, the magnitude of the bias is economically small, with a coefficient of -0.09, which implies that increasing prevalence by 10 percent reduces utilization by 0.9 percent — implying a negative bias of 0.9 percent. We also include additional specifications that control for endogeneity[39], service price[40], and MSA fixed effects. Including service price as a control variable, model (2), shows a negative relationship between service price and utilization, but has practically no effect on the prevalence coefficient. Including MSA fixed effects, model (3), has very little effect on either the price or prevalence estimates. As a final check, we include instruments for treated prevalence and service prices.[41] As an instrument for price, we use service prices for other diseases,

[39] Since those areas with higher levels of utilization may also have higher levels of treated prevalence, there might be positive bias in the empirical relationship between $SUI_{d,r}$ and $PREV_d^r$.

[40] Not controlling for service prices may also introduce a positive bias, since $SUI_{d,r}$ and $PREV_d^r$ are potentially both negatively correlated with price (i.e. high price areas have lower utilization and fewer individuals receive diagnosis).

[41] Although MSA fixed effects are included, there are still potential endogeneity concerns with both prevalence and service price. For prevalence, doctors may raise utilization by assigning more diagnoses, so high utilization areas may have high prevalence, causing a correlation between ε_d^r and $\log(PREV_d^r)$. The instruments for treated prevalence are based on the treated prevalence for other diseases. The justification for this instrumenting strategy is that diagnosis rates of other diseases are likely assigned by distinct

since the quality of treatment for different diseases is likely distinct, but there are likely common cost factors that are correlated with price (e.g. nurses' wages). The results that apply instrumental variables appear in model (4), which shows a negative but insignificant relationship between prevalence and utilization.

Table 11. Regressions of log($SUI_{d,r}$) on the Log of Treated Prevalence

	(1)	(2)	(3)	(4)
log($PREV_{d,r}$)	-0.0912***	-0.100***	-0.0798***	-0.0229
	(-8.38)	(-9.30)	(-9.39)	(-0.78)
log($SPI_{d,r}$)		-0.167***	-0.131***	-0.363***
		(-7.62)	(-7.72)	(-5.93)
MSA Fixed Effects	No	No	Yes	Yes
Price Control	No	Yes	Yes	Yes
Instrumental Variable	No	No	No	Yes
R^2	0.0328	0.0498	0.0402	0.00575
N	26348	26321	26321	26151

t statistics in parentheses * p<0.05, ** p<0.01, *** p<0.001

Notes. Results include clusters for standard errors at the MSA level. These results are robust to a number of alternative specifications, such as the inclusion of MPC Disease Category-MSA fixed effects or removing outliers from the analysis. Specification (4) includes instrumental variables for both price and prevelance. The instruments for price are the SPI measures for the other ETGs in the same MPC cateogry weighted by expenditure share, and the SPI measures for diseases in other MPC classes in the MSA weighted by expenditure share. The instruments for prevalence include prevalence for other diseases in the MPC for the MSA weighted by expenditure share and the prevelance for diseases in other MPC classes in the MSA weighted by expenditure share. Similar results are found if only the prevalence and price in other MPC classes is used as an instrument.

For each of the four specifications, the magnitude of the coefficient on prevalence is quite small and suggests that regional diagnosis differences may not be problematic. This can also be revealed through the very high correlation (0.99) between the residuals of specification (4) and the $SUI_{d,r}$. This indicates that controlling for prevalence makes little difference to the underlying indexes.[42] Although the bias appears small, it may be

physicians, which is consistent with only a small amount of the variation in prevalence across diseases being explained by MSA fixed-effects, as demonstrated by Table A10.3 in the appendix. However, the prevalence rates that are common across disease categories are more likely due to random cultural factors. For distinct reasons, it is also possible for the price variable to be endogenous. For instance, a higher price may indicate a higher product quality for that service, which would introduce a positive bias on the price coefficient.

