NOTE:

These materials were prepared by subcontractors for consideration by the *Committee on Geographic Variation in Health Care Spending and Promotion of High-value Care*. These analyses were commissioned and overseen by the Committee. However, the findings and views expressed in the subcontractor reports do not necessarily reflect those of the NRC/IOM or the Committee. Neither the methodology nor the subcontractor reports have been subject to formal institutional review for the Interim Report. As the committee continues to review the findings from the analyses contained herein, we invite you to provide feedback on the content of these reports. Please note that any comments will be entered into the project's public access file, and will be available for public review.

Provide Feedback

Harvard Medical School Department of Health Care Policy

Geographic Variation in Health Care Spending, Utilization, and Quality among the Privately Insured

Final Technical Report
Presented to: The Institute of Medicine
Committee on Geographic Variation in Health Care Spending and Promotion of High-Value Care

August 29, 2012

Michael McKellar[a]
Mary Beth Landrum[a], Ph.D.
Teresa Gibson[a,b], Ph.D.
Bruce Landon[a,c], M.D., M.B.A.
Sivia Naimer[a], M.S.
Michael Chernew[a], Ph.D.

[a] Department of Health Care Policy, Harvard Medical School, Boston, MA
[b] Thomson Reuters Health care, Ann Arbor, MI
[c] Division of Primary Care and General Internal Medicine, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA

Abstract

Background

Researchers have documented substantial variation in practice patterns, utilization, spending, and quality for health care services across the United States for four decades. Whereas a great deal of work has explored geographic variation in the Medicare population, much less has explored the commercial market. Our analysis explores variation in commercial health care spending, utilization, price, and quality.

Data/ Sample

The primary data source is 2005-2010 Thomson Reuters MarketScan Commercial Claims and Encounters Database (MarketScan). MarketScan includes 113 million person-year observations over the three-year primary study period (2007-2009) submitted by both private health plans and employers. This data is supplemented with input cost index data from the Center for Medicare and Medicaid Services (CMS) and other sources (e.g., Census data).

Methods

We assess variation between Hospital Referral Regions (HRRs) using Ordinary Least Squares (OLS) regression to adjust for patient level differences in demographics, health status, and insurance characteristics. Our primary measure of HRR spending is the average of residuals within each HRR from the OLS regression, adjusted to reflect the sampling variation associated with sample size within HRRs. As a sensitivity analysis, we assess the variation among a smaller geographic unit (the Hospital Service Area (HSA)), and examine the variation within HRR. The primary analysis is based on spending adjusted for differences in the cost of inputs such as wages across areas, though we also assess variation in total spending and in a variety of measures of quantity (i.e., utilization). This allows us to assess how much variation in spending is due to variation in prices versus quantities. Finally, using similar methods, we assess the variation in selected measures of quality across the HRRs and examine spending for 15 cohorts of individuals with selected medical conditions.

Results

There is considerable variation in spending between HRRs. The standard deviation in spending is $450 per enrollee per year, while spending at the 75th percentile HRR is 18 percent greater than spending at the 25th percentile HRR. This gap in spending is insensitive to inclusion of covariates, although health status, age, and sex explain a substantial proportion of the HRR variation in utilization. Using the preferred model (adjusting for age, gender, and health status) the standard deviation in spending drops from $447 to $403 (10%) when we adjust for input prices, and falls to $284 (30%) when we hold prices constant. We also observed substantial variation in price across markets (coefficient of variation in price is 0.19), even after adjusting for age, sex, and health status. Moreover, although variation is price is considerably greater than variation in quantity, there is a strong inverse correlation between price and quantity. We observed slightly higher variation across HSAs and MSAs and within HRR variation is moderate. The standard deviation in HSA spending within HRRs is $236 (which represents 28% of total variation at the HSA level), while the average ratio of the highest spending to lowest spending HSA within an HRR is around 1.3. The variation in quality is generally smaller than the variation in spending with the 75th percentile HRR generally less than 10% above that in the 25th percentile, though for some quality measures the range of variation is greater. Finally, the correlation between quality and spending is generally weak.

Conclusion

Health care spending among commercially insured beneficiaries varies widely. We observed substantial variation in both price and utilization across markets; however, price effects dominate after accounting for differences in health status. Moreover, because the variation in spending is not strongly related to quality measures, the benefit from greater spending is unclear.

Contents

Introduction 3

Background 3

Contribution of this Work and Research Questions 4

Methods 4

Overview 4

Data Sources 4

Geographic Units 6

Measurement of Spending 6

Measurement of Utilization 8

Aggregate Measures 8

Specific Utilization Measures 8

Measurement of Price 9

Implied Measure 10

Laspeyres Price Index 10

Adjusting for Input Prices 10

Partial Year Enrollment 11

Covariates 11

Covariate Models 13

Measuring Variation 14

Calculating Within-HRR Variation 15

Decomposition of Variation in Spending into Price and Quantity 15

Decomposition of Variation in Price into Input Price and Markup 16

Variation in the Likelihood of no-use and Spending Dependent on Use 16

Comparing Disease Cohorts 16

Quality Measures 18

Individual Measures 18

Quality Aggregation 22

Variation in Market Level Variables 23

Processing and Quality Assurance 25

Results 25
Sample Characteristics 25
Correlation 28
Variation in Spending 289
Stability over Time 37
Aggregate Variation in Utilization 38
Aggregate Variation in Price 42
Consistency across Cohorts 43
Consistency across Geographic Units 46
Within-Geography Variation 46
Quality Analysis 47
Market Level Analysis 59
Conclusion 64
Limitations 65
Future Research 65
Acknowledgements 67
Appendix 68
Sensitivity Analysis 1: Not Apportioning Claims that are Missing Procedure Codes 69
Sensitivity Analysis 2: Dropping Enrollees with Capitated Claims 70
Sensitivity Analysis 3: Omitting Enrollees without Drug Data 70
Sensitivity Analysis 4: Differences between Data Providers 71
Sensitivity Analysis 5: Shrinking Estimates 72
Supplemental Analysis 1: HCUP 72
Bibliography 74

Introduction

Researchers have documented substantial variation in practice patterns, utilization, spending, and quality for health care services across the United States for four decades. Most of this variation cannot be explained by the underlying sickness or preferences of the population, indicating that substantial improvements in quality and efficiency are possible. The research presented here attempts to shed light on a number of issues related to geographic variation in the commercially insured population, with the purpose of informing current and future policy.

Background

Dartmouth's seminal work on geographic variation estimates that unadjusted Medicare spending varies by as much as 55% between the highest and lowest quintile regions in the last six months of life, a finding corroborated by the Medicare Payment Advisory Board and others.[1-3] Some variation is a reflection of the costs of providing a service (input costs) between areas. In the Medicare context, payment rates are based on a fee schedule that incorporates the differing input costs of providing care, and thus the observed variation in that population is largely a reflection of use. Use is dependent on a number of underlying factors, including the underlying disease burden (or health status) of the population in each area. To give a sense of the role of health status in explaining variation, MedPAC estimates that total variation between the top and bottom decile MSAs declines roughly 20 percentage points when health status is taken into account. However, even after controlling for health status, there remains a 30% difference in spending between the top and bottom decile regions.

The extent of remaining unexplained variation, as well as substantial research, strongly suggests that practice patterns differ across areas, and that these differences drive a portion of total variation. Wennberg and Gittelsohn's research dating back to the 1970s documents substantial differences in practice patterns for similar patients within Vermont, indicating that practice patterns may differ even within relatively small areas.[4] More recent research has confirmed variation exists in the treatment of specific procedures such as rates of Cesarean deliveries, antibiotic prescriptions, hospitalizations, and types of dialysis access.[5-9] In addition, evidence suggests that higher spending regions do not receive better quality of care.[2,10]

Whereas a great deal of research has explored geographic variation in the Medicare population, much less has been performed to evaluate variation in the commercial market. Although Medicare represents a disproportionate share of health care expenditure per capita compared to younger populations, the private insurance market represents around 45 percent of United States health insurance expenditure.[11] In addition, several factors suggest that the variation in the commercial market may differ substantially from Medicare. For instance, the most prevalent or costly conditions in the under-65 population differ substantially from Medicare beneficiaries, who often face several chronic conditions. It is feasible that providers in some areas are relatively more efficient at providing care for the chronic diseases that most often afflict the elderly, but are relatively less efficient at providing care for acute episodes. This differential efficiency could mean the highest spending Medicare regions are not necessarily the highest spending commercial regions. In addition, commercial insurers do not set prices based on a standardized administrative fee schedule. Commercial insurers negotiate prices with individual providers or provider networks, allowing greater potential divergence in price even within the same region. Only recently has any research documented the extent and nature of variation among a national sample of commercially insured, non-elderly Americans. This work suggests that the extent of variation is similar and perhaps greater among the commercially insured, and areas with relatively high Medicare spending do not necessarily have high commercial spending.[12]

Contribution of this Work and Research Questions

Our analysis explores many of the underlying factors previously hypothesized to underlie variation in commercial health care spending, utilization, and quality. In addition, we test aspects of variation that go well beyond the scope of most published work. We begin by evaluating the role of price, quantity, and underlying input price in explaining spending variation. We then look at the extent to which a number of potentially explanatory variables, such as health status, explain variation and explore the sensitivity of these results to different levels of geography, the consistency of variation over time, and the type of service provided (i.e. inpatient or outpatient services). We also test how variation differs across the treatment of enrollees with specific diseases. We conclude by assessing the extent of variation in several quality indicators and their relationship with spending.

Methods

Overview

Our general approach is to estimate geographic variation using ordinary least squares regression models. We explore the extent and consistency of variation using a variety of approaches, comparing several covariate groupings (referred to as *clusters*), dependent variables (e.g. medical spending versus drug spending), and population groups (e.g. enrollees with diabetes versus enrollees with lower back pain). In most cases, we test the sensitivity of results by correlating area-level estimates, although we also present data on how the magnitude of variation differs across specifications.

Data Sources

Thomson Reuters MarketScan Database

The primary data source is 2005-2010 Thomson Reuters MarketScan Commercial Claims and Encounters Database (MarketScan). MarketScan includes 113 million person-year observations over the three-year primary study period (2007-2009) submitted by both private health plans and employers. The data captures employees and dependent family members covered under the same plan. From 2007-2009, MarketScan represented $450 billion in total expenditure, including all inpatient and outpatient claims incurred by each person, as well as drug claim information for around 80% of the sample. Claim information includes both the date of service as well as the location of service (inpatient facility setting, inpatient professional setting, outpatient setting, or pharmaceutical), procedure rendered and an indicator for whether the claim was paid through a capitated contract. Inpatient facility claims are assigned Medicare-equivalent diagnosis-related groups (DRGs) using DRG Grouper software package released annually. Inpatient professional claims and outpatient claims are assigned CPT procedure codes (we lack the modifier codes), and drug claims are assigned 11-digit NDC identifiers. In addition to health care claims, the database contains enrollee demographic information including age, sex, zip code of residence, type of insurance plan (HMO, PPO, etc.), out-of-pocket spending, as well as monthly enrollment fields that identify the portion of the year each enrollee is covered.

Comparisons of under-65 population in the MarketScan data with estimates of the U.S. population from the Medical Expenditure Panel Survey (MEPS) and Kaiser Family Foundation State Health Facts reveal that the distribution of age and sex of the enrollees in the MarketScan data is similar to that in the overall population with employer sponsored coverage. For example, in the 2006 MarketScan database, 51.5% of enrollees were female, whereas the MEPS 2006 estimated that 50.4% of beneficiaries with employer sponsored insurance were female. Likewise, enrollees aged 19 to 64 accounted for 71.4% of the MarketScan under-65 population, compared to 72.7% in the Kaiser Family Foundation State Health Facts estimates for the under-65 population with employer sponsored insurance in 2007.

United States Census Bureau: Demographic data

MarketScan does not contain enrollee-level data for race or family income. These data are added at the zip code level using information from the 2000 and 2010 Decennial Censuses for median family income and race, respectively. Due to confidentiality restrictions, Thomson Reuters could not provide precise zip code level data so instead each measure is aggregated by zip codes into quartiles weighted by the proportion of the population falling into each group.

At the time of study, 2010 Decennial Census data was not yet available at the 5 digit zip code level for median family income. Instead, 2000 Decennial Census data at the 5-digit zip code level are inflated to 2010 using trends in the Census county level data. Calculation of the county growth rate required using the American Community Survey (ACS) three-year estimates for 2010 data, which do not exist for all counties. For the 41% of counties without 2010 data, we use the average growth rate for all counties that had complete data.

Center for Medicare and Medicaid Services (CMS) Wage Index and Relative Value Unit files

We adjust spending for variation in input prices across different markets. These adjustments are made using the input price indices used to calculate Medicare prices. Specifically, CMS releases annual public-use hospital wage-index files that we use to approximate input costs of inpatient facility services in an area. CMS sets these wage-index values at the Core Based Statistical Area (CBSA)-level or larger, with many states receiving only one wage value. We use each yearly file in our study period (2007-2009) to estimate input prices. In addition to the hospital wage-index, we utilize CMS's public-use PPRVU datasets to estimate the input price of providing outpatient and inpatient professional services in an area. These data contain both CBSA-specific Geographic Price Cost Index (GPCI) values as well as procedure-specific relative value units (RVUs). We utilize the last release of these data in each calendar year for our study period.

Area Resource File (ARF)

The ARF is a publically available database of health professional statistics, including data from the American Hospital Association, the United States Census Bureau, and the American Medical Association. Data includes the number and specialty of physicians, the number of hospitals, the number of hospital beds, and population estimates by county.

American Hospital Association (AHA) Annual Survey Database

The AHA contains hospital-level statistics on the type of hospital, the number of hospital beds, teaching status, government affiliation, and specialty hospital designation.

HealthLeaders Interstudy (Interstudy)
Interstudy is an insurance industry survey containing county-level estimates of private and public insurance plan types. Specifically, it contains the proportion of the population covered by private HMOs, PPOs, and POSs, as well as the number of Medicare beneficiaries, Medicaid beneficiaries, and the uninsured population.

Health Resources and Services Administration (HRSA) Health Professional Shortage Areas (HPSAs)
The HRSA identifies areas and populations that are considered underserved based on access to medical services. For the purposes of this report, we focus on only area designations.

Geographic Units

Our primary geographic unit is a Hospital Referral Region (HRR), developed by Dartmouth. HRRs are commonly used in geographic variation literature, and are constructed based on the referral patterns for cardiovascular and neurosurgical specialty procedures.[1,2,13] HRRs are assigned to enrollees based on the zip code of residence. The MarketScan database captures claims for individuals in each of the 306 HRRs. In addition to HRRs, we also explore variation at the Hospital Service Area (HSA) level as well as the Metropolitan Statistical Area (MSA). HSAs are localized hospital markets that contain one or more hospitals that account for the majority of Medicare hospitalizations for assigned zip codes. HSAs are perfectly nested in HRRs, which gives rise to a within-HRR analysis described below. MarketScan contains data on enrollees in all 3,436 possible HSAs, but due to the small size of HSAs compared to HRRs or MSAs, enrollees are much more likely to leave their resident HSA to receive care.

Unlike HRR and HSAs, MSAs are not defined by health care resources and referral patterns, but are based on population centers. For rural counties not included in a Metropolitan (or Micropolitan) Statistical Area, we assign the state of residence. This results in coverage in all of the possible 441 MSA/ state designations.

Measurement of Spending

The data capture all inpatient and outpatient spending for an enrollee, including payments by employers, employee cost-sharing, and any coordination of benefit payments from other insurance coverage. Claims are fully adjudicated and reflect the amount reimbursed to the provider of services. Observed spending is based on transaction prices for all services used including inpatient medical, outpatient medical, and outpatient prescription drug spending. *Total spending* refers to the aggregation of all three components, whereas *medical spending* omits drug spending. All spending is inflation-adjusted to 2009 dollars using the Bureau of Labor Statistics GDP Deflator.

Multiple observations per claim
Services are often billed by providers in parts so that one service may result in multiple observations in administrative data. In the vast majority of instances, these appear to be the same service provided in the same setting or a correction to a paid amount. For this reason, claims of the same procedure code for the same person that occur on the same day are collapsed to one observation. We label this a "claims day". Because we do not observe procedure modifier codes, spending on related services (e.g., assistant surgeons) occurring on the same day will be attributed to the procedure and thus counted as "price" as opposed to quantity. Similarly, if there happen to be multiple services of the same type

delivered to the same person, on the same day (e.g. office visits of the same complexity, physical therapy, etc.) these will all be counted as one claims day and the added spending would be treated as an increase in price, not quantity.