[42] Another check on this relationship between utilization and prevalence was conducted to determine if overall prevalence may be associated with utilization. For this analysis, we exclude MSA fixed effects and include the log of overall prevalence along with state fixed effects. In fact, we found a strong and significant negative relationship between overall prevalence and utilization, consistent with a potential

beneficial to re-examine these issues at the micro level in future work (similar to the type of analysis conducted by Dunn and Shapiro (2011)). An advantage of focusing on more micro level data is that additional demographic and regional controls may be included in the model to account for alternative variables that may impact utilization patterns for each individual and the potential selection bias may be addressed using standard micro-econometric techniques, such as a Heckit procedure.

9 Medical-Care Expenditures and Quality Measures

There is considerable variation in spending, service prices and service utilization, but it is unclear what this variation means for consumer surplus, since greater spending may be associated with high-quality care. Although this relationship between spending and quality does not appear to be present in Medicare markets,[43] less research has been conducted in the commercial sector. Here we use a set of procedural quality measures constructed from the MarketScan data to examine whether there is an association between quality and the MCE, SUI and SPI. The quality measures are constructed using Healthcare Effectiveness Data and Information Set (HEDIS) guidelines from the National Committee for Quality Assurance (NCQA). Additional details regarding the construction of the quality measures are included in the appendix.

The top portion of Table 12 shows the correlation between the log of specific quality measures and the log of the $MCE_{d,r}$, $SPI_{d,r}$, and $SUI_{d,r}$ for the corresponding diseases. For instance, we compare the quality measure indicating beta-blocker treatment with indexes where d = hyptertension and d = ischemic heart disease. Looking at correlations between disease-specific quality measures and their corresponding disease indexes, there does not appear to be a consistent pattern. While the correlation for diabetes testing for pediatric patients and cholesterol testing for patients with ischemic vascular disease show some positive and significant relationships, we see that many of the other correlations are insignificant or significant and negative.

bias. However, correcting for this potential bias, we find that the corrected utilization measures remained quite close to the episode-based measures (a correlation of 0.97). In addition, the negative relationship between overall prevalence and utilization shrinks considerably if we remove outliers from the sample, but the relationship with the disease-specific prevalence measure remains. Finally, one should keep in mind that this relationship between overall prevalence and utilization does not include MSA fixed effects, so there may be other MSA-specific factors causing this relationship, which may not signify any bias.

[43] For example, one study in this area by Fisher et al. (2003) shows that patients in high spending areas do not have better health outcomes or greater satisfaction scores.

In the bottom portion of Table 12 we include additional quality measures that assess preventative measures. Since preventative measures are not disease-specific we compare this measure to the aggregate MCE_r, SPI_r and SUI_r measures. There is a strong positive relationship between the SUI_r and the preventative quality measure. For completeness, we also compare the MSA-level indexes to a composite of the four non-preventative measures listed in the top portion of the table. For the composite of non-preventative quality measures, there appears to be a significant correlation with SPI_r. This latter correlation may be indicative of a range of demographic and physician factors that may be correlated with both prices and quality of treatment at the MSA level.

Table 12. Correlation between Quality and Episode Indexes: MCE, SPI and SUI

	Log(MCE)	Log(SPI)	Log(SUI)
Log Quality Measure - Persistence of Beta-Blocker Treatment after a Heart Attack			
Hypertension	-0.192*	0.1393	-0.227**
p-value	0.0784	0.2036	0.0367
Ischemic Heart Disease	-0.1389	-0.0256	-0.1122
p-value	0.2048	0.8162	0.3067
Log Quality Measure - HbA1c Test for Pediatric Patients			
Diabetes	0.2224**	0.1921*	0.1809*
p-value	0.0408	0.0783	0.0975
Log Quality Measure - Complete Lipid Profile for Patients 18 years and older with Ischemic Vascular Disease			
Ischemic Heart Disease	0.1212	0.27**	-0.1572
p-value	0.2692	0.0124	0.1509
Log Quality Measure - Those with Back Pain Not Reporting an MRI within first 6 Months			
Joint Degeneration - Back	0.0687	0.1919*	-0.0314
p-value	0.532	0.0784	0.7752
Log Preventative Measures - Composite (Summation of Two Indexes of Preventative Care)			
Aggregate Indexes	0.1712	-0.1205	0.471**
p-value	0.1173	0.2718	0
Log Non-Prev. Quality Measures - Composite (Summation of the Four Quality Indexes Listed Above)			
Aggregate Indexes	0.3052**	0.3536**	0.0226
p-value	0.0045	0.0009	0.8372
Log Quality Measure - Composite (Summation of the Four Quality Indexes and the Two Preventative Care Indexes)			
Aggregate Indexes	0.2323**	0.0249	0.3969**
p-value	0.0324	0.8211	0.0002

* 90% signficance level ** 95% signficance level

Notes. For the disease measures reported here, we aggregate over diseases by defining the diseases as the ETG Category, rather than the ETG-severity combination. Similar results are found if one defines deseases based on the ETG-severity combination.

The limited correlations between quality and measures of price and utilization at the disease level suggest no clear pattern, but the strong positive relationships between spending and quality at the aggregate level is an interesting finding that warrants further exploration. The positive relationship between spending and quality at the aggregate level differs from what has been found looking at Medicare markets that show little relationship between overall spending and quality. This result also differs from the recent work by Turbyville et al. (2011), which shows very little relationship between utilization and quality for commercial markets. Although this result is interesting, it is important to keep in mind a number of limitations. First, we only look at simple correlations and there may be other explanatory factors affecting these relationships. Second, we have just a handful of quality measures that may not accurately reflect the true quality in the market. Third, these measures do not look at outcomes, which would be preferable measures of quality. Therefore, much more work needs to be done to determine if these correlations are causal.

10 Conclusion

Unlike Medicare markets where the service prices play a limited role in explaining variation in expenditures across markets, the variation in service prices in the commercial sector appear to be as important as utilization. Focusing on variation on a disease-by-disease basis, we find that the coefficient of variation for the typical disease to be around 0.22. Although there is large variation in both service prices and service utilization measures across diseases between markets, aggregate differences in the typical expenditure for disease treatment are not that different from aggregate price indexes observed for other goods and services, with a coefficient of variation of around 0.10.

The observed variation in spending across markets suggests the possibility of inefficiencies in some markets and potential savings. Several exercises are conducted to measure the economic importance of this variation. We find that the potential savings from controlling utilization on a disease-by-disease basis may have a substantial effect on overall expenditures. For example, those areas with the highest quartile of utilization for a disease may find savings of around 30 percent by adopting the utilization practices from the lower utilization areas. This implies large potential gains from controlling utilization, perhaps through bundled payments. However, differences in provider practice patterns across areas may not be the only cause of lower utilization. Sensitivity to market price may also play an important role, since we find a negative correlation between utilization and prices across markets, so lower utilization may partly be driven by higher service prices.

Although the observed variation in spending across markets suggests that there are potential savings from adopting the pricing or utilization patterns in low-spending areas, it is possible that greater utilization or service prices may be indicative of higher quality in commercial markets. In this study we find mixed evidence regarding measures of quality in each of the three indexes. Although measures of quality and the associated spending on related diseases is unclear, we find a strong positive relationship with aggregate measures of spending and composite measures of quality. The positive correlation between the overall utilization measure and quality measures of preventative care is particularly strong, but the underlying cause of this relationship is unclear.