Excluded data

Any enrollee who is observed with a negative payment observation after aggregating to the claim-day is excluded from the analysis for that year. It is impossible to tell whether these claims are corrections meant to apply on a different date, or if there has been a data entry error. This exclusion reduces the total sample by less than 0.5 percent over the three year period. In addition, we exclude enrollees who are covered through Medicare (through long-term disability) or who have missing enrollment data, which represents another 0.5 percent reduction in sample.

Outpatient claims missing procedure codes

A sizable number of outpatient services performed in a facility are not assigned CPT codes. These claims often have special revenue codes that denote broad service categories such as "imaging," but are not assigned a more detailed CPT code. We assign spending (s) associated with claims without procedure codes (m), to other CPT codes (c) based on the outpatient procedures billed by physicians on the same day.

When no outpatient claims for professional services with procedure codes are present, the spending amount is not changed. This spending is still included in the analyses, but is not adjusted for input price differences (i.e. not adjusted for underlying cost of providing services in an area) and is treated as if the input price is one. This spending also enters the aggregate quantity measure (a measure of spending with prices held constant) as observed. We find almost identical results when not distributing missing procedure code spending in this way (see Sensitivity Analysis 1 in the Appendix).

Imputing spending for capitated claims

Spending for capitated claims is reported in MarketScan in various ways by employers and health plans. On average, approximately 6% of individuals are enrolled in capitated products, and the range per HRR is 0% to 55% of enrollees, with a median of <1%. Some claims are entered as an estimated FFS equivalent, while others are assigned the lump sum capitated amount with all subsequent claims associated with zero dollar payment amounts. In other cases it is possible that only the cost sharing amount is entered. We cannot determine which method is used and do not believe the dollar amounts listed for the capitated claims consistently recorded or reflect a reimbursed amount for the service. Moreover, we do not observe the capitation payment. To address this inconsistency, we drop the dollar amount for all claims paid by capitation and replace the spending with an imputed amount. Our imputation strategy sets the price for each observed capitated claim (p) equal to the national average price for the service ($\hat{p}$), multiplied by the HRR-specific price index (i).

This imputation is performed across all capitated claims including those with zero dollar associated payments. Due to a small number of regions with high capitation, this method generates slightly different results than if all capitated enrollees are dropped, especially when comparing the ranks of regions that have relatively high levels of capitation (see Sensitivity Analysis 2 in the Appendix).

Missing Drug Data

Due to contractual arrangements, around 20% of enrollees with medical (non-drug) data do not have drug data. For the remaining 80% we are able to assess drug spending even if spending is zero. Yet, we cannot determine if the people with no drug data have unreported drug coverage from another source (which we suspect) or lack drug coverage. To maximize our sample size, we estimate our medical spending models on the full sample and run separate models of drug spending on the subset of patients with drug data. We standardize area-specific estimates of drug spending to the population with medical spending data by centering all covariates with the mean in the medical sample so that spending estimates are comparable. To estimate total spending we add the per capita adjusted HRR (HSA, MSA) drug spending estimates to the per capita adjusted HRR (HSA, MSA) medical spending estimates. Sensitivity analysis suggests that input price adjusted medical spending for the 80 percent with drug data is substantially higher than for the 20 percent without drug data, although the extent of HRR variation is similar and relatively well correlated (see Sensitivity Analysis 3 in the Appendix).

Measurement of Utilization

Unlike spending, which can be observed by summing the payments associated with all claims, obtaining an aggregate utilization measure across different types of procedures involves an explicit weighting mechanism because units of service are not constant. Our primary approach is to create an aggregate measure of utilization that is weighted based on the average spending for each service, although we recognize that any inaccuracy in the weights will have an impact on our conclusions. We validate the use of this aggregate utilization approach by comparing it against several sentinel measures such as office visits and imaging procedures, as well as a different aggregate approach based on the HRR-specific price index.

Aggregate Measures

For inpatient and outpatient claims, we develop an aggregate quantity measure index by holding price constant based on the national average price for each procedure. Measures per person per year, quantity (q) is equal to the sum of all procedures (i) multiplied by the procedure's mean national price ($\hat{p}$).

For drug claims, we standardize quantity based on the number of days supplied in each prescription. For supplies of 30 days or less, the quantity is standardized to one. For supplies of more than 30 days, quantity is standardized to a 30 day equivalents. For instance, a 90 day supply is equivalent to three, whereas a seven day supply is equivalent to one. This method accounts for the different nature of acute medications and chronic medications.

In addition to the above approach, we test the sensitivity of our results by calculating quantity as spending divided by price. Price in this instance is calculated using a Laspeyres price index, described below.

Specific Utilization Measures

In addition to the aggregate measure, we compute the counts of several sentinel services outlined in Table 1. Most of these services are counted for each enrollee analogous to the method for counting multiple observation lines that have the same procedure code. For inpatient days, we use a length of stay variable to estimate the total number of days and then classify it as either medical or surgical, depending on the type of service provided. Inpatient admissions are counted in the same way, but without regard to the length of stay. Diagnostic imaging procedures are only captured in outpatient claims and include a range of over 600 procedure codes that include X-rays, MRIs, ultrasounds, and CT scans as well as a number of other diagnostic tests. Emergency department visits are identified by the place of service (outpatient hospital or office) and one of five qualifying CPT codes.

Table 1. List of sentinel services

Total inpatient days
Total inpatient admissions
Inpatient surgical days
Inpatient surgical admissions
Inpatient medical days
Inpatient surgical admissions
Diagnostic imaging procedures
Emergency department visits

Table 1 lists the sentinel services selected for analysis.

In addition to analyzing the stability of the aggregate quality measure, we evaluate the extent to which variation in utilization is consistent across types of procedures. In addition to the sentinel services listed above, we look at utilization patterns for services considered potentially discretionary or that may vary substantially by practice style. Evidence suggests that utilization of some of these services is often predicated on the underlying physician workforce as opposed to patient characteristics.[7,14] Table 2 lists the services, which include hip and knee replacement, the number of laparoscopic versus open cholecystectomies, hysterectomies, nuclear and non-nuclear cardiac stress tests, as well as bilateral cardiac catheterizations. All are identified by corresponding procedure codes with no more than one of any type of discretionary service being counted on the same day.

Table 2. List of discretionary services

Hip Replacement
Knee Replacement
Laparoscopic versus Open Cholecystectomy
Hysterectomy
Nuclear cardiac stress test
Non-nuclear cardiac stress test
Bilateral Cardiac Catheterization

Table 2 lists potentially discretionary services selected for analysis.

Measurement of Price

The product of price and quantity determines total spending according to the formula:

$$spending = price * quantity$$

Holding spending constant, any estimation error in quantity will be reflected in a proportional estimation error in price, creating an artificial inverse correlation between price and quantity. With this in mind, we explore two methods of calculating price. Our primary approach is to use our estimates of spending and aggregate quantity to calculate implied price at the HRR level. We test the sensitivity of this approach against one where we estimate price more directly using a Laspeyres price index (which is limited by use of only a subset of services due to sample size limitations; see below for details). We examine the sensitivity of our estimated decomposition into the price and quantity components to our choice of method.

Implied Measure

Given that spending=price*quantity, our *implied* price measure is calculated as observed spending divided by observed quantity at the HRR level.

Laspeyres Price Index

As an alternative to the implied approach to measuring price described above, we develop a price index using a market basket of services defined yearly. The market basket consists of the top 100 inpatient services (defined as DRGs) in terms of total spending, top 100 outpatient professional services (designated by CPT codes), and any additional inpatient or outpatient professional services that fall within the top 200 services overall. Due to substantial revisions to codes during our study period, we calculate individual market baskets for each year separately. For each year, we use the market basket to calculate a Laspeyres price index for each HRR. The index value (d) for hrr (h) is equal to the sum of the hrr mean price ($\hat{p}_h$) of all market basket services (i) multiplied by the national weight for that service (w).

$$d_h = \sum_i \hat{p}_{h,i} * w_i$$

For any services in the market basket that do not occur in a given region, the national average price is imputed. The procedure weight is the proportion of the market basket represented by that service, in terms of total national spending (s).

The resulting index value is converted to a proportion of the national average, so that a value of 1 indicates that the prices in a region are exactly the national average, and a value of 1.1 indicates that prices are 10% higher than the national average.

Adjusting for Input Prices

For inpatient facility claims, input price is set equal to the Hospital Wage Index in an enrollee's county of residence. Following the CMS method for adjusting prices geographically, we adjust a portion of total spending by the wage index and add in the remaining unadjusted portion to estimate input-price adjusted spending for inpatient facility services. This proportion is the labor-related share published annually in the Hospital Inpatient Prospective Payment System Final Rule. In 2007, the labor-related share was 0.697, so our estimate of input price adjusted spending (s_p) is equal to the share multiplied by the spending on all inpatient facility services (s_f), divided by the wage index (w) plus the non-labor share (0.303) multiplied by spending on inpatient facility:

$$s_p = \left(0.697 * \frac{s_f}{w}\right) + (0.303 * s_f)$$

For outpatient services, as well as inpatient professional services, we combine the procedure-specific Relative Value Unit (RVU) weight with county estimates of the Geographic Practice Cost Index (GPCI). For a given service, CMS publishes RVUs broken into three components: 1) a practice expense RVU, 2) and work RVU, and 3) a malpractice RVU. Due to the fact that services can have differential input costs across areas, we estimate the relative share of each component of the RVU and multiply it by the equivalent GPCI weight. For instance, in 2007 the work, practice expense, and malpractice RVUs for a new-patient office visit of moderate complexity were 1.34, 0.43, and 0.09, respectively. Thus the relative share of each of these components was 0.72, 0.23, and 0.05. This final adjustment factor takes the sum of these shares, multiplied by the equivalent GCPI component. Input price adjusted spending for outpatient and inpatient facility services is equal to the sum of the observed spending on each claim, divided by that claim's specific adjustment factor. Some outpatient claims, such as anesthesia services, are not paid through the same RVU system in Medicare, and for these claims we divide the observed spending by the practice expense GPCI in the county. Drug spending and durable medical equipment spending are not input price adjusted, as we assume that the production costs of these products are not localized to the markets in which they are purchased.

Partial Year Enrollment

Due largely to employees joining and leaving firms, as well as firms joining or leaving the MarketScan database, we observe approximately 20% of people enrolled for less than 12 months in any year. We employ a three-part method to correct for partial year enrollment. We 1) annualize spending and utilization measures by dividing by the portion of the year enrolled, 2) include an independent variable indicating the portion of the year enrolled, and 3) weight observations in spending and utilization regressions by the portion of the year enrolled.

Covariates

Age and Sex

For the aggregate spending and utilization models, age (0-64) is broken into five-year age bands. Sex is designated male or female, and we include an interaction between sex and each five-year age band. For cohort analysis and quality analysis, age bands are defined to fit the population being studied.

Race/ ethnicity and Income

Race/ethnicity is reflected in five categories: 1) White (non-Hispanic), 2) Black (non-Hispanic), 3) Hispanic (of any race), 4) Asian, and 5) other (non-Hispanic). Each category is broken into quartiles, so that quartile one of the white category indicates that the zip code is in the top 25% of zip codes of whites as a proportion of the total population. Each quartile is coded with a dummy indicator, and enters the model with the fourth quartile in each category as the reference group.

Benefit Generosity

To measure the variation in benefit generosity across enrollees, we utilize health plan level average cost-sharing for five common services. Each measure encompasses the total out-of-pocket payment per claim day, aggregating the copayment, coinsurance, and deductible amount paid by the patient. These out of pocket (OOP) measures are averaged within a health plan offered by an employer to reflect the benefit structure faced by the patient. When plans within an employer were small (less than 100 enrollees) plans were combined to create more stable estimates. The five measures are included in Table 3.

Table 3: List of benefit generosity measures

Measure
OOP per prescription for brand name medications
OOP per prescription for generic medications
OOP for ER visits per ER visit
OOP for inpatient admissions per admission
OOP for office visits per office visit[a]

Table 3 contains a list of the five benefit generosity measures.

Health Status

We create a health status variable using Verisk Health DxCG software (version 3.10). The program uses the age, sex, and diagnoses codes in claims for each enrollee in the data and calculates the relative risk of that individual, with the mean risk of the population equal to 1.[b] To ameliorate the concern of endogeneity of a health status measure based on observed claims, we use an enrollee's previous-year claims to estimate a predicted risk for the current year. For enrollees without previous experience, we impute relative risk based on the enrollee's age and gender.

Data Source

The MarketScan data is contributed by either large employers or health plans. Sensitivity analysis indicates that these two samples are relatively well correlated (see Sensitivity Analysis 3).

Enrollment

Enrollment is measured as the percent of the year enrolled. We include an enrollment indicator in each model described below.

Insurance Plan Type

Plan types include: Health Maintenance Organization (HMO), Preferred Provider Organization (PPO), Point of Service (POS), Consumer Driven or High Deductible Health Plan (CDHDHP), and traditional indemnity.

---

[a] Ideally specialty and primary care office visit copayments would be separated. However, MarketScan does not reliably capture provider specialty information.

[b] Due to imputation of risk scores for enrollees without prior experience, who are given age and sex based scores that are generally much closer to one, the mean in our population is lower than one.

Covariate Models

Many factors affect spending and consumption of heath care services, some of which are easily observed (age and sex) and others of which are difficult to observe (patient preferences). We use seven different sets of independent variables to assess their relative ability to explain geographic variation in spending. The baseline model includes only year dummies and adjustments for partial year enrollment, while the other (referred to as covariate *clusters*) include different combinations of available independent variables. The definitions of these seven models are presented in Table 4.

Table 4. Model definitions, "Y" indicates inclusion.

	Baseline	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Partial Year Indicator	Y	Y	Y	Y	Y	Y	Y
Year Dummies	Y	Y	Y	Y	Y	Y	Y
Age Bands	N	Y	Y	Y	Y	Y	Y
Gender	N	Y	Y	Y	Y	Y	Y
Age/Gender Interaction	N	Y	Y	Y	Y	Y	Y
Health Status	N	N	Y	N	N	Y	N
Race/ Ethnicity Measures	N	N	N	Y	N	Y	N
Income	N	N	N	N	Y	Y	N
Benefit Generosity Measures	N	N	N	N	N	N	Y

Health Plan Type Indicators	N	N	N	N	N	N	Y
Data Source (Employer or Health Plan)	N	N	N	N	N	N	Y

Table 4 describes the seven covariate groups used in analysis.

Measuring Variation

As described above, we employ OLS regression models to estimate area-specific measures of spending and utilization. For the aggregate analyses, we estimate area-level spending (and utilization) by first fitting an OLS model without area-level fixed effects. We then estimate area-level effects as the average of the residuals from this model. This is almost identical to a fixed-effects model that includes dummy variables for each area.[c] We approximate standard errors for the area-level effects as the root mean square error from the regression model divided by the square root of number of enrollees in the area. Note this does not account for error in estimating regression coefficients and also does not account for multiple observations on enrollees and thus somewhat overstates true precision of the estimate. We summarize variation across areas by examining the distribution of these area-specific estimates (such as the standard deviation, the coefficient of variation, or ratio of upper and lower quartiles or deciles). In disease cohorts we use residuals from OLS models to estimate area-level effects and report analogous standard errors from the fixed-effects model.

Areas with a relatively small sample have less reliable estimates of spending because the small samples will make the estimate of the area fixed effect particularly susceptible to the idiosyncratic illness and spending of this group of patients. This possibility requires a strategy to correct estimates of the area fixed effects for noise in the estimates arising from sampling variation. A solution to the noise problem is to "shrink" the estimates using an empirical Bayes framework. These estimators, often referred to as shrinkage estimators because extreme values are pulled towards the overall mean, minimize prediction error and better account for regression to the mean effects than standard estimates. These estimates are also often referred to as reliability-adjusted estimates. Reliability-adjusted estimates η_j^{EB}, of an estimated area effect (Aj) is $\eta_j^{EB} = R_j A_j + (1-R_j)M$, where M is the national average and the degree to which an estimate is "shrunk" (R) is given by:

$$R_j = H/(H + \left(\frac{\theta}{n_j}\right))$$

Here, H is the variance of the area effects (i.e. the between area variance), θ is the within-area variance of the outcome measure, and n_j is the number of observations for HRR j. This shrinkage factor has an

[c] Our use of the residual approach compared to the fixed effect methods is purely for expedience. The computational complexity and computing time for running the fixed effects method was substantial, without any meaningful difference in estimates. In a sensitivity analysis (not reported), we find that HRR estimates in a one-year comparison of the residual method to a fixed effect method are correlated over 0.98.

intuitive interpretation of the "reliability" of each estimate of the HRR effect. HRRs with smaller reliability will have their estimated H_j shrunk towards the national mean – closer implicitly to that of other HRRs; for those with $R_j = 1$, there will be no shrinkage. The degree of shrinkage depends positively on the residual variance θ and negatively on the size of the each HRR (n_j). We use an empirical Bayes approach to compute the between and within area terms (sometimes referred to as the signal and noise variance). In practice, performing empirical Bayes estimation for linear mixed models is very straightforward as all the quantities required for shrinkage are estimated as part of a standard OLS regression. The noise variance (θ/n_j) for each area is equal to the mean square error from the OLS regression model divided by the sample size. We estimate the signal variance as the observed variance of the area-level rates adjusted for sampling variability by subtracting mean sampling error variance:

$$H = var(A_j) - mean(\vartheta/n_j)$$

Calculating Within-HRR Variation

It is possible that substantial variation exists even with HRRs. To test the extent of within-HRR variation, we regress HSA input price adjusted spending estimates on HRR dummies, weighted by the population in HSAs. The resulting r-squared value provides a measure of the variance in HSA spending explained by HRRs.