There are some additional areas where more research may be beneficial. First, preliminary findings from this study show a negative and significant relationship between the SUI and the SPI measures. This negative relationship between service price and service utilization appears to be quite strong in our analysis and robust to alternative specifications and IV techniques. However, more work should be done to study if this negative correlation is spurious or whether it signifies something real such as a demand relationship (e.g. Dunn (2011)). Second, it is unclear how closely these expenditures are actually linked to the amount consumers pay for their overall medical care. Profit margins and administrative costs of insurers may drive a wedge between the full price of services studied here (that is, the full amount paid to the provider by the insurer and the patient) and the out-of-pocket costs to consumers (that is, the premium and the out-of-pocket costs). Third, the study between these indexes and quality may be greatly improved. For instance, future work may benefit from studying outcome measures of quality, rather than the process measures used in this paper, and more work needs to be done to uncover the source of the positive correlations between spending and quality observed in this study. Fourth, this paper characterizes MCE, SPI, and SUI differences across markets, but does not attempt to explain underlying reasons for these differences. Future work may benefit by looking for explanatory causes for the observed variation across markets (e.g. Dunn and Shapiro (2011)). While there are a number of areas to extend our research, we believe the framework presented here to define episode expenditures, service prices, and service utilization may be valuable for analyzing a variety of related topics in the future.

References

[1] 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

[2] Anderson, Gerard and Peter Sotir Hussey. 2001. "Comparing Health System Performance in OECD Countries." *Health Affairs*.

[3] 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.

[4] Berndt E.R., Cutler D.M., Frank R.G., Newhouse J.P., Triplett J.E. 2000. "Medical Care Prices and Output", in: A. Culyer and J. P. Newhouse (Eds.), Handbook of Health Economics, Elservier, Amsterdam. pp. 119-180.

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

[6] 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.

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

[8] 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.

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

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

[11] Dunn, Abe, Eli Liebman, Sarah Pack, and Adam Shapiro. 2010. "Medical Care Price Indexes for Patients with Employer-Provided Insurance: Nationally-Representative Estimates from MarketScan Data." Bureau of Economic Analysis, *Working Paper.*

[12] Dunn, Abe. 2011. "The Demand for Medical Care: Measuring Price Sensitivity from Cross Regional Data." *Working Paper.*

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

[14] 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.

[15] 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.

[16] Folland and Stano. 1989. "Sources of Small Area Variation in the Use of Medical Care." *Journal of Health Economics*, pgs 85-107.

[17] 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.

[18] 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.

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

[20] Medicare Payment Advisory Commission (MedPac). 2009. Report to the Congress: Measuring Regional Variation in Service Use.

[21] Newhouse, Joe P., and the Insurance Experiment Group, (1993), *Free For All? Lessons from the Health Insurance Experiment*, Harvard University Press, Cambridge.

[22] Phelps, Charles E. 1995. "Welfare Loss from Variations: Further Considerations." *Journal of Health Economics* 14, pgs 253-260.

[23] Phelps, Charles E. 1997. "Good Technologies Gone Bad: How and Why the Cost-Effectiveness of Medical Intervention Changes for Different Populations." *Medical Decision Making*, 17 pgs 107-117.

[24] Shapiro I.L., Shapiro M., Wilcox D. 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. pp. 411-438.

[25] 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: 45-53.

[26] Turbyville, Sally, Meredith Rosenthal, L. Gregory Pawlson, and Sarah Scholle. 2011. "Health Plan Resource Use - Bringing Us Closer to Value-Based Decisions." *The American Journal of Managed Care*, 17(1) pgs e1-e7.

[27] Welch, Gilbert, Sandra Sharp, Dan Gottlieb, Jonathan Skinner, and John Wennberg. 2011. "Geographic Variation in Diagnostic Frequency and Risk of Death Among Medicare Beneficiaries." *Journal of the American Medical Association*, 305(11) pgs 1113-1118