Decomposition of Variation in Spending into Price and Quantity

Total spending is dependent upon the quantity and price of services. We use the aggregate measures of price and quantity described above to calculate the relative proportion of spending variation that can be attributable to price and quantity. Any error in the calculation of quantity will be reflected by a proportional error in the calculation of price. For this reason, we employ three methods. The primary method is to use the aggregate quantity measure to calculate an implied price. As a first sensitivity, we use an analogous method except that instead of calculating quantity directly, we calculate price directly using the Laspeyres price index. We then use these estimates to calculate an implied quantity. Finally, we use both the aggregate quantity measure as well as the price index to calculate an implied spending value. In each case, the decomposition method is identical, but the method for calculating price, spending, and quantity varies.

The variance in spending can be decomposed using the following formula:

$$var[s] = \hat{q}^2 * var[p] + \hat{p}^2 * var[q] + higher\ order\ terms$$

The higher order terms include the covariance between quantity and price. Because the higher order terms can be negative, we report the relative share of variance attributable to quantity as:

$$relative\ share\ attibutable\ to\ quantity = \hat{p}^2 * var[q]/(\hat{p}^2 * var[q] + \hat{q}^2 * var[p])$$

Or, conversely:

$$relative\ share\ attributable\ to\ price = 1 - share\ attribuable\ to\ quantity$$

Decomposition of Variation in Price into Input Price and Markup

We further break down price into input price (i.e. the cost of producing a service) and markup, the price above input price. We calculate input price adjusted spending by adjusting observed spending by the Medicare Area Wage Index (for inpatient facility claims) and the Medicare GPCI (for inpatient professional and outpatient claims). Using this input price adjusted spending, we calculate input price as:

$$price_i = input\ price\ adjusted\ spending/quantity$$

Where *i* is input price. We then use input price to calculate markup:

$$price = price_i * markup$$

or

$$spending = price_i * markup * quantity$$

We further break down variation in price to variation in input price and variation in mark-up:

$$var[p] = \hat{\imath}^2 * var[m] + \widehat{m}^2 * var[i] + higher\ order\ terms$$

Variation in the Likelihood of no-use and Spending Dependent on Use

Variation in spending may be driven by differences in the likelihood of consuming any health care and differences in spending conditional on use. To test the extent to which differences in the extensive and intensive margins explain variation, we evaluate a linear probability model of non-use in combination with a model that includes only people with claims. Both models are adjusted for age, sex, and health status. We exploit the multiplicative nature of these models to determine the proportion of total variation explained by each component. Specifically, given that total spending in an area is equal to the average spending conditional on use multiplied by the likelihood of use.

Comparing Disease Cohorts

Even if substantial variation exists between geographic units, it is not certain that areas that have high spending overall have high spending across all conditions. For instance, areas with high spending on orthopedic conditions may not have high spending on cardiac conditions. We isolate 15 disease cohorts to explore the differences in variation for treating populations with various chronic and acute diseases. Unlike the aggregate analysis, the cohort analysis is predicated on use because only people with observed claims can be identified. Table 5 lists the 15 disease cohorts, the number of enrollees in each cohort, the episode length that an enrollee's claims are counted after the observation period, if a clean period was imposed, and whether or not gender is included in models.

Table 5: List of cohorts

Disease	Enrollees (thousands)	Episode Length (months)	Clean Period?	Include Gender Variable?
Stroke	62	12	Y	Y
Diabetes	3,648	12	N	Y
Pneumonia	221	3	Y	Y
Rheumatoid Arthritis	280	12	N	Y
Depression	2,195	12	N	Y
Congestive Heart Failure	275	12	N	Y
Acute Myocardial Infarction	81	12	Y	Y
Coronary Heart Disease	1,130	12	N	Y
Chronic Obstructive Pulmonary Disease	391	12	N	Y
Cataract Surgery	188	3	N	Y
Lower Back Pain	5,281	12	N	Y
Cholecystectomy	75	6	N	Y
Lung Cancer	13	12	Y	Y
Prostate Cancer	63	12	Y	N
Breast Cancer	42	12	Y	N

Table 5 lists the 15 disease cohorts used in our analysis, along with the number of episodes (in thousands), the length of the episode, whether a clean period was imposed, and whether or not models include a gender indicator.

For 13 of the 15 conditions, the episode length is a set time period after a qualifying event (such as a heart attack or initial diagnosis of breast cancer). Due to the lead-up procedures that occur before cholecystectomy, the episode is extended three months before the qualifying event. For cataract surgery, if a second qualifying event is observed within the three months following a qualifying event, the episode is extended for three months past the second event. For six of the cohorts (stroke, pneumonia, AMI, lung cancer, prostate cancer, and breast cancer) a clean period was imposed that restricted the eligible group to those without qualifying events in a prior period. The purpose of this is to capture claims following new episodes of these diseases, and not the ongoing care of patients who already had these conditions.

Due to the relatively low occurrence of several of these conditions in the under 65 population, we do not evaluate all cohorts at the MSA or HSA levels. Specifically, we only perform MSA analysis on the nine cohorts with over 100,000 enrollees and do not perform cohort analysis at the HSA level.

Quality Measures

We construct a series of quality measures to test the magnitude and stability of variation in quality of care. For each measure, we observe two indicators: whether an enrollee is eligible for the quality measure and whether the enrollee met the quality standard.

Eligibility Requirement
Partial year enrollees are deemed ineligible if their enrollment dates are insufficient to determine if they meet the quality measure.

Individual Measures

Our final measure set includes 31 individual measures: 12 process measures which were chosen from the National Committee for Quality Assurance (NCQA) HEDIS@2011 Health care Effectiveness Data and Information Set or was recommended by relevant specialty society; readmissions within 30 days of discharge (also based on HEDIS specifications); 12 Prevention Quality Indicators (PQIs) measuring potentially avoidable hospitalizations (these measures were developed by the Agency of Health care Quality and Research (AHRQ) and were designed to measured quality of ambulatory care); and seven Patient Safety Measures (PSIs) (these measures were also developed by the AHRQ to measure adverse events associated with medical errors). Note while AHRQ has also developed individual and composite measures for inpatient mortality and pediatric safety, we have not analyzed these measures because mortality and pediatric safety events are very rare in our commercial population. For example, we observed fewer than 900 total pediatric safety events each year compared to 15,000-20,000 in the adult population. Table 6 contains a list and description of all quality measures, as well as whether we risk adjust quality scores.

Many of our quality measures are based on an episode of care and enrollees can have multiple episodes in a given calendar year. In a few cases we observed extreme and unlikely number of episodes. We excluded a small number of enrollees with greater than four cataract surgeries (0.2%), and greater than five admissions (0.6%).

Enrollees can also have multiple preventable admissions (PQIs) in a year. While PQIs are typically measured as rates (number of avoidable admissions per 1,000 enrollees), multiple avoidable admissions were very rare in our population. For ease of interpretation, we collapsed all PQI indicators into a binary summary equal to one if the enrollee had one or more admissions of specific type.

Table 6: List and definitions of quality measures

Measure	Description	Risk Adjust?
Screening Mammography	% of women age 42-65 who had a mammogram to screen for breast cancer within last 2 years	N
Radiation therapy following breast conserving surgery for breast cancer	% receiving radiation therapy following lumpectomy	N
Treatment with DMARD for rheumatoid arthritis	% with a disease modifying anti-rheumatic drug	N
Treatment with bronchodilator for COPD	% receiving bronchodilator within 30 days of inpatient stay or ER visit	N
Treatment with antidepressant for major depression	% with antidepressant therapy for 84 days after new episode of major depression	N
Treatment with antiplatelet for stroke	% with antiplatelet medication on discharge after ischemic stroke or TIA	N
Treatment with beta blockers for AMI	% with beta blockers for 6 months following an AMI	N
Antibiotics for pneumonia	% appropriate empiric antibiotic for community-acquired bacterial pneumonia	N
Hemoglobin A1C for Diabetes	% patients with diabetes with hemoglobin A1C measurement in calendar year	N
Eye exam for Diabetes	% patients with diabetes with eye exam in calendar year	N
Appropriate treatment for low back pain	% with low back pain without an imaging study within 28 days of diagnosis	N
Lack of complications following cataracts surgery	% without postoperative complications within 30 days of cataract surgery requiring additional surgical procedures	N
Readmissions	% of patients readmitted to hospital within 30 days of discharge	Y

Diabetes short-term complications (PQI #01)	% with hospitalization for short-term diabetes complication	Y
Diabetes long-term complications (PQI #03)	% with hospitalization for long-term diabetes complication	Y
COPD in older adults (PQI #05)	% 40 and older with admission for COPD or asthma	Y
Asthma in younger adults (PQI #15)	% 18-40 with admission for COPD or asthma	Y
Hypertension (PQI #07)	% with admission for hypertension	Y
CHF (PQI #08)	% with hospitalization for CHF	Y
Dehydration (PQI #10)	% with hospitalization for dehydration	Y
Bacteria Pneumonia (PQI #11)	% with hospitalization for bacterial pneumonia	Y
Urinary Tract Infections (PQI #12)	% with hospitalization for urinary tract infection	Y
Angina (PQI #13)	% with hospitalization for angina without a procedure	Y
Uncontrolled diabetes (PQI #14)	% with hospitalization for uncontrolled diabetes	Y
Lower extremity amputation (PQI #16)	% with hospitalization for lower-extremity amputation	Y
Pressure ulcer (PSI #03)	% admissions with pressure ulcer	Y
Iatrogenic pneumothorax (PSI #06)	% admissions with Iatrogenic pneumothorax	Y
CVC-related blood stream infection (PSI #07)	% admissions with central venous catheter with CVC-related blood stream infection	Y

Postoperative hip fracture (PSI #08)	% surgical admissions with postoperative hip fracture	Y
Postoperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI #12)	% surgical admissions with Postoperative Pulmonary Embolism or Deep Vein Thrombosis	Y
Postoperative Wound Dehiscence Rate (PSI #14)	% surgical admissions with Accidental Puncture or Laceration	Y
Accidental Puncture or Laceration Rate (PSI #15)	% admissions with Accidental Puncture or Laceration	Y

Table 6 contains descriptions of the 32 quality indicators and whether or not they are risk adjusted.

Risk adjustment.
Many of our quality indicators measure processes applied to specific populations (for example breast cancer screening, anti-depressant therapy for major depression, etc.). As they have strict eligibility criteria, they do not require adjustment for patient characteristics. However, some indicators should account for differences in risk across areas (for example PQI measures and readmissions). For these measures we adjusted for age, gender, age/gender interactions and health status measured by a DxCG score. See Table 6 for identification of which individual quality measures will be risk adjusted using these variables.

We employ an OLS model analogous to that used in the spending and utilization analyses to adjust for patient characteristics, estimating the average area-level residual. However, some of our dependent variables are extremely rare (the PQI and PSI indicators) and linear models may result in negative HRR estimates. Thus for rare outcomes, we use a straight-forward adjustment method that has been applied frequently in provider profiling activities (this method is also recommended by AHRQ to risk adjust PSI indicators). Specifically, we fit a logistic regression model to individual outcomes, without HRR fixed effects, but with relevant patient risk adjustment variables. We then use the estimated regression coefficients and observed distribution of covariates to compute the expected number of events in each HRR (i.e., $exp_i = \Sigma\, p_{ij}$ where p_{ij} is the predicted probability of event for patient j in HRR i based on the fitted model). Then we create a risk-adjusted rate for each HRR equal to the observed number of events divided by the expected number of events, multiplied by the average rate of the indicator across the entire population:

$$\textit{Risk-adjusted rate}_i = (obs_i / exp_i) * \textit{reference rate}$$

where obs_i is the observed number of events in HRR_i, exp_i is the expected number estimated using the logistic regression and reference rate is the average rate in our population (i.e., total number of observed events divided by total population).

We compute a standard error for the risk-adjusted rate as:

$$SE\_risk\text{-}adjusted\ rate = (reference\ rate/exp_i)*sqrt(\Sigma\ p_{ij}*(1-p_{ij}))$$

Note the above formula does not account for estimation error in p_{ij} and also does not account for multiple observations on enrollees and thus somewhat overstates true precision of the estimate.

Reliability-Adjustment

Because many of the individual quality indicators are based on small numbers of eligible cases or on rare events, we estimate HRR-level effects using Empirical Bayes methods.[15] These estimators, often referred to as shrinkage estimators because extreme values are pulled towards the overall mean, minimize prediction error and better account for regression to the mean effects than standard estimates. These estimates are also often referred to as reliability-adjusted estimates (see for example the AHRQ PSI composite measure report).[16] Reliability-adjusted estimates are computed as:

$$RAE_i=[rate_i*reliability\ weight_i] + reference\ population\ rate*[1-reliability\ weight_i]$$

where $rate_i$ is the observed or risk-adjusted rate in HRR i and reliability $weight_i$ = (Signal variance) / (signal variance + noise $variance_i$).

This formula puts additional weight on HRR-specific estimates when either the average signal to noise ratio is high or sample size is large (so that the noise variance is low).

We use an empirical Bayes approach to compute the signal and noise variance. The noise variance for each HRR is equal to the square of the standard error of the rate (computed as described above for risk-adjusted measures and = $1/n_i*sqrt\ (p_{i*}(1-p_i))$ where p_i is the observed rate for process measures). We estimate the signal variance as the observed variance of the area-level rates adjusted for sampling variability by subtracting mean sampling error variance:

$$Signal\ variance = var(rate_i) - mean(noise_i)$$

Quality Aggregation

We also report composite quality measures. We aggregate individual measures within four general domains: 1) preventable admissions (PQI composites), 2) patient safety (PSI composites), 3) process measure composites, and 4) readmission. We compute PQI and PSI composite measures by adapting the framework proposed by AHRQ to our setting. Specifically, following recommendations by AHRQ we compute three PQI composites: overall, acute, and chronic. Each of these is computed as the average risk-adjusted rate of relevant individual measures. The overall composite is the average of all 12 individual PQIs, the acute composite includes dehydration admissions, bacterial pneumonia admissions and urinary tract infection admissions; the chronic composite includes diabetes short-term complications, diabetes long-term complications, COPD admissions in older adults, hypertension, CHF, angina without a procedure, asthma in younger adults, and lower-extremity amputations. Note this approach of averaging risk-adjusted rates for each admission effectively uses the prevalence of the conditions to weight the individual measures. Again following AHRQ recommendations, we compute a PSI composite as a weighted average of standardized individual measures (i.e. standardized by dividing the risk-adjusted rate by the reference population rate so that individual measures are a ratio of observed to expected counts).[16] Weights were computed as the relative frequency of the adverse event (numerator).

Because there are no gold standards for weighting individual process measures into a composite, we create several different composites for the process measures. Our first approach is to compute an HRR level summary measure equal to the fraction of measures where the quality standard was met among all measures for which individuals in the HRR were eligible (often called an opportunity weighted composite). There are several potential issues with this simple aggregation. First, some quality targets may be more easily attained than others. Second, this simple aggregation weights individual quality indicators according to the number of patients eligible for the measure and does not easily allow for additional emphasis on measures that are more salient or more highly variant across areas that may be better markers of area-level quality. Thus we also create a second composite that equally weights standardized individual measures. In each of these, enrollees who are eligible for multiple measures contribute more heavily compared to those eligible for fewer measures. Finally, if quality markers are not correlated with each other, aggregating into a summary may mask differences across areas (for example, if some areas perform well on diabetes care but poorly in treatment for low back pain). Thus in addition to univariate composite quality measures, in the aggregate cohort, we explore the use of an empirical approach that uses observed correlations among individual quality measures to aggregate similar measures.