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

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

11 Appendix

11.1 Service-Specific Indexes

Table A7.1 MCE Index Across Service Types - $MCE_{r,s}$

	MSA	MCE_r	Inpatient Hospital	Outpatient Hospital	Office - General MD	Other	Pharmacy
1	Akron, OH	0.917	0.838	1.230	0.741	0.727	0.980
2	Atlanta-Sandy Springs-Marietta, GA	0.944	0.844	0.879	0.843	1.000	0.930
3	Augusta-Richmond County, GA-SC	0.865	0.870	0.916	0.654	0.953	0.787
4	MSA in Texas	1.033	1.100	0.783	1.027	1.126	1.043
5	MSA in the Northeast	0.848	0.947	0.679	0.705	0.693	1.048
6	MSA in the South	0.876	0.797	0.646	0.755	0.845	1.088
7	Birmingham-Hoover, AL	1.016	1.271	0.858	0.701	1.257	1.040
8	Boston-Quincy, MA	1.069	1.012	1.587	0.751	0.928	0.889
9	Bridgeport-Stamford-Norwalk, CT	1.049	0.961	1.051	0.936	1.019	0.968
10	Cambridge-Newton-Framingham, MA	1.049	0.967	1.527	0.768	0.875	0.898
76	Tampa-St. Petersburg-Clearwater, FL	0.929	0.908	0.808	0.896	0.927	1.038
77	MSA in Ohio	0.990	0.909	1.215	0.800	0.912	1.154
78	Tulsa, OK	0.987	1.006	1.029	1.088	0.727	1.018
79	Virginia Beach-Norfolk-Newport News, VA-	1.017	0.940	1.061	0.897	1.021	1.034
80	Warren-Farmington Hills-Troy, MI	0.876	0.746	0.901	0.998	0.697	1.018
81	MSA in the South	0.949	0.878	0.967	0.848	0.780	1.026
82	Vest Palm Beach-Boca Raton-Boynton Beac	0.976	0.987	0.862	0.890	0.917	1.040
83	Wichita, KS	0.949	0.594	0.745	1.867	1.051	1.090
84	Wilmington, DE-MD-NJ	0.969	1.004	0.834	0.841	1.009	1.122
85	Youngstown-Warren-Boardman, OH-PA	0.797	0.690	0.851	0.696	0.825	0.925
	COV	0.098	0.215	0.247	0.246	0.212	0.077

Table A7.2 SPI Index Across Service Types - $SPI_{r,s}$

	MSA	SPI_r	Inpatient Hospital	Outpatient Hospital	Office - General MD	Other	Pharmacy
1	Salinas, CA	1.406	2.246	1.528	0.934	1.250	1.028
2	Milwaukee-Waukesha-West Allis, WI	1.273	1.314	1.352	1.385	1.135	1.033
3	MSA in California	1.271	1.484	1.406	1.128	1.309	1.045
4	Oakland-Fremont-Hayward, CA	1.265	1.528	1.286	1.116	1.426	1.035
5	San Diego-Carlsbad-San Marcos, CA	1.164	1.315	1.235	0.966	1.259	1.038
6	Wichita, KS	1.153	1.102	1.395	1.028	0.895	0.983
7	MSA in the Midwest	1.149	1.087	1.196	1.153	1.118	1.068
8	Minneapolis-St. Paul-Bloomington, MN-WI	1.143	1.226	1.059	1.303	1.032	0.980
9	Dallas-Plano-Irving, TX	1.124	1.006	1.288	1.070	1.125	1.016
10	Peoria, IL	1.120	0.882	1.280	1.154	1.235	0.852
76	Cleveland-Elyria-Mentor, OH	0.921	0.787	0.844	0.927	0.872	1.071
77	MSA in Pennsylvania	0.915	0.806	0.715	1.117	1.018	0.994
78	Warren-Farmington Hills-Troy, MI	0.912	0.787	0.712	0.986	0.887	1.094
79	Providence-New Bedford-Fall River, RI-MA	0.897	0.980	0.633	0.989	0.898	0.952
80	Detroit-Livonia-Dearborn, MI	0.896	0.775	0.710	0.984	0.872	1.048
81	MSA in Ohio	0.894	0.769	0.773	0.879	0.795	1.132
82	Akron, OH	0.888	0.786	0.792	0.884	0.803	1.073
83	Kingsport-Bristol-Bristol, TN-VA	0.887	0.816	0.791	1.010	0.837	0.870
84	Pittsburgh, PA	0.855	0.762	0.659	0.921	0.799	1.048
85	Youngstown-Warren-Boardman, OH-PA	0.834	0.692	0.740	0.826	0.670	1.097
	COV	0.095	0.206	0.206	0.119	0.155	0.066