Each composite measure is a weighted average of K individual measures:

$$Comp_i = w' R$$

where w is a Kx1 vector of weights and R is a Kx1vector of reliability adjusted quality measures. We compute the standard error of the composite errors as:

$$SE\_comp = sqrt(w' \, var(R) \, w)$$

Variation in Market Level Variables

Several factors related to provider and insurance market structure likely play a significant role in determining health care expenditure. For instance, competition among providers should theoretically drive down prices. Furthermore, evidence suggests that provider concentration and primary care concentration is associated with lower spending.[17] We explore two approaches to estimating the association between market level measures and the components of spending (quantity, price, and mark-up). The first is to merge area level estimates to individuals and run least squares regressions exactly as we described above, taking the average of residuals as the area level estimate. In addition, as these factors are not patient level characteristics, we run second-order regressions to evaluate the association between market level characteristics and market level variation.

Table 7. List, source, and level of construction of market level variables

Measure	Source	Geography
Total population	ARF	HRR and HSA
Active MDs per 1,000	ARF	HRR and HSA
Active primary care physicians per 1,000	ARF	HRR and HSA
Active specialists per 1,000	ARF	HRR and HSA
Ratio of PCPs to specialists	ARF	HRR and HSA
Number of hospital beds per 1,000	ARF	HRR and HSA
Proportion of population fully insured by commercial PPO	Interstudy	HRR and HSA
Proportion of population fully insured by commercial HMO	Interstudy	HRR and HSA
Proportion of population fully insured by commercial POS	Interstudy	HRR and HSA
Proportion of population uninsured	Interstudy	HRR and HSA
Proportion of population covered by Medicare	Interstudy	HRR and HSA
Proportion of population covered by Medicaid	Interstudy	HRR and HSA
Medicare malpractice GPCI	CMS	HRR and HSA
Herfindahl index (HHI) of competition based on the distribution of beds in each market	AHA	HRR
Presence of a teaching hospital	AHA	HRR
Presence of a specialty hospital	AHA	HRR
Presence of a government owned hospital	AHA	HRR
Health professional shortage area	HRSA	HRR and HSA

Table 7 contains a list of market level variables, the source of the measure, and levels of geography at which they are constructed

ARF, Interstudy, and CMS data are only available at the county level, whereas HRRs (and HSAs) are defined as a combination of zip codes. Counties are not nested in HRRs, so we have constructed a county to HRR (and HSA) crosswalk weighted by proportion of the HRR in each county following instructions provided by Dartmouth. Each source that contains measures at the 5-digit FIPS code level is merged by FIPS to the crosswalk. The weights are then applied to each measure, and the resulting values collapsed to the HRR level.

For the Medicare malpractice GPCI, the cross-walking procedure is identical to that used in the input price adjustment method. Each “locality” name is matched to a county name and 5-digit FIPS code. Counties are perfectly nested in localities. This FIPS code file is cross-walked to HRR (HSA) as described above.

Data is not available for each year in the three year period for the ARF, but values are stable across time. Therefore, for each measure we take an average value for as many years available from 2007-2009. For HRR analysis, we employ two methods of analysis. First, we merge the market level characteristics to enrollee-level data and calculate area level effects as the average of residuals. We analyze the extent to which the coefficient of variation on spending, quantity, price, and mark-up changes once these covariates are held constant.

In addition, we run second-order regression with each area (HRR or HSA) as an observation. Markets for hospitals services are likely larger than markets for physician services, both of which are smaller than insurance markets. Therefore we also construct measures at the HSA level and run the second-order regression with HSA as the unit of observation. We regress the market level measures on spending, quantity, price, mark-up, and the quality composites.

Processing and Quality Assurance

Data was processed using Unix SAS v9.2 and Stata/MP 11.2. Code related to data processing steps was reviewed both internally as well as by a member of the Thomson Reuters team. In addition, code and output was reviewed by an external agency for quality assurance purposes.

Results

To ease the reader through the results section, each is presented with a guiding research question.

Sample Characteristics

Guiding Question: What are the characteristics of enrollees in the MarketScan database?

Our primary analysis includes 113 million enrollee/year observations during our study period (not adjusted for partial year enrollment). Table 8 displays basic sample characteristics. On average, MarketScan enrollees are young, with a mean age of 32.6 and are almost evenly split by gender. Around 20 percent of the sample is enrolled in the data for less than a year. Given the age of the sample and previous analysis by Thomson Reuters, most of this attrition is thought to be due to enrollees moving as opposed to death, although we do not reliably observe the cause of attrition. 6.1 percent of the sample has at least one capitated claim during the year. These capitated claims have spending imputed as described above.

Table 8 also displays sample averages of the benefit generosity measures and the distribution of insurance plan types. Average out of pocket expenditures for the five services fit within the range of expected values, with generic medications costing around 3.5 times less than brand name medications, and with office visits and emergency department visits costing less than inpatient admissions, where patients commonly face deductibles or coinsurance. Insurance plans are predominately Preferred

Provider Organizations (66.7%), Health Maintenance Organizations (15.4%), and Point of Service plans (8.8%), with much smaller samples of traditional indemnity or consumer driven health plans/high deductible health plans (5.7% collectively). These estimates of benefit generosity and the distribution of insurance plan type is similar to that reported in the 2007 Kaiser Family Foundation/Health Research and Educational Trust Employer Health Benefits report.[18]

Table 8. Overall sample characteristics

	Average (n=113 million enrollee/years)
Age	32.6
Percent Female	51.3%
Percent with Drug Coverage Data	76.9%
Percent Capitated	6.1%
Percent of Data from Employers	43.5%
Percent Enrolled Full Year	84.9%
Average out of pocket (AOOP) for Brand Drug	$26.9
AOOP for Generic Drug	$7.3
AOOP for Office Visit	$24.3
AOOP for Inpatient Admission	$631.4
AOOP for Emergency Department Visit	$126.5
Percent in HMOs	15.4%
Percent in PPOs	66.7%
Percent in POSs	8.8%
Percent in CDHP/HDHP	2.7%
Percent in Traditional Indemnity	3.0%

Table 8 depicts summary statistics for several key variables in the MarketScan Database.

Guiding Question: How much variation is there in enrollee and insurance traits across HRRs?

Table 9 presents unadjusted summary statistics across the 306 HRRs. The median sample size across HRRs is over 200,000 person/years, indicating that fairly precise estimates are possible. Both age and sex are tightly distributed around a mean age of 32.9 and mean percent female around 51.0%. Health status is similarly distributed, with an interquartile percent difference of about 9%. Capitation, on the other hand, is highly skewed with a very few areas driving up the average value. While the median area has less the 1% of its enrollees with any capitated claims, the maximum is above 55%. This indicates that our imputation method will only have a meaningful effect on the handful of HRRs with a large proportion of capitation (see Sensitivity Analysis 2 in the Appendix).

Table 9. HRR summary statistics (includes those eligible for cluster 2)

	Average	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile
Population	368,965	58,389	99,367	204,968	418,271	831,335
Age	32.9	30.9	31.8	32.8	33.8	35.2
Percent Female	51.0%	49.4%	50.0%	50.9%	51.7%	52.8%
Health Status: DxCG score	0.93	0.85	0.88	0.92	0.97	1.03
Percent Capitated	5.1%	0.0%	0.1%	0.7%	4.2%	14.9%
Percent with Drug Coverage Data	79.5%	63.7%	71.3%	81.3%	88.0%	92.3%
Portion of Year Enrolled	85.0%	81.9%	83.5%	85.3%	86.7%	88.0%
HMO	12.1%	0.7%	2.7%	7.4%	15.8%	30.7%
PPO	69.0%	47.3%	60.5%	71.3%	80.4%	85.9%
POS	8.2%	1.6%	2.9%	5.9%	10.4%	15.4%
CDHP/HDHP	3.4%	1.0%	1.6%	2.8%	4.6%	6.8%
Indemnity	3.5%	0.7%	1.3%	2.1%	4.3%	7.9%
Percent of Data from Employers	50.4%	13.5%	31.3%	54.3%	70.9%	79.0%
AOOP Brand	26.8	21.7	25.3	27.1	28.7	30.2
AOOP Generic	7.3	6.2	6.8	7.4	7.8	8.1
AOOP Office Visit	24.7	20.5	23.0	25.2	26.6	28.0
AOOP ED Visit	132.0	76.8	107.2	128.4	156.9	183.7
AOOP Inpatient Admission	664.5	379.9	521.8	656.2	811.0	918.2

Table 9 depicts summary statistics across the 306 HRRs. This table depicts the variation in several measures, such as differences in age, sex, insurance type, and benefit generosity.

Correlation

Our main findings:

- Price and quantity are negatively correlated.
- Spending is positively correlated with price.
- Spending is uncorrelated with quantity.
- Health status is highly correlated with quantity, but with a much weaker association with spending.
- Inpatient and outpatient spending are moderately correlated, but prescription drug spending is almost completely uncorrelated with inpatient and outpatient spending.
- While area-level spending is positively correlated from one year to the next, the correlation drops across multiple years.
- Most of the utilization indicators are moderately positively correlated with one another; however, some measures show much weaker correlations (e.g. correlation between count of office visits and inpatient admissions is 0.10).
- The two methods of calculating aggregate quantity have moderate positive correlation.
- Spending is positively correlated across the 15 disease cohorts, although the extent of correlation varies substantially.
- The extent of variation is consistent across HRRs, HSAs, and MSAs. The smaller units (HSAs and MSAs) show slightly higher coefficients of variation, but in general are consistent with findings at the HRR level.
- Our results indicate that the ratio of highest to lowest HSAs within HRR varies substantially.

Variation in Spending

Our main findings:

- There is significant variation in spending across HRRs.
- Adjusting for input prices reduces the variation by about 10% across all models.
- Controlling for age, sex, and health status does not significantly reduce variation in spending.
- Much of the variation in spending can be attributable to price. This is consistent when variation in spending is deconstructed by type of service (inpatient and outpatient).
- Covariates of age, sex, health status, race, income, and insurance characteristics do not explain a great deal of the variation in input price adjusted medical spending across HRRs.
- While inpatient and outpatient spending are moderately correlated, spending in neither category correlates with prescription drug spending.
- Most of the variation in spending is driven by variation in spending conditional on use as opposed to differences in the number people using care.
- Some, but not all, of the change in spending is related to people entering or exiting the database.

Guiding Question: What is the relationship between spending, price, and quantity?

Figures 1-3 show correlation plots of spending versus implied price, spending versus quantity, and implied price (spending/quantity) versus quantity, after adjusting for age, sex, and health status. The relationship of these three measures impacts the decomposition of variation described below. Specifically, price and quantity (Figure 3) are negatively correlated, which causes a sizable negative covariance term in the spending decomposition equation defined above. Likewise, spending is positively correlated with price and relatively uncorrelated with quantity. Figure 4 shows that health status is highly correlated with quantity, but with a much weaker association with spending. There are concerns that health status and quantity are highly endogenous, as an enrollee's previous years use is the basis of health status.

Figure 1. Relationship between spending and implied price, with and without controls for age, sex, and health status

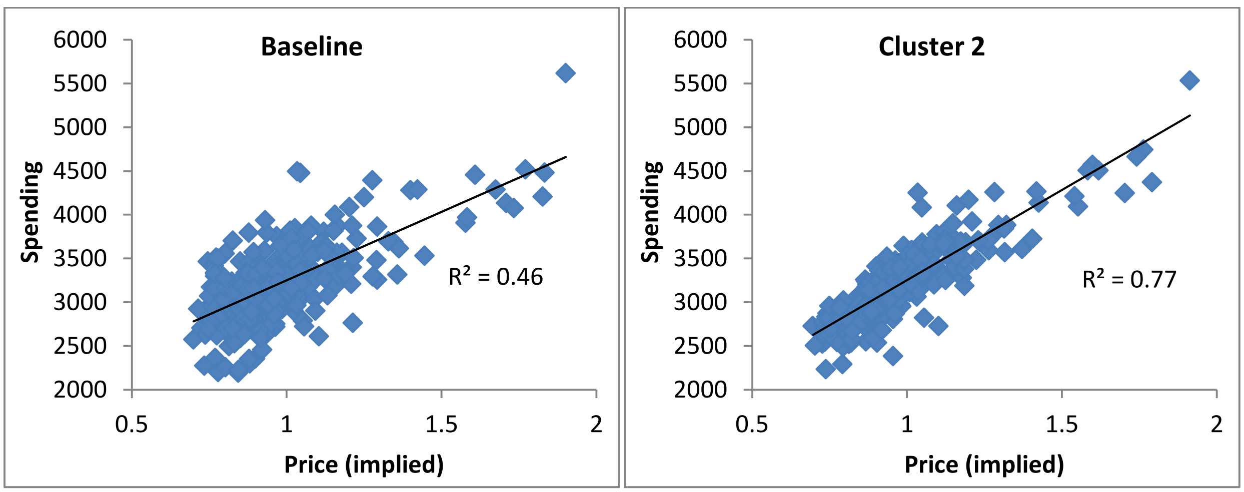

Figure 1 displays a correlation of spending and price.

Figure 2. Relationship between spending and quantity, with and without controls for age, sex, and health status.

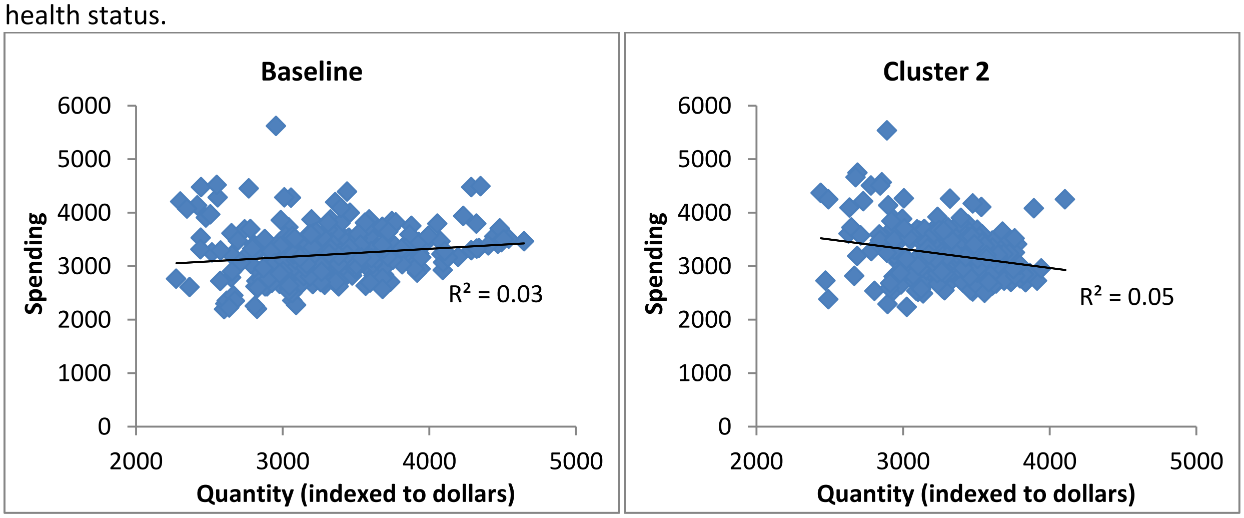

Figure 2 displays a correlation between spending and aggregate quantity.

Figure 3. Correlation of HRR level Price and Quantity, with and without controls for age, sex, and health status.

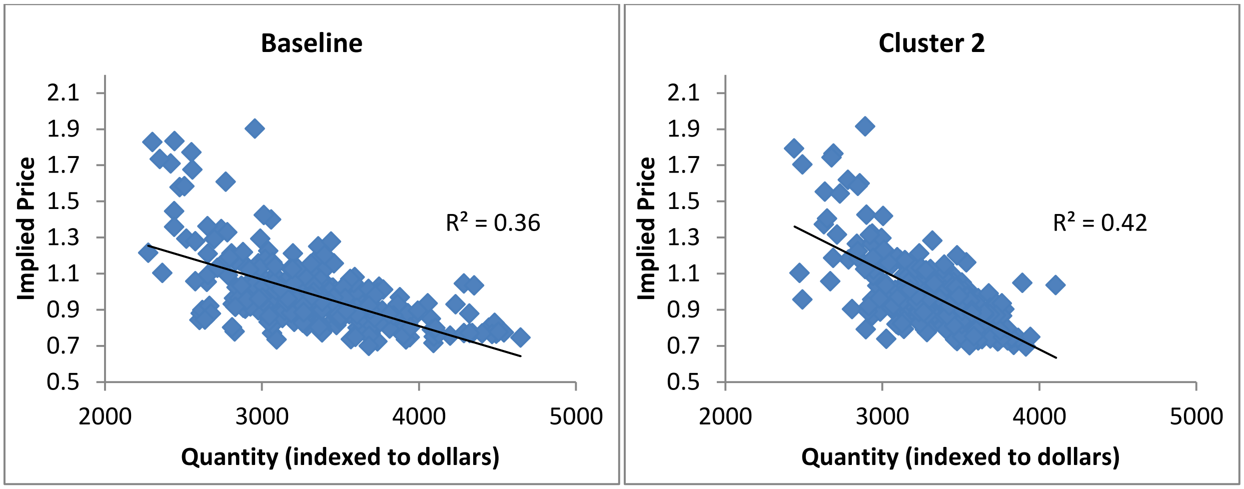

Figure 3 displays the association between HRR aggregate quantity and implied price.

Figure 4. Correlation of health status and spending compared to the correlation of health status and aggregate quantity.

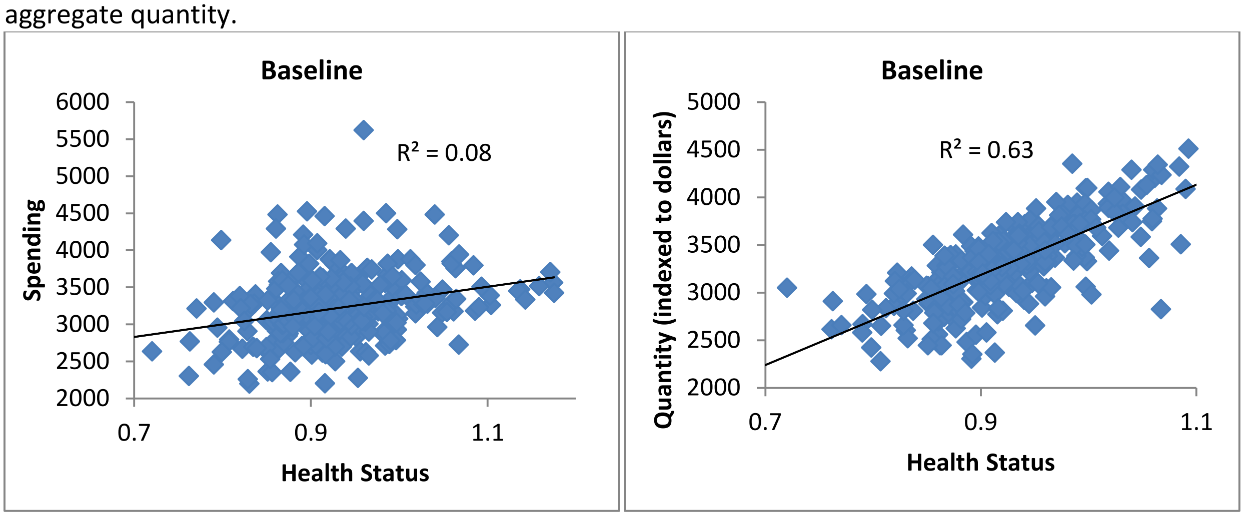

Figure 4 displays the association between HRR health status and spending as well as health status and aggregate quantity.

Guiding Question: How much variation exists in unadjusted spending, input price adjusted spending, and quantity?

Table 10 displays a summary of spending variation that is not adjusted for covariates. The top 10th percentile spends 39 percent more than the bottom 10th percentile on medical spending, 71 percent more on drug spending, 36 percent more on input price adjusted medical spending, and consumes 40 percent more quantity.

Table 10. HRR Variation in Spending (unadjusted)

	Medical Spending	Drug Spending	Input Price Adjusted Medical Spending	Quantity (Indexed to dollars)
Mean	3,222	828	3,321	3,336
Standard Deviation	450	164	417	453
Coefficient of Variance	0.14	0.20	0.13	0.14
Ratio of 75th percentile to 25th percentile	1.18	1.27	1.16	1.18
Ratio of 90th percentile to 10th percentile	1.39	1.71	1.36	1.40

Table 10 presents a summary of unadjusted metrics of variation for medical spending, drug spending, input price adjusted spending, and quantity. Models are corrected for partial year enrollment and contain year dummies. Markets are ranked separately for each outcome variable.

Guiding Question: How does adjusting for prices affect variation?

Controlling for differences in input price reduces the coefficient of variation in spending by around 10 percent across multiple models. However, variation in quantity is only reduced by 3 percent without adjustment for covariates. When controlling for age, sex, and health status, the coefficient of variation is reduced by 38% percent compared to spending.

Table 11. Changes in Variation for Spending, Input Price Adjusted Spending, and Quantity across three Models

		Baseline	Cluster 2	Cluster 6
Spending	Coefficient of Variation	0.14	0.14	0.14
Input Price Adjusted Spending	Coefficient of Variation	0.13	0.12	0.13
	Percent reduction in the coefficient of variation compared to spending	10.2%	12.7%	7.5%
Quantity	Coefficient of Variation	0.14	0.09	0.11
	Percent reduction in the coefficient of variation compared to spending	2.7%	38.3%	17.5%

Table 11 displays the coefficient of variation in spending, input price adjusted spending, and quantity across the baseline (unadjusted) model, a model controlling for cluster 2 covariates, and a model adjusting for cluster 6 covariates. It also shows the difference in this ratio when compared against the spending model.

Using cluster 2 covariates, the relative share of variation attributable to price and quantity is somewhat sensitive to the method of calculating price and quantity (Table 12). Whereas price represents around four times as much of an effect as quantity when using the aggregate quantity measure and implied price, it only represents around twice the effects when price is calculated with the Laspeyres index and quantity is implied. On the other hand, the relative effect of quantity is similar to the base model when using implied spending (price index multiplied by the aggregate quantity measure).

Table 12. Decomposition of variation in spending, by method of calculating price, quantity, and spending.

	Coefficient of Variation in Spending	Coefficient of Variation in Quantity	Coefficient of Variation in Price	Relative Share Attributable to Quantity	Relative Share Attributable to Price
Using aggregate quantity, implied price, and observed spending	0.14	0.09	0.19	16%	84%
Using price index, implied quantity, and observed spending	0.14	0.17	0.23	35%	65%
Using aggregate quantity, price index, and implied spending.	0.18	0.17	0.23	12%	88%

Table 12 displays the relative share of variation attributable to price and quantity across three methods of calculating price, quantity, and spending. All models are adjusted for age, sex, and health status.

Guiding Question: Is the relative effect of price and quantity consistent in inpatient and outpatient care?

When using the aggregate quantity measure derived by setting prices constant across all services, price varies more than quantity, and does so to a similar degree for both inpatient and outpatient services (Table 13).

Table 13. Decomposition of variation in spending by type of service. Relative proportion of spending Variance Due to Spending or Price.

	Quantity	Price
Total Medical	16%	84%
Inpatient	18%	82%
Outpatient	21%	79%

Table 13 compares results of the spending composition into quantity and price effects broken down by place of service.

Guiding Questions: What is the relative role of input prices and mark-up in explaining variation in price? Is this consistent in inpatient and outpatient care?

Variation in mark-up (difference between transaction price and input price) explains around 5 times more of the variation in price as does differences in input price. This is more pronounced in outpatient services than in inpatient services (Table 14).

Table 14. Decomposition of variation in price by type of service. Relative proportion of price variance due to input price or mark-up.

	Mark-up	Input Price
Medical	84%	16%
Inpatient	75%	25%
Outpatient	89%	11%

Table 14 displays the relative effect of mark-up and input price in explaining variation in price for medical, inpatient, and outpatient spending. All models are adjusted for age, sex, and health status.

Guiding Question: How much variation can be explained by various covariates?

Health status had the largest impact on input price-adjusted spending at the individual level; an increase in risk score by 10% increases spending by $362 (see Appendix Table 1 for regression coefficients). Inclusion of health status in the model increases R-squared by 7 percentage points. Insurance type was also an important predictor of spending at the individual level (for example, enrollees in high deductible plans spent $351 less compared to point of service plans). However, because of modest variation in these characteristics across HRRs (Table 9), little additional variation in input price adjusted medical spending can be explained by the 6 covariate clusters (Table 15). The ratio of 75th to 25th percentile is consistently 1.15 to 1.17 across all clusters and the baseline, unadjusted model. In addition, each of the covariate clusters is correlated above 0.9 with cluster 2, and the baseline is correlated 0.82. This indicates that the included covariates do not explain a great deal of the variation in input price adjusted medical spending across HRRs.

Table 15. HRR variation in input price adjusted medical spending

	Baseline	Cluster 1 (age & gender)	Cluster 2 (age, gender, health status)	Cluster 3 (age, gender, race)	Cluster 4 (age, gender, income)	Cluster 5 (age, gender, health status, race, income)	Cluster 6 (age, gender, insurance characteristics)
R-squared	<0.00	0.01	0.08	0.01	0.01	0.08	0.01
Standard Deviation	416.6	394.7	403.2	372.5	389.1	380.7	412.3
Coefficient of Variation	0.13	0.12	0.12	0.12	0.12	0.12	0.13
Ratio of 75th percentile to 25th percentile	1.16	1.16	1.17	1.15	1.16	1.15	1.16
Ratio of 90th percentile to 10th percentile	1.36	1.33	1.34	1.30	1.32	1.32	1.36
Correlation to Cluster 2	0.82	0.93	1.00	0.92	0.93	0.96	0.91

Table 15 summarizes the extent of variation in input price adjusted medical spending across the 306 HRRs using various clusters of covariates.

Guiding Question: Is the extent of variation consistent across inpatient, outpatient, and drug spending?

Variation is more pronounced in input price adjusted inpatient spending than for input price adjusted outpatient or drug spending. However, drug spending is more dispersed at the tails, as the ratio of the top to bottom 10th percentiles rises disproportionally compared to inpatient or outpatient services.

Table 16. Consistency of variation across type of service

	Inpatient	Outpatient	Prescription Drug
Average	954	2,358	814
Standard Deviation	164	298	122
Coefficient of Variation	0.17	0.13	0.15
Ratio of 75th percentile to the 25th percentile	1.27	1.16	1.19
Ratio of 90th percentile to the 10th percentile	1.50	1.37	1.48

Table 16 displays several measures of variation across inpatient, outpatient, and drug spending. Inpatient and outpatient spending adjusted for input prices, and all models control for age, sex, and health status (cluster 2 covariates).

Guiding Question: Are inpatient, outpatient, and drug spending correlated across areas?

Spending is not consistently high or low across categories. Inpatient and outpatient spending are moderately correlated (0.50), but prescription drug spending is almost completely uncorrelated with inpatient and outpatient spending.

Table 17. Correlation of Spending by type of service, cluster 2, HRR

	Inpatient	Outpatient	Prescription Drug
Inpatient	1		
Outpatient	0.50	1	
Prescription Drug	0.07	0.05	1

Table 17 displays pairwise correlations of input price adjusted inpatient, input price adjusted outpatient, and prescription drug spending. All models are adjusted for age, sex, and health status.

Guiding Question: Is variation on the intensive or the extensive margin (are more people using care, or is there more spending conditional on use)?

There is substantial variation in both total medical spending conditional on use as well as the probability of no use (no spending). However, spending conditional on use is highly correlated with overall spending (0.98), whereas the probability of no-use is uncorrelated with total spending. This implies that most of the variation in spending is driven by variation in spending conditional on use as opposed to differences in the number of people using care.

Table 18. Variation in the intensive and extensive margins, HRR

	Total Medical Spending	Medical Spending Conditional on Use	Likelihood of No-use
Average	3,210	3,633	0.18
Standard Deviation	447	614	0.03
Coefficient of Variation	0.14	0.17	0.16
Ratio of 75th percentile to the 25th percentile	1.19	1.22	1.20
Ratio of 90th percentile to the 10th percentile	1.36	1.46	1.46
Correlation with Cluster 2	-	0.98	0.02

Table 18 displays HRR variation in total medical spending, medical spending conditional on use, and the likelihood of no-use. All models adjusted for age, sex, and health status.

Stability over Time

Guiding Question: Are areas that are high spending in one year necessarily high spending in the next year?

Area-level spending is positively correlated from one year to the next, ranging from 0.69 from 2009 to 2010 to 0.88 from 2008 to 2009 (Table 19). However, the correlation of area level estimates drops across multiple years. For instance, the correlation of 2006 estimates to 2007 estimates is 0.77, but the correlation of 2006 estimates to 2010 estimates is 0.57. Table 20 displays the same results, but includes only enrollees who are enrolled in MarketScan throughout the period. Comparing the two populations, it appears that some but not all of the change in spending is related to people entering or exiting the database.

Table 19. Comparison of 2006 to 2010 estimates of input price adjusted spending, adjusted for cluster 2 covariates

	2006 (n=31,276,572)	2007 (n=32,517,151)	2008 (n=40,814,948)	2009 (n=39,571,126)	2010 (n=44,641,103)
2006	1	-	-	-	-
2007	0.77	1	-	-	-
2008	0.73	0.81	1	-	-
2009	0.70	0.80	0.88	1	-
2010	0.57	0.68	0.64	0.69	1

Table 19 displays pair-wise correlation results of yearly input price adjusted medical spending estimates for 2006-2010. All models are adjusted for age, sex, and health status (cluster 2 covariates).

Table 20. Comparison of 2006 to 2010 estimates of input price adjusted spending, adjusted for cluster 2 covariates. Includes enrollees enrolled through the entire 2006 to 2010 period.

	2006 (n=9,948,503)	2007 (n=9,948,503)	2008 (n=9,948,503)	2009 (n=9,948,503)	2010 (n=9,948,503)
2006	1	-	-	-	-
2007	0.73	1	-	-	-
2008	0.71	0.82	1	-	-
2009	0.68	0.78	0.85	1	-
2010	0.70	0.80	0.79	0.81	1

Table 20 displays pairwise correlation results of yearly input price adjusted medical spending estimates for 2006-2010. All models are adjusted for age, sex, and health status (cluster 2 covariates).

Aggregate Variation in Utilization

Our main findings:

- All individual utilization measures vary substantially across HRRs even among sentinel services such as emergency department visits and inpatient medical admissions.

- While most correlations are positive, the magnitude of the correlation implies that high utilization of one type of service does not necessarily imply high utilization of another. Because utilization of different types of services is not strongly correlated, aggregate quantity which averages across services is less variable than individual measures.

- The two methods of calculating aggregate quantity have a moderate positive correlation, explaining why results may be somewhat sensitive to how quantity is counted.

Guiding Question: What is the extent of variation across several utilization measures?

All individual utilization measures vary substantially across areas (the coefficient of variation is larger than 10%). Among sentinel services emergency department visits and inpatient medical admissions and days have the largest variation while rates of prescription drug fills, office visits and imaging were more stable. As expected discretionary services including hip replacement, low back surgery, nuclear stress tests and hysterectomies vary substantially across areas (coefficient of variation 30-40%)

Table 21: Summary of adjusted quantity, annualized

	Average	Standard Deviation	Coefficient of Variation	Ratio of 75th percentile to 25th percentile	Ratio of 90th percentile to 10th percentile
Aggregate Measure	3315.92	284.39	0.09	1.11	1.24
Prescription Drug Fills[d] (per capita)	12.82	1.60	0.12	1.15	1.34
Inpatient Day (per 1,000)	213.15	37.30	0.17	1.24	1.55
Inpatient Admissions (per 1,000)	62.51	9.39	0.15	1.21	1.45
Inpatient Medical Days (per 1,000)	123.95	26.53	0.21	1.30	1.69
Inpatient Medical Admissions (per 1,000)	39.59	7.45	0.19	1.26	1.58
Inpatient Surgical Days (per 1,000)	89.66	12.04	0.13	1.19	1.39
Inpatient Surgical Admissions (per 1,000)	23.05	3.29	0.14	1.22	1.43
Emergency Department Visits (per 1,000)	287.66	76.34	0.27	1.46	2.04
Office Visits (per 1,000)	3180.79	315.66	0.10	1.14	1.30
Imaging Procedures (per	1123.73	122.36	0.11	1.15	1.33

[d] Due to missing drug data in around ~20% of enrollees, per capita drug fills are based on a smaller sample.

1,000)					
Knee Replacement (per 1,000)	1.41	0.38	0.27	1.43	2.05
Hip Replacement (per 1,000)	0.69	0.12	0.17	1.26	1.53
Hysterectomies (per 1,000)	2.79	0.92	0.33	1.54	2.61
Lower Back Surgeries (per 1,000)	1.93	0.55	0.28	1.48	2.17
Nuclear Stress Test (per 1,000)	17.91	7.41	0.41	1.76	3.18
Nonnuclear Stress Test (per 1,000)	26.38	6.97	0.26	1.40	2.00
Bilateral Cardiac Cathertization (per 1,000)	0.52	0.17	0.33	1.53	2.19
Closed Cholecystectomy (per 1,000)	0.14	0.05	0.36	1.43	2.03
Open Cholecystectomy (per 1,000)	3.79	1.30	0.34	1.57	2.53

Table 21 displays summary statistics for the aggregate quantity measure and 19 individual measures. The table includes estimates of variation for each of the measures. All measures are adjusted for age, sex, and health status (cluster 2 covariates). Markets are ranked separately for each outcome variable.