Table A7.3 SUI Index Across Service Types - $SUI_{r,s}$

	MSA	SUI_r	Inpatient Hospital	Outpatient Hospital	Office - General MD	Office MD - Speciality	Other	Pharmacy
1	Cambridge-Newton-Framingham, MA	1.194	1.190	1.216	1.173	1.195	1.237	1.177
2	Boston-Quincy, MA	1.183	1.175	1.205	1.171	1.188	1.215	1.172
3	Gary, IN	1.127	1.136	1.156	1.103	1.131	1.137	1.108
4	MSA in Mississippi	1.122	1.147	1.175	1.105	1.111	1.150	1.066
5	MSA in Ohio	1.118	1.123	1.156	1.106	1.132	1.141	1.076
6	MSA in Pennsylvania	1.114	1.071	1.131	1.120	1.110	1.111	1.152
7	Pittsburgh, PA	1.090	1.106	1.134	1.071	1.076	1.113	1.057
8	MSA in the Midwest	1.073	1.104	1.082	1.072	1.072	1.076	1.057
9	Kalamazoo-Portage, MI	1.073	1.089	1.080	1.088	1.094	1.060	1.054
10	St. Louis, MO-IL	1.064	1.096	1.075	1.053	1.054	1.068	1.054
76	MSA in the South	0.920	0.897	0.890	0.949	0.937	0.889	0.978
77	MSA in the South	0.920	0.931	0.896	0.918	0.931	0.921	0.942
78	Atlanta-Sandy Springs-Marietta, GA	0.919	0.909	0.911	0.941	0.927	0.916	0.937
79	Las Vegas-Paradise, NV	0.918	0.919	0.912	0.946	0.928	0.908	0.927
80	New York-White Plains-Wayne, NY-NJ	0.911	0.913	0.868	0.930	0.909	0.935	0.949
81	MSA in California	0.902	0.887	0.901	0.910	0.901	0.916	0.919
82	Des Moines, IA	0.897	0.911	0.891	0.896	0.883	0.876	0.930
83	Nassau-Suffolk, NY	0.896	0.899	0.860	0.910	0.890	0.908	0.938
84	Salinas, CA	0.887	0.911	0.903	0.881	0.887	0.898	0.861
85	MSA in the South	0.875	0.885	0.860	0.873	0.851	0.849	0.930
	COV	0.064	0.073	0.078	0.059	0.066	0.071	0.059

Table A10.1 MSA Medical-Care Price Indexes andTreated Prevalence Index - Population-Based and Episode-Based Measures - SORTED BY THE EPISODE MEASURE - MCE_r

MSA	Rank $PREV_r$	$PREV_r$	Rank MCE_r	MCE_r	Rank MCE^{pop}_r	MCE^{pop}_r	Rank SUI_r	SUI_r	Rank SUI^{pop}_r	SUI^{pop}_r
Milwaukee-Waukesha-West Allis, WI	73	0.923	1	1.293	1	1.193	39	1.004	68	0.932
Salinas, CA	64	0.956	2	1.247	2	1.179	84	0.887	81	0.851
MSA in the Midwest	74	0.917	3	1.238	8	1.143	8	1.073	48	0.988
Oakland-Fremont-Hayward, CA	70	0.936	4	1.219	5	1.149	53	0.972	69	0.921
MSA in the Midwest	79	0.900	5	1.169	16	1.060	12	1.062	58	0.962
Minneapolis-St. Paul-Bloomington, MN-WI	84	0.863	6	1.160	39	0.996	26	1.018	80	0.867
Gary, IN	82	0.873	7	1.143	44	0.992	3	1.127	49	0.987
Fort Worth-Arlington, TX	31	1.011	8	1.142	3	1.168	22	1.034	24	1.052
Indianapolis, IN	43	0.994	9	1.142	6	1.144	27	1.018	30	1.021
MSA in California	50	0.980	10	1.132	10	1.110	81	0.902	77	0.888
Memphis, TN-MS-AR	51	0.979	76	0.897	81	0.881	74	0.926	71	0.913
Kingsport-Bristol-Bristol, TN-VA	15	1.079	77	0.887	60	0.964	32	1.013	12	1.095
Providence-New Bedford-Fall River, RI-MA	20	1.040	78	0.883	71	0.928	30	1.016	18	1.064
Warren-Farmington Hills-Troy, MI	3	1.180	79	0.876	25	1.035	51	0.979	4	1.155
MSA in the South	12	1.099	80	0.876	58	0.966	75	0.923	29	1.022
Detroit-Livonia-Dearborn, MI	11	1.107	81	0.867	65	0.951	44	0.993	13	1.093
MSA in Michigan	1	1.194	82	0.867	27	1.024	67	0.939	10	1.115
Augusta-Richmond County, GA-SC	9	1.126	83	0.865	47	0.984	71	0.930	20	1.057
MSA in the South	7	1.134	84	0.848	56	0.968	77	0.920	26	1.049
Youngstown-Warren-Boardman, OH-PA	4	1.163	85	0.797	70	0.928	55	0.969	8	1.130