Guiding Question: Is utilization of one service associated with the utilization of another?

Most of the seven utilization indicators presented in Table 22 are positively correlated in the range of 0.4-0.6. However, some measures show a much weaker correlations. For example, the correlation between count of office visits and the count of inpatient admissions is 0.10. While most correlations are positive, the magnitude of the correlation implies that high utilization of one type of service does not necessarily imply high utilization of another. Because utilization of different types of services is not strongly correlated, aggregate quantity which averages across services is less variable than individual measures. Because of the large volume of imaging services, aggregate quantity is most highly correlated with utilization of imaging procedures.

Table 22. Stability of quantity, controlling for cluster 2 covariates, HRR

	Aggregate Quantity	Aggregate Quantity (Estimated with Laspeyres price index)	Total inpatient admissions	Total inpatient days	Office visits	Emergency Department Visits	Imaging Procedures
Aggregate Quantity (Output price adjusted spending)	1	-	-	-	-	-	-
Aggregate Quantity (Estimated with Laspeyres price index)	0.53	1	-	-	-	-	-
Total inpatient admissions (count)	0.53	0.39	1	-	-	-	-
Total inpatient days (count)	0.46	0.27	0.90	1	-	-	-
Office visits (count)	0.36	<-0.01	0.10	0.14	1	-	-
Emergency Department Visits (count)	0.52	0.35	0.50	0.44	0.05	1	-
Imaging Procedures (count)	0.81	0.39	0.40	0.35	0.51	0.52	1

Table 22 displays pair-wise correlations of several of the sentinel utilization measures and the two aggregate measures. All models are adjusted for age, sex, and health status.

Guiding Question: How similar are the two approaches to calculating aggregate utilization?

The two methods of calculating aggregate quantity have a moderate positive correlation (0.53), explaining why results may be somewhat sensitive to how quantity is counted (Figure 5). Because aggregate quantity measured using output price adjusted spending is more highly correlated with individual services than quantity based on a market basket approach, we prefer the approach that aggregates across all services.

Figure 5: Correlation of two aggregate quantity measures

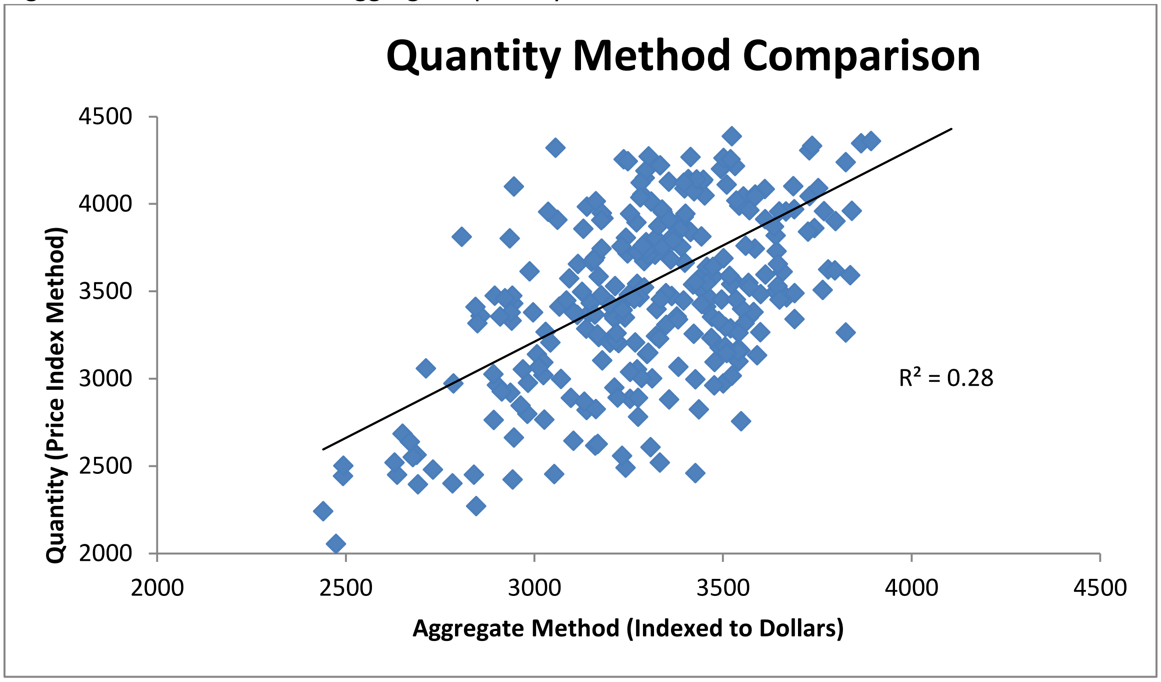

Figure 5 displays the correlation between aggregate quantity calculated with the price index and our base case scenario. Both composites are adjusted for age, sex, and health status.

Aggregate Variation in Price

Guiding Question: How similar are the two approaches of calculating price?

Figure 6 demonstrates that the implied method of calculating price, and the price index method are highly correlated (correlation=0.83).

Figure 6: Correlation of price index method and implied (residual) method to measure price

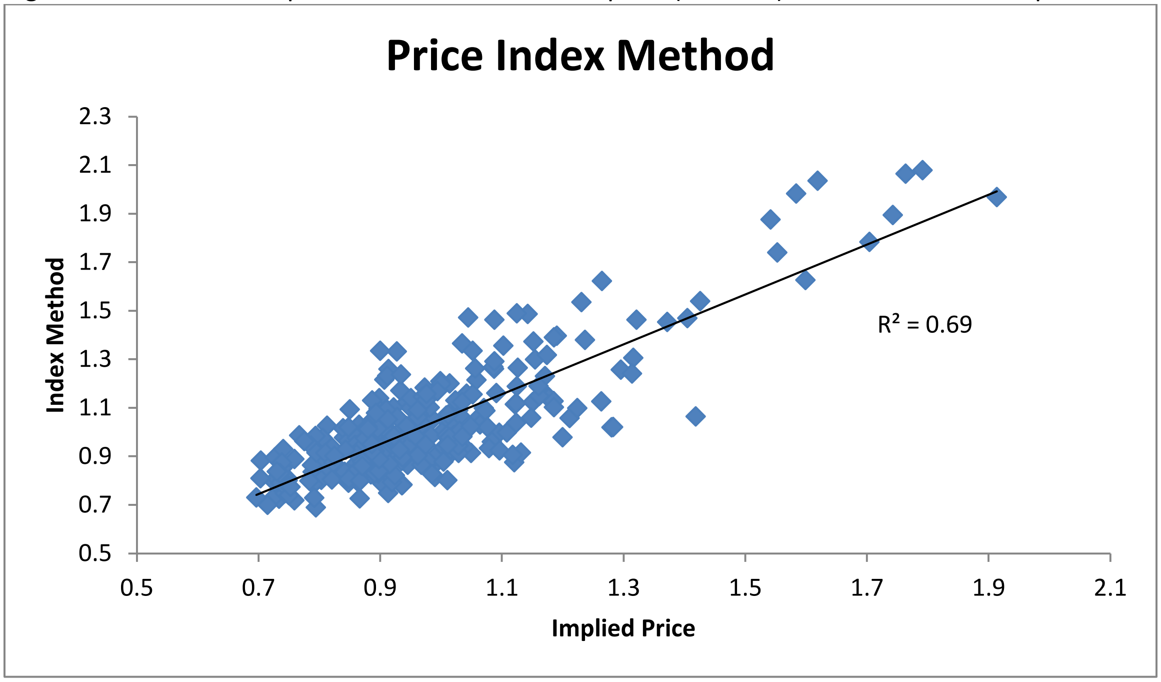

Figure 6 shows the correlation between the implied price and price calculated with the Laspeyres price index. Both models are adjusted for age, sex, and health status.

Consistency across Cohorts

Guiding Question: Is there a similar extent of variation across treatment of individuals with certain conditions?

There is generally more variation among the 15 cohorts than in aggregate spending. The coefficient of variation ranges from 0.08 for lung cancer, to 0.37 for cataracts.

Table 23. Variation within Cohorts, per member per month spending by HRR

	Overall	*Stroke*	*Diabetes*	*Pneumonia*	*Rheumatoid Arthritis*	*Depression*	*CHF*	*AMI*	*CHD*	*COPD*	*Cataracts*	*Lower Back Pain*	*Cholecystectomy*	*Lung Canc.*	*Prost. Canc.*	*Breast Canc.*
Sample Size (thousandth)	113m	62	3,648	221	280	2,195	275	81	1,130	391	188	5,281	75	13	63	42
Average Spending	181	2,167	903	3,395	1,193	859	3,150	4,811	1,898	1,792	1,126	831	3,318	8,697	2,766	6,039
Standard Deviation	16	197	104	427	145	110	358	648	396	258	416	120	788	738	331	813
Coefficient of Variation	0.09	0.09	0.12	0.13	0.12	0.13	0.11	0.13	0.21	0.14	0.37	0.14	0.24	0.08	0.12	0.13
Ratio of 75^{th} percentile to the 25^{th} percentile	1.12	1.15	1.15	1.16	1.18	1.17	1.19	1.32	1.20	1.55	1.20	1.33	1.09	1.16	1.20	1.15
Ratio of 90^{th} percentile to the 10^{th} percentile	1.24	1.35	1.36	1.34	1.37	1.32	1.37	1.71	1.44	2.36	1.47	1.79	1.22	1.34	1.38	1.35

Table 23 displays the relative extent of variation in input price adjusted medical spending across the 15 disease cohorts. All models are adjusted for age, sex, and health status.

Guiding Question: Are high (or low) spending areas for one condition similarly high (or low) for others?

Spending is positively correlated across the 15 disease cohorts, although the extent of correlation varies substantially. For instance, the correlation between input price adjusted medical spending on patients with cataract surgery and input price adjusted spending on patients with lung cancer is 0.24.

Table 24. Correlation of cohorts input price adjusted spending, HRR

	Overall	*Stroke*	*Diab.*	*Pneum.*	*Rh. Arthr.*	*Depr.*	*CHF*	*AMI*	*CHD*	*COPD*	*Catar.*	*Lower Back Pain*	*Chol.*	*Lung Canc.*	*Prost. Canc.*	*Br. Canc*
Sample (thousands)	113m	62	3,648	221	280	2,195	275	81	1,130	391	188	5,281	75	13	63	42
Overall	1.00															
Stroke	0.40	1.00														
Diabetes	0.82	0.49	1.00													
Pneumonia	0.43	0.30	0.54	1.00												
Rheum. Arthritis	0.49	0.34	0.57	0.37	1.00											
Depression	0.76	0.43	0.81	0.53	0.49	1.00										
CHF	0.57	0.44	0.66	0.45	0.39	0.55	1.00									
AMI	0.64	0.40	0.66	0.47	0.40	0.63	0.54	1.00								
CHD	0.76	0.45	0.79	0.41	0.48	0.70	0.69	0.74	1.00							
COPD	0.67	0.49	0.73	0.54	0.50	0.67	0.63	0.62	0.70	1.00						
Cataract	0.58	0.31	0.59	0.27	0.33	0.54	0.36	0.36	0.57	0.47	1.00					
Lower Back Pain	0.75	0.39	0.72	0.44	0.45	0.69	0.52	0.59	0.62	0.64	0.35	1.00				
Cholecystectomy	0.51	0.34	0.54	0.38	0.36	0.53	0.39	0.59	0.58	0.52	0.36	0.40	1.00			
Lung Cancer	0.36	0.33	0.40	0.33	0.34	0.37	0.32	0.39	0.38	0.42	0.24	0.38	0.34	1.00		
Prostate Cancer	0.50	0.29	0.53	0.34	0.36	0.48	0.39	0.50	0.55	0.44	0.33	0.43	0.36	0.37	1.00	
Breast Cancer	0.61	0.31	0.62	0.37	0.43	0.60	0.37	0.46	0.55	0.43	0.52	0.49	0.42	0.46	0.49	1.00

Table 24 displays pairwise correlations in input price adjusted medical spending between the 15 disease cohorts and the overall sample.

Consistency across Geographic Units

Guiding Question: Is the extent of variation similar across different geographic units?

The extent of variation is consistent across HRRs, HSAs, and MSAs. The smaller units (HSAs and MSAs) show slightly higher coefficients of variation, but in general are consistent with findings at the HRR level.

Table 25. Consistency of variation across geography

	HRR	HSA	MSA
Regions (n)	306	3,436	441
Average	$3,319	$3,362	$3,299
Standard Deviation	$403	$454	$496
Coefficient of Variation	0.12	0.14	0.15
Ratio of 25th percentile to the 75th percentile	1.17	1.17	1.19
Ratio of 10th percentile to the 90th percentile	1.34	1.39	1.40

Table 25 displays measures of variation in input price adjusted medical spending across HRRs, HSAs, and MSAs.

Within-Geography Variation

Guiding Question: How much variability is there within an HRR?

HSAs are nested within HRRs. To estimate within-HRR variation we compute the extent to which HHRs can predict HSA level variation. The unexplained portion of this regression is the portion of variation due to within-HRR variation. Around 70% of the variation in HSA spending is explained by variation in HRR spending, when weighted by the size of the HSAs. Moreover, the standard deviation in HSA spending within HRRs is $236 (28% of total variation at the HSA level). As an alternative metric, we also summarize the ratio of the highest spending HSA to the lowest spending HSA within each HRR. We limit this to HRRs that contain more than 3 HSAs (n=261). The results (reported in Table 26) demonstrate that the ratio of highest to lowest HSAs within HRR varies substantially. The median ratio is 1.31, with the 10th and 90th percentiles 1.14 and 1.59 respectively.

Table 26. Ratio of the highest spending HSA to lowest spending HSA within each HRR containing more than 3 HSAs (n=261)

	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile
	1.14	1.21	1.31	1.45	1.59

Table 26 displays summary statistics of the ratio of the highest spending HSA to the lowest spending HSA with each HRR containing more than 3 HSAs.

Quality Analysis

Our main findings:

- Individual quality measures tend to vary less than spending, although there is substantial variability among a handful of indicators.

- Among the 10 process measures best measured with administrative data, we observe the most variation across HRRs in screening mammography and testing of hemoglobin A1C in patients with diabetes.

- Risk-adjusted rate of readmission is 1.21 times higher in the 90th percentile versus 10th percentile HRR.

- As expected, rates of preventable admissions in this non-elderly commercially insured population are substantially lower than those observed in other populations.

- Due to small number of admissions and low event rates, most patient safety indicators are not reliably measures at the HRR level. Nonetheless, we observed moderate variation in patient safety indicators across HRRs.

- HRR-level quality measures have generally only low to moderate correlation.

- Consistent with other populations, the 12 preventable admission rates were highly correlated with each other, while patient safety indicators were not highly correlated.

- In a factor analysis fit to all 31 measures, we found that 3 composites can explain 36.4% of total variation. The first factor, which explains 25% of the variation, is associated with avoidance of all 12 preventable admissions, screening mammography and appropriate treatment for low back pain. The second factor is associated with radiation therapy following breast conserving surgery, treatment with DMARD for RA, treatment with bronchodilator for COPD, avoiding pressure ulcers and avoiding postoperative pulmonary embolism or DVT. The third factor was associated with treatment for depression, treatment with beta blockers, appropriate antibiotics for pneumonia, and avoiding readmissions.

- Our results suggest that areas with higher prices have higher quality. Most quality composites are moderately positively correlated with total spending. These correlations are reduced with input price adjustment. Further, holding price constant, most quality composites are negatively correlated with quantity (i.e. areas with high utilization have lower quality).

Guiding Questions: How much variation is there between various quality indicators, and does the extent of variation differ across indicators?

We report summaries of average quality and variation across HRRs for the 31 individual quality measures in Table 23. The individual quality measures tend to vary less than spending, although there is substantial variability among a handful of indicators.

Process Measures: Quality varies across the 12 process measures. For example, on average 99% of cataract surgeries did not have major complications while only 66% of women age 42 to 65 had a screening mammogram in the prior 2 years. Some of this variation is appropriate as not all women choose to undergo mammography screening and still others have contraindications that are not always apparent in claims data (e.g., prior mastectomy). Sample size and reliability also vary substantially across the individual measures. While these measures are based on HEDIS specifications, claims data may not reliably capture all processes. For example, stand-alone vision insurance likely explains part of the low rates of eye exams for enrollees with diabetes. Similarly, we do not observe over the counter aspirin antiplatelet therapy for stroke.

Among the 10 process measures best measured with administrative data, we observe the most variation across HRRs in screening mammography and testing of hemoglobin A1C in patients with diabetes. For example, the rate of screening mammography is 1.22 times higher in the 90th compared to the 10th percentile HRR (71.6% versus 58.8%). Similarly rates of hemoglobin A1C testing are 1.20 times higher in 90th versus 10th percentile HRR (92.1% versus 76.5%).

Readmissions: On average, commercially insured enrollees with an admission are readmitted within 30 days of discharge 7.9% of the time. Risk-adjusted rate of readmission is 1.21 times higher in the 90th percentile versus 10th percentile HRR (8.5% versus 7.1%).

Preventable Admissions: As expected, rates of preventable admissions in this non-elderly commercially insured population are substantially lower than those observed in other populations, ranging from 3.4 admissions for lower extremity amputation per 100,000 in the population to 105.5 admissions for COPD per 100,000 adults over the age of 40. However, because of large number of adult enrollees in each HRR (which serve as the denominator for the measure), avoidable hospitalization rates are measured reliably in most HRRs. In addition, avoidable admissions rates vary substantially across HRRs. For example rates of admissions for hypertension were 3.23 times higher in 90th compared to 10th percentile HRR (37.3 versus 11.6 per 100,000).

Patient Safety Indicators: Rates of patient safety events were also very low in this population ranging from 0.2 postoperative hip fractures per 1000 surgical admissions to 11.8 postoperative pulmonary embolism or deep vein thrombosis per 1000. Due to small number of admissions and low event rates, most patient safety indicators are not reliably measures at the HRR level. Nonetheless, we observed moderate variation in patient safety indicators across HRRs.

Table 27: Geographic variation in individual quality measures

	Average Sample Size	Average Reliability	Mean	Standard Deviation	Coefficient of Variation	Ratio of 75 to 25th percentiles	Ratio of 90^{th} to 10th percentiles
Process Measures							
Screening Mammography	46,372	99.4%	65.3%	5.3%	0.08	1.10	1.22
Radiation therapy following breast conserving surgery for breast cancer	28	44.6%	87.5%	4.6%	0.05	1.07	1.12
Treatment with DMARD for rheumatoid arthritis	360	71.2%	89.5%	3.2%	0.04	1.04	1.09
Treatment with bronchodilator for COPD	96	56.5%	80.1%	4.6%	0.06	1.06	1.12
Treatment with antidepressant for major depression	616	77.2%	63.8%	4.7%	0.07	1.09	1.19
Treatment with antiplatelet for stroke	120	54.1%	40.0%	5.1%	0.13	1.20	1.37
Treatment with beta blockers for AMI	95	42.0%	66.3%	3.3%	0.05	1.07	1.13
Antibiotics following pneumonia diagnosis	1,326	88.4%	68.9%	5.2%	0.08	1.09	1.18
Hemoglobin A1C for Diabetes	11,505	99.3%	84.8%	7.5%	0.09	1.09	1.20
Annual eye exam for Diabetes	11,505	99.2%	33.0%	9.1%	0.28	1.46	2.01
Appropriate treatment for low	10,262	98.5%	72.7%	6.2%	0.09	1.13	1.24

back pain							
Lack of complications following cataracts surgery	610	40.9%	99.0%	0.5%	0.005	1.01	1.01
Readmissions	8,020	79.1%	7.9%	0.8%	0.10	1.12	1.21
Preventable Admissions							
Diabetes short-term complications (PQI #01)	184,394	54.9%	20.1 per 100,000	3.9 per 100,000	0.19	1.27	1.57
Diabetes long-term complications (PQI #03)	188,663	69.4%	30.8 per 100,000	7.6 per 100,000	0.25	1.35	1.77
COPD in older adults (PQI #05)	116,092	86.5%	105.5 per 100,000	37.6 per 100,000	0.357	1.63	2.60
Asthma in younger adults (PQI #15)	72,228	59.6%	23.4 per 100,000	8.1 per 100,000	0.35	1.49	2.18
Hypertension (PQI #07)	184,394	78.7%	23.3 per 100,000	9.9 per 100,000	0.42	1.82	3.23
CHF (PQI #08)	188,663	78.3%	38.5 per 100,000	12.0 per 100,000	0.31	1.48	2.09
Dehydration (PQI #10)	184,394	82.7%	53.0 per 100,000	19.1 per 100,000	0.36	1.56	2.34
Bacteria Pneumonia (PQI #11)	188,663	86.3%	94.6 per 100,000	28.7 per 100,000	0.30	1.55	2.19
Urinary Tract Infections (PQI #12)	184,394	72.6%	39.7 per 100,000	10.5 per 100,000	0.26	1.45	1.99
Angina (PQI #13)	188,663	73.8%	12.9 per 100,000	5.9 per 100,000	0.45	1.74	2.77
Uncontrolled diabetes (PQI #14)	188,663	71.3%	8.1 per 100,000	4.2 per 100,000	0.52	2.08	3.68
Lower extremity amputation (PQI	188,663	22.6%	3.4 per 100,000	0.4 per 100,000	0.12	1.13	1.31

#16)							
Patient Safety Indicators							
Pressure ulcer (PSI #03)	2,961	55.1%	5.5 per 1000	1.5 per 1000	0.29	1.42	1.93
Iatrogenic pneumothorax (PSI #06)	11,334	30.5%	0.4 per 1000	0.1 per 1000	0.20	1.26	1.62
CVC-related blood stream infection (PSI #07)	10,337	17.8%	0.6 per 1000	0.05 per 1000	0.09	1.10	1.24
Postoperative hip fracture (PSI #08)	3,546	16.6%	0.2 per 1000	0.04 per 1000	0.21	1.27	1.63
Postoperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI #12)	5,261	75.6%	11.8 per 1000	3.3 per 1000	0.28	1.40	1.86
Postoperative Wound Dehiscence Rate (PSI #14)	1,051	4.8%	0.6 per 1000	0.04 per 1000	0.07	1.06	1.15
Accidental Puncture or Laceration Rate (PSI #15)	11,782	59.7%	3.7 per 1000	0.7 per 1000	0.20	1.31	1.66

Table 27 displays descriptive statistics and measures of variability across a number of quality indicators.

Guiding Question: Do areas perform similarly across different types of quality indicators?

HRR-level quality measures were generally only moderately correlated. For example, pairwise correlations between the 10 process measures varied from -0.18 (radiation therapy following breast conserving therapy and appropriate treatment for low back pain) to 0.52 (antidepressant treatment and treatment with beta blockers following an AMI), but most were <0.30. However, as seen previously in all other populations, the 12 preventable admission rates were highly correlated with each other. Also consistent with previous analyses, patient safety indicators were not highly correlated (pairwise correlation all < 0.20). Correlations across measure types (for example, process measures with preventable hospitalizations) were also generally low, with a few exceptions.

We report estimated factor loadings in Table 24. Consistent with low correlation pairwise correlations in process measures, fitting a factor analysis to the 10 process measure, we find that two factors explain only 31% of total variance in process quality scores. The first factor is most strongly related to screening mammography, antidepressant treatment for major depression, treatment with beta blockers for AMI and appropriate treatment for low back pain. The second factor is related to radiation therapy following breast conserving surgery, treatment with DMARD for RA, treatment with bronchodilator for COPD and appropriate antibiotics for pneumonia. The lack of cohesiveness clinically and low proportion of total variance explained by these two factors does not support empirical based aggregation of these measures. Two factors can explain 68.7% of variation in 12 PQI measures. The factors divide the measures into chronic and acute admissions, supporting the use of two composites combing rates for chronic and acute admissions respectively. In addition, the high correlation among all 12 PQI measures supports the value of an overall composite for preventable admissions. Because of low correlation among patient safety measures, aggregating these measures is not supported by empirical analyses (a single factor only accounts for 10% of total variation). Finally, in a factor analysis fit to all 31 measures, we found that 3 composites can explain 36.4% of total variation. The first factor, which explains 25% of the variation, is associated with avoidance of all 12 preventable admissions, screening mammography and appropriate treatment for low back pain. The second factor is associated with radiation therapy following breast conserving surgery, treatment with DMARD for RA, treatment with bronchodilator for COPD, avoiding pressure ulcers and avoiding postoperative pulmonary embolism or DVT. The third factor was associated with treatment for depression, treatment with beta blockers, appropriate antibiotics for pneumonia, and avoiding readmissions. Again, because of the lack of clinical cohesiveness of the factors and low variance explained we do not report empirically derived composite measures. In addition, composite measures of process quality and patient safety should be interpreted cautiously as averaging uncorrelated individual indicators may mask area-level variation in quality.

Table 28: Relationship among quality measures: Factor loadings.

	Process only Factor 1	Process only Factor 2	PQI Factor 1	PQI Factor 2	PSI Factor	All measures Factor 1	All measures Factor 2	All measures Factor 3
% of variance explained	18.4%	12.9%	34.2%	24.5%	10.2%	24.7%	6.4%	5.3%
Screening Mammography	0.41	--				-0.39	--	--
Radiation therapy following breast conserving surgery for breast cancer	--	0.36				--	-0.37	--
Treatment with DMARD for rheumatoid arthritis	--	0.58				--	-0.44	0.41
Treatment with bronchodilator for COPD	--	0.50				--	-0.41	--
Treatment with antidepressant for major depression	0.82	--				-0.47	--	0.50
Treatment with beta blockers for AMI	0.58	--				--	--	0.55
Antibiotics for pneumonia	0.46	0.68				-0.44	0.38	0.46
Hemoglobin A1C for Diabetes	--	--					--	--
Appropriate treatment for low back pain	0.56	--				-0.62	--	--
Lack of complications following cataracts	--	--				--	--	--

surgery								
Readmissions						--	--	0.31
Diabetes short-term complications (PQI #01)			0.63	0.35		0.75	--	--
Diabetes long-term complications (PQI #03)			0.87	--		0.85	0.31	--
COPD in older adults (PQI #05)			0.60	0.52		0.85	--	--
Asthma in younger adults (PQI #15)			0.59	--		0.65	--	--
Hypertension (PQI #07)			0.70	0.35		0.82	--	--
CHF (PQI #08)			0.74	0.35		0.79	--	--
Dehydration (PQI #10)			0.31	0.77		0.70	--	--
Bacteria Pneumonia (PQI #11)			0.29	0.84		0.76	--	--
Urinary Tract Infections (PQI #12)			0.46	0.70		0.79	--	--
Angina (PQI #13)			0.41	0.32		0.47	--	--
Uncontrolled diabetes (PQI #14)			0.69	0.43		0.77	--	--
Lower extremity amputation (PQI #16)			0.41	--		0.34	--	--
Pressure ulcer (PSI #03)					0.31	--	0.50	--
Iatrogenic pneumothorax (PSI #06)					0.34	--	--	--
CVC-related blood stream infection (PSI					--	--	--	--

#07)								
Postoperative hip fracture (PSI #08)					--	--	--	--
Postoperative Pulmonary Embolism or Deep Vein Thrombosis Rate (PSI #12)					0.59	--	0.47	--
Postoperative Wound Dehiscence Rate (PSI #14)					--	--	--	--
Accidental Puncture or Laceration Rate (PSI #15)					--	-0.30	--	--

Table 28 displays descriptive statistics and measures of variability across a number of quality indicators.

Across all opportunities, processes are obtained 66% of time (Table 29). Excluding the two measures not accurately measured using claims data, this rate increases to 71%. Using the opportunity weighted composite, quality is 14% higher in the top 90^{th} percentile compared to the lowest 10^{th} percentile. However, the opportunity weighted measures is strongly related to screening mammography because of the large number of women eligible for this measures. There is substantially less variation when quality is averaged across the 10 measures. Rates of avoidable admissions in the 90^{th} percentile area are double those in 10^{th} percentile areas and rates of patient safety events are approximately 50% higher.

We find moderate correlation across domains of quality (Table 30). Process measures are negative correlated with preventable admissions (i.e. increased use of recommended preventive and chronic care is associated with *avoidance* of potentially preventable admissions). Rates of readmissions within 30 days of an inpatient episode are not highly correlated with other quality measures.

Table 29: Geographic variation in composite quality measures

	Mean	Standard Deviation	Coefficient of Variation	Ratio of 75 to 25th percentiles (difference for empirical composites)	Ratio of 90^{th} to 10th percentiles (difference for empirical composites)
Opportunity Weighted Composite	65.2%	4.1%	0.06	1.09	1.16
Indicator Average Composite	70.9%	2.1%	0.03	1.04	1.07
Opportunity Weighted Composite Restricted Measures	70.2%	3.7%	0.05	1.07	1.14
Indicator Average Composite Restricted Measures	77.8%	2.3%	0.03	1.04	1.07
Preventable Admissions Composite	453.4 per 100,000	118.3 per 100,000	0.26	1.43	1.99
Preventable chronic admissions composite	266.0 per 100,000	74.2 per 100,000	0.28	1.43	2.06
Preventable acute admissions composite	187.3 per 100,000	53.2 per 100,000	0.28	1.46	2.02
Patient safety composite	3.2 per 1000	0.6 per 1000	0.17	1.22	1.46

Table 29 displays descriptive statistics and measures of variability across a number of composite quality indicators.

Table 30. Correlation of quality measures

	Opportunity Weighted Composite Restricted Measures	Indicator Average Composite Restricted Measures	Preventable admissions composite	Preventable chronic admissions composite	Preventable admissions acute composite	Patient Safety composite	Readmissions
Opportunity Weighted Composite Restricted Measures	1.00						
Indicator Average Composite Restricted Measures	0.70	1.00					
Preventable Admissions Composite	-0.53	-0.58	1.00				
Preventable chronic admissions composite	-0.53	-0.64	0.95	1.00			
Preventable acute admissions composite	-0.41	-0.34	0.86	0.66	1.00		
Patient safety composite	-0.23	-0.29	0.15	0.30	-0.15	1.00	
Readmissions	0.11	0.21	-0.10	-0.09	-0.08	0.23	1.00

Table 30 displays pairwise correlation across a number of composite quality measures, weighted by HRR population.

Guiding Question: Are aggregate composites of quality associated with price, spending, use, or input price adjusted spending?

Most quality composites are moderately positively correlated with total spending. These correlations are reduced with input price adjustment. Further, holding price constant, most quality composites are negatively correlated with quantity (i.e. areas with high utilization have lower quality). These results suggest that areas with higher prices have higher quality.

Table 31. Correlation of quality composites and spending, input price adjusted spending, and multiple utilization measures.

	Medical Spending	Input Price Adjusted Spending	Quantity	Total inpatient days	Total office visits	Imaging	Prescription Fills
Opportunity Weighted Composite Restricted Measures	0.18	0.11	-0.13	-0.51	0.05	-0.02	0.02
Indicator Average Composite Restricted Measures	0.27	0.22	-0.22	-0.44	-0.24	-0.26	0.00
Preventable Admissions Composite	-0.25	-0.07	0.44	0.79	0.20	0.47	0.36
Preventable chronic admissions composite	-0.23	-0.13	0.44	0.77	0.22	0.46	0.28
Preventable acute admissions composite	-0.25	0.02	0.37	0.69	0.15	0.41	0.42
Patient safety composite	0.04	-0.18	0.00	0.14	0.05	0.01	-0.31
Readmissions	0.02	-0.07	-0.09	-0.02	-0.07	-0.04	-0.19
Mammography	0.19	0.11	0.06	-0.38	0.12	0.11	0.13
Antibiotics following pneumonia diagnosis	0.30	0.30	-0.06	-0.38	-0.23	-0.23	0.21
Hemoglobin A1C test	0.00	0.06	-0.26	-0.08	0.03	-0.06	-0.05
Imaging following complaint of lower back pain	0.24	0.07	-0.57	-0.44	-0.36	-0.61	-0.53

Table 31 displays the correlation of various composite quality measures and spending, quantity, and input price adjusted spending.

Market Level Analysis

Guiding Question: How much variation exists in market level measures?

Market level characteristics tend to vary significantly more than variation in spending, utilization, price, and quality. The Herfindahl-Hirschman Index (HHI) of competition in the hospital market was calculated based on number of hospital beds. Table 32 displays the mean and various metrics of variation for 18 market level variables analyzed at the HRR level.

Table 32. Summary Statistics of Market Level Variables

	Mean	Standard Deviation	Coefficient of Variation	Ratio of 75th to 25th Percentiles	Ratio of 90th to 10th Percentiles
Total population	501,443	911,571	1.82	4.89	18.77
Active MDs per 1,000	2.837	1.328	0.47	1.59	2.50
Active primary care physicians per 1,000	1.037	0.332	0.32	1.53	2.07
Active specialists per 1,000	1.800	1.025	0.57	1.71	2.84
Ratio of PCPs to specialists	0.623	0.128	0.21	1.31	1.70
Number of hospital beds per 1,000	3.967	1.675	0.42	1.73	2.81
Proportion of population fully insured by commercial PPO	0.143	0.049	0.34	1.57	2.40
Proportion of population fully insured by commercial HMO	0.081	0.082	1.01	5.13	18.79
Proportion of population fully insured by commercial POS	0.016	0.021	1.30	4.76	21.63
Proportion of population uninsured	0.143	0.044	0.31	1.61	2.17
Proportion of population covered by Medicare	0.152	0.031	0.20	1.31	1.67
Proportion of population covered by Medicaid	0.155	0.052	0.34	1.54	2.32
Medicare malpractice GPCI	0.891	0.384	0.43	1.77	2.84

HHI	2,780	1,664	0.60	2.25	4.80
Presence of a teaching hospital	0.450	0.496	1.10	-	-
Presence of a specialty hospital	0.106	0.297	2.81	-	-
Presence of a government owned hospital	0.442	0.495	1.12	-	-
Health professional shortage area	0.337	0.344	1.02	-	-

Table 32 displays summary statistics of market level variables.