Table A10.2 MSA Medical-Care Price Indexes andTreated Prevalence Index - Population-Based and Episode-Based Measures - SORTED BY THE POPULATION MEASURE - MCE^{pop}_r

MSA	Rank $PREV_r$	$PREV_r$	Rank MCE_r	MCE_r	Rank MCE^{pop}_r	MCE^{pop}_r	Rank SUI_r	SUI_r	Rank SUI^{pop}_r	SUI^{pop}_r
Milwaukee-Waukesha-West Allis, WI	73	0.923	1	1.293	1	1.193	39	1.004	68	0.932
Salinas, CA	64	0.956	2	1.247	2	1.179	84	0.887	81	0.851
Fort Worth-Arlington, TX	31	1.011	8	1.142	3	1.168	22	1.034	24	1.052
Birmingham-Hoover, AL	6	1.138	31	1.016	4	1.166	14	1.050	1	1.203
Oakland-Fremont-Hayward, CA	70	0.936	4	1.219	5	1.149	53	0.972	69	0.921
Indianapolis, IN	43	0.994	9	1.142	6	1.144	27	1.018	30	1.021
MSA in Alabama	2	1.190	48	0.973	7	1.143	52	0.977	5	1.147
MSA in the Midwest	74	0.917	3	1.238	8	1.143	8	1.073	48	0.988
Dallas-Plano-Irving, TX	44	0.993	12	1.129	9	1.132	31	1.016	31	1.017
MSA in California	50	0.980	10	1.132	10	1.110	81	0.902	77	0.888
Cleveland-Elyria-Mentor, OH	52	0.976	63	0.935	76	0.921	25	1.021	41	1.005
Phoenix-Mesa-Scottsdale, AZ	68	0.948	57	0.956	77	0.912	43	0.996	61	0.952
MSA in the South	41	0.998	74	0.899	78	0.910	85	0.875	79	0.883
MSA in the South	38	1.001	73	0.901	79	0.909	76	0.920	67	0.932
Wichita, KS	62	0.956	59	0.949	80	0.904	72	0.927	78	0.883
Memphis, TN-MS-AR	51	0.979	76	0.897	81	0.881	74	0.926	71	0.913
Riverside-San Bernardino-Ontario, CA	76	0.903	55	0.957	82	0.868	73	0.927	83	0.843
MSA in Mississippi	85	0.832	35	1.003	83	0.833	37	1.007	84	0.838
Las Vegas-Paradise, NV	77	0.902	70	0.907	84	0.827	79	0.918	85	0.836
Pittsburgh, PA	81	0.894	71	0.907	85	0.817	7	1.090	51	0.981

11.2 Quality Measures

We construct six quality measures from the claims data using methods outlined by National Committee for Quality Assurance (NCQA). We focus on quality measures that may be constructed from administrative claims data. These quality measures are described in greater detail in NCQA Measure Technical Specifications.[44]

1. *Persistence of Beta-Blocker Treatment after a Heart Attack* - The percentage of patients 18 years of age and older during the measurement year who were hospitalized and discharged alive from July 1 of the year prior to the measurement year to June 30 of the measurement year with a diagnosis of acute myocardial infarction (AMI) and who received persistent beta-blocker treatment for six months after discharge (See Page 38).

2. *HbA1c Test for Pediatric Patients* - Percentage of pediatric patients with diabetes who had an HbA1c test in a 12-month measurement period (See Page 27).

3. *Complete Lipid Profile* - Percentage of patients 18 years and older with ischemic vascular disease who had a complete lipid profile (See Page 45).

4. *Use of Imaging Studies for Low Back Pain* - The percentage of patients with a primary diagnosis of low back pain who did not have an imaging study (plain X-ray, MRI, CT scan) within 28 days of the diagnosis (Page 12).

5. *Cervical Cancer Screening* - The percentage of women 21–64 years of age who received one or more Pap tests to screen for cervical cancer. (within a 3 year period). (Page 94).

6. *Breast Cancer Screening* - The percentage of women 40–69 years of age who had a mammogram to screen for breast cancer. (within a 2 year period). (Page 92).

11.3 Robustness Checks

To check the robustness of the results presented in this paper, we generated the tables presented here under a number of alternative assumptions (Tables 2 through 7). The following is a list of robustness checks. Unless noted, no qualitative results changed.

1. Isolated 2006 and 2007 separately. The results are quite similar in each of the separate years. The key advantage from combining years is that we are able to use more observations from each MSA.

2. For selecting diseases to be included in the sample we alter the threshold for the number of episodes observed in the data. Recall that the threshold applied here

[44] http://www.ncqa.org/tabid/59/Default.aspx

was 10,000. Similar results are attained if the threshold falls to 5,000 (accounting for 86 percent of spending) or is increased to 50,000 (accounting for 50 percent of spending). The problem with a lower threshold is that there will not be a sufficient number of episodes to attain an accurate price. In contrast, the problem with too high a threshold is that it accounts for a more limited fraction of overall spending.

3. Removed the population weights.

4. Adjusted the sample of cities by changing the threshold from 20,000 to 30,000 enrollees per MSA (dropping 24 cities).

5. Dropped the top and bottom 2.5 percent of episodes based on the episode expenditure price.

6. Aggregated diseases to the ETG category level and left out differences in severity.

7. Aggregated diseases to the Major Practice Category (MPC) level.

11.4 Prevalence and MSA Factors

As discussed in Section 7, the key difference between the episode-based measure and the population-based measure is the treated prevalence for each diseases. Therefore, it may be useful to examine the source of the variation for treated prevalence. Table A9.3 presents analysis that looks at how much variation in treated prevalence for each disease may be explained by MSA fixed effects, MSA-MPC fixed effects, and MSA-ETG Category fixed effects (identical to the analysis in Table 4). We find that MSA fixed effects explain relatively little of the variation in prevalence across areas, with an R^2 of 0.19. Including MSA-Major Disease Category fixed effects more than doubles the R^2 measure (and adjusted R-squared measure), indicating that factors that affect the prevalence of a disease in an area tend to be specific to the type of disease.

Table A10.3 Decomposing the Sources of Treated Prevalence Variation

Log(Treated Prevalence)

	R^2	Adj R^2	MSE	N
MSA-FE	0.1924	0.1767	0.24635	11305
MSA-MPC-FE	0.4781	0.3997	0.21036	11305
MSA-Disease-FE	0.8212	0.6916	0.15077	11305

Notes. Based on a regressions on log(PREV) for those diseases that have more than one severity and includes disease-severity fixed effects. Similar results are found w hen one includes all diseases and compares the fit of the model w ith MSA fixed effects to the fit w ith MPC-MSA fixed effects.