Guiding Questions: What is the association between market level variables and the components of spending? Does the association between market level traits and spending differ if we use HSAs as (versus HRRs) as our market definition?

The definition of the market (HRR versus HSA) appears to affect the significance and magnitude of associations. However, as there is more error inherent to the cross walking procedure that matches counties to HSAs, and as patients are much more likely to go out of their HSA to receive care (compared to HRRs), these estimates may not be reliable.

Table 33. Second-Order Regression Output for HRRs and HSAs

	Spending		Quantity		Price		Mark-up	
	HRR	HSA	HRR	HSA	HRR	HSA	HRR	HSA
R-Squared	0.16	0.06	0.42	0.22	0.40	0.22	0.42	0.22
Population (ln)	2.89	-0.51	2.68	3.36***	0.00	-0.011***	-0.0334***	-0.022***
Physicians Per 1,000	3.04	2.74***	-4.914***	-1.84***	0.0308***	0.013***	0.0213***	0.0096***
Ratio of PCP to specialists	-27.39	0.026	-38.06***	-0.0007	0.04	6.55e-05	0.02	-0.00032
Hospital beds per 1,000	-2.73	-0.81***	1.895*	-0.13	-0.0172**	-0.0021***	-0.0117**	-0.0020***
HHI	-0.001		0.001		0.000		0.000	
Malpractice GPCI	-11.96*	0.31	33.13***	27.7***	-0.170***	-0.080***	-0.168***	-0.091***
Shortage Area indicator	0.70	-0.0051	-11.90***	-10.78***	0.04	0.030***	0.0530**	0.031***
	-6.58		-4.26		-0.01		0.00	

Government hospital indicator								
Specialty hospital indicator	-5.91		1.74		-0.03		-0.01	
Teaching hospital indicator	-15.94***		8.146**		-0.0911***		-0.0726***	
Percent uninsured	-71.15	-61.00***	-121.4***	-135.9***	0.24	0.20***	0.23	0.18***
Percent Medicare	-171.8*	-75.32***	39.35	28.73*	-0.711*	-0.29***	-0.764**	-0.29***
Percent Medicaid	47.99	-18.47	-24.87	-7.54	0.30	-0.033	0.14	-0.0146
Percent with commercial HMO coverage	-68.66	-41.42***	54.27*	36.99***	-0.392**	-0.21***	-0.488***	-0.26***
Percent with commercial PPO coverage	95.80***	72.26***	-81.96***	-92.28***	0.828***	0.53***	0.490***	0.28***
Percent with commercial POS coverage	-171.20	-137.7***	-58.07	-18.03	-0.53	-0.44***	-0.937**	-0.76***
Constant	375.2***	366.5***	325.6***	306.6***	1.146***	1.17***	1.685***	1.37***

*p<0.1, **p<0.05, ***p<0.01

Table 33 displays coefficients on market level variables regressed on spending, quantity, price, and mark-up. Note that several hospital based measures were not included in the HSA analysis due to data availability.

Guiding Question: How much does variation in spending (and its components) decrease when market level traits are taken into account?

Controlling for the menu of market level variables reduces the coefficient of variation modestly (5.3%) for spending, but significantly for quantity, price, and mark-up (26.0%, 22.8%, and 26.7%). The substantial negative correlation between price and quantity explains why the variation in spending does not decrease significantly even though the variation in its components (price and quantity) does decrease significantly.

Table 34. Variation in Spending, Quantity, Price, and Mark-Up after Controlling for Market Level Measures

		Spending	Quantity	Price	Mark-Up
Cluster 2 (age, sex, and health status). No market level covariates.	Coefficient of Variation	0.14	0.09	0.19	0.16
Cluster 2 with market level covariates.	Coefficient of Variation	0.13	0.06	0.15	0.12
	Percent reduction in the coefficient of variation compared to Cluster 2 without market level covariates.	5.3%	26.0%	22.8%	26.7%

Table 34 displays measures of variation for the elements of spending with and without controlling for market level variables.

Guiding Question: What is the relationship between market level traits and quality?

The insurance market appears to be significantly associated with quality, as is malpractice risk.

Table 35a. Second-Order Quality Regression Output

	Opportunity Weighted Composite Restricted Measures	Indicator Average Composite Restricted Measures	Preventable Admissions Composite	Preventable chronic admissions composite	Preventable acute admissions composite
R-Squared	0.29	0.49	0.47	0.43	0.49
Population (ln)	-0.00175	-0.00118	1.72e-06	6.12e-05	-5.95e-05*
Physicians Per 1,000	0.00274	0.00118	-4.70e-05	-2.31e-05	-2.39e-05
Ratio of PCP to specialists	0.0174	-0.00501	0.000277	6.14e-05	0.000216
Hospital beds per 1,000	0.000366	-0.00117	0.000238***	0.000109***	0.000128***
HHI (based on concentration of hospital beds)	2.94e-06**	6.23e-07	-3.93e-08	-3.00e-09	-3.63e-08**
Malpractice GPCI	-0.0165***	-0.0196***	0.000918***	0.000645***	0.000273***

Shortage Area indicator	0.00491	0.00935***	-0.000139	-5.94e-05	-7.98e-05
Government hospital indicator	-0.00191	-0.00484**	1.40e-05	3.51e-05	-2.11e-05
Specialty hospital indicator	0.00796	0.00254	-7.03e-05	-7.58e-05	5.52e-06
Teaching hospital indicator	0.00215	-0.000662	0.000190	0.000174*	1.59e-05
Percent uninsured	-0.105**	-0.157***	-0.000153	-0.00111	0.000962
Percent Medicare	0.200***	0.0191	-0.00122	0.000111	-0.00133
Percent Medicaid	-0.126***	-0.155***	0.00523***	0.00404***	0.00118**
Percent with commercial HMO coverage	0.0455	-0.0175	-0.00280**	-0.00194***	-0.000860*
Percent with commercial PPO coverage	0.186***	0.0843***	-0.00487***	-0.00302***	-0.00185***
Percent with commercial POS coverage	-0.120	-0.145***	-0.00510*	-0.00156	-0.00354***
Constant	0.692***	0.853***	0.00305***	0.000925	0.00213***

*p<0.1, **p<0.05, ***p<0.01

Table 35a displays coefficients on market level variables regressed on spending, quantity, price, and mark-up.

Table 35b. Second-Order Quality Regression Output, Continued

	Patient safety composite	Readmissions
R-Squared	0.33	0.16
Population (ln)	0.000228***	0.00137**
Physicians Per 1,000	2.21e-05	0.00106**
Ratio of PCP to specialists	0.000166	0.00379
Hospital beds per 1,000	1.86e-05	9.13e-05
HHI (based on concentration of hospital beds)	2.35e-08	4.68e-07
Malpractice GPCI	0.000293***	7.12e-05

Shortage Area indicator	4.05e-05	-0.000271
Government hospital indicator	-3.12e-05	1.21e-05
Specialty hospital indicator	-0.000152	0.000499
Teaching hospital indicator	3.43e-05	0.000789
Percent uninsured	-0.00259***	-0.0534***
Percent Medicare	-0.00210*	-0.0316*
Percent Medicaid	0.000155	-0.0111
Percent with commercial HMO coverage	-0.00169***	-0.0177*
Percent with commercial PPO coverage	-0.000422	-0.00165
Percent with commercial POS coverage	0.00125	-0.00377
Constant	0.000780	0.0712***

*p<0.1, **p<0.05, ***p<0.01

Table 35b displays coefficients on market level variables regressed on spending, quantity, price, and mark-up.

Conclusion

Health care spending among commercially insured beneficiaries varies widely, with the top 10th percentile spending 39% more than the bottom 10th percentile on medical services, without adjusting for age, sex, or other demographic or insurance related covariates. A small proportion of this can be explained by variation in covariates such as age, sex, and health status, although substantial variation persists. Input prices do not explain a sizable proportion of variation either.

We observed substantial variation in both price and utilization across markets, although price effects dominated in a model that controls for health status, age, and sex. This said, due to the negative covariance between price and quantity, adjusting spending for prices does not markedly decrease the extent of variation.

Individual utilization measures, such as counts of inpatient admissions, office visits, and emergency department visits, vary more across HRRs than do either aggregate measure, but are only weakly correlated with one another. This suggests that areas that have high utilization rates of one type of service do not have high rates of utilization for others (i.e., there may be a substitution effect across various types of services).

Quality tends to vary less than spending across HRRs, but varies markedly depending on the measure. In general, areas that perform well on one measure are no more likely to perform well on another. Moreover, because the variation in spending is not strongly related to quality measures, the benefit from greater spending is unclear.

Limitations

The MarketScan population is comprised of enrollees in large employers and group health plans. These individuals may not be representative of the entire commercial population. For instance, it is likely that people who are insured in the small group market may have different consumption patterns. In addition, there are several variables that we do not capture that may play a significant explanatory role. For instance, it is possible that differences in patient preferences explain some portion of the variation.

In addition, there is potential for some level of measurement error due to the imprecise nature of claims data. The imputation strategy for capitated claims may not accurately reflect the actual price of services, which could lead to misestimates in area level results, especially for regions with high proportions of capitation. Sensitivity analysis suggests that areas with high levels of capitation are sensitive to this strategy compared to one that drops capitated enrollees, but the general conclusion about the extent of variation is not affected (see Sensitivity Analysis 2)

Similarly, the method of distributing spending associated with missing procedure codes across other services on the same day may falsely increase the estimated price of certain services. However, sensitivity analysis indicates that this strategy does not have an impact on area level estimates of quantity or price (see Sensitivity Analysis 1).

There are also limitations to the measurement of certain covariates. In particular, the two socioeconomic measures are based at the population level and may estimate individual characteristics poorly. In addition, the measure of health status is explicitly based on consumption of health care, which may not accurately capture the true underlying health status of the population. We plan to test the correlation of claims-based health status measures with NHANES measures as part of a supplemental analysis.

Future Research

Several extensions to the existing research would be useful. First, and most importantly, by design this analysis was descriptive, relating area level spending to individual and market traits. Because the analysis was not designed to identify causal relationships it is difficult to assess how spending would change in response to various policy options. For example, we do not explicitly address how spending among the commercially insured would change in response to changes in Medicare (or Medicaid) payment rates. As a result it is difficult to extrapolate from this analysis to assess how spending by the commercially insured would respond to a value modifier. Similarly, we do not address how other changes in policy (e.g., expansion of Accountable Care Organizations, changes to MA payment rates, or expansion of Medicaid eligibility) will affect spending. Finally, we do not examine how other market activities (e.g. provider mergers or insurer competition) might affect commercial spending.

Secondly, and related to our lack of causal analysis, we were not able to assess certain important market dynamics, such as the extent to which differences in site of care (which can affect the prices for any given service because often prices are higher for services provided in hospital outpatient departments relative to when provided in physician offices) affect geographic variation. Similarly, we do not explicitly examine spillovers between sectors. For example, MedPAC has presented some evidence suggesting that pressure from commercial payers can reduce Medicare spending and certainly Medicare spending can affect spending by commercial payers. We have not explored these issues in this work.

Thirdly, this research is predicated on the idea that geography is a meaningful unit of analysis for policy. While it is certainly the case that geographic variation research can highlight important issues (e.g. uncertainty in practice patterns, existence of waste, the importance of price variation), but that does not imply that geography is the appropriate unit for policy intervention. Providers within any geographic area may behave differently. Some may be efficient and others wasteful. Although we were able to investigate the sensitivity of our results to different geographic units, and quantify variation across HSAs within HRRS, we were not able to identify individual providers or provider systems. As a result, we cannot comment on provider level variation. There are many reasons to believe that provider systems are a better unit for policy interventions. Incentives will operate more directly at the provider system level and interventions at that level may avoid the prisoners dilemma associated with intervening at the geographic level. When the intervention is so broad, individual providers may find it in their interest to practice aggressively even if collectively they entire group of providers in an area have an incentive to practice conservatively.

Fourthly, this research focused on assessing the variation in spending at a point in time. We did not address variation in spending growth. Over time, spending growth represents a bigger policy challenge than the variation in the level of spending. It is likely that the factors driving the variation in spending differ from those driving variation in spending growth. Our analysis cannot speak to the latter issue.

Finally, there are a number of technical issues that arose which deserve further attention. These include the difficulty in separating variation in true health status from variation in coding. Our NHANES analysis, delayed by data access issues, will shed some light on that issue, but will likely not be definitive. A better understanding of the observed inverse correlation between price and quantity would also be valuable because in the health care sector one would not expect the relationship to be as strong as we observed. We conducted several sensitivity analyses that suggest this finding is not simply a measurement issue, but a stronger explanation would be valuable. Further work on variation in quality would also be useful. Our quality measures are limited and aggregation of quality is difficult. More work in this area would be valuable.

Acknowledgements

None of this would have been possible without the excellent team of people working at every level of this project. In particular, we would like to thank Jim Livingston and Yanmei Liu for expert programming assistance, and Emily Corcoran for project management. We would also like to thank Emily Ehrlich, Amanda Farr, Kay Miller, Brian Moore and Gary Allen at Thomson Reuters for their expert guidance and collaboration when working through the extensive MarketScan database. Finally, we would like to thank the Institute of Medicine staff for their coordination of this research effort, and the Committee for providing invaluable methodological feedback.

Appendix

Coefficients from two OLS regression models are presented in Appendix Table 1. Both models use input price adjusted medical spending as the dependent variable. Cluster 2 contains controls for age, sex, enrollment, year, and health status. Cluster 6 includes age, sex, enrollment, year, and insurance variables.

Appendix Table 1: HRR Medical Regression Coefficients

	Cluster 2	Cluster 6
Sample Size (Drug Sample)	112,903,225	103,196,815[a]
R-squared	0.08	0.01
Age/ sex included?	Yes	Yes
Partial year enrollment	1 Models adjusted for age, sex, and health status.,414 (51.89)	1,765 (60.42)
2007 Dummy	-293.4 (18.17)	-465.7 (19.36)
2008 Dummy	-275.2 (13.38)	-324.8 (15.56)
DxCG	3,617 (38.5)	-
HMO[b]	-	-192.4 (53.93)
PPO	-	-51.35 (42.24)
CDHP/HDHP	-	-351 (49.18)
Indemnity	-	-209.3 (60.12)
Avg Brand Rx OOP	-	-3.25 (1.90)
Avg Generic Rx OOP	-	-7.08 (5.81)
Avg Office Visit OOP	-	-11.44 (1.56)
Avg Emergency Department OOP	-	0.02 (0.14)
Avg Inpatient Admission OOP	-	-0.04 (0.02)

Standard errors in parenthesis.
Appendix Table 1 depicts the coefficient estimates of covariates in cluster 2 and cluster 6 models using input price adjusted medical spending as the dependent variable.

[a] Sample size in cluster 5 is ~10 million less than other models due to lack of benefit generosity data.
[b] Insurance plan type compared to Point of Service Plans.

Sensitivity Analysis 1: Not Apportioning Claims that are Missing Procedure Codes

A substantial number of outpatient claims are not associated with procedure codes, often because they are paid as a separate facility bill. In order to more accurately capture the price of each service, we spread spending on claims that are missing procedure to other claims on the same day. This procedure does not alter spending results because all dollars are aggregated to the person-year level, regardless of the procedure associated with the claim. However, the apportioning method could affect both input price adjustment and the aggregate quantity measure because both rely on procedure-specific information. In the case of input price adjusted spending, apportioning causes some claims to have greater associated spending which are then differentially adjusted based on Relative Value Units. There is more potential for substantial differences in the aggregate quantity measure, as quantity is weighted by the average spending on individual procedures. The comparison of input price adjusted spending models calculated with and without the apportioning step is displayed in Appendix Table 2, and aggregate quantity models displayed in Appendix Table 3.

Appendix Table 2. Sensitivity of input price adjusted HRR medical spending estimates to not distributing spending on claims that are missing procedure codes

	With Spending Distributed	Without Spending Distributed
Sample Size	112,903,225	112,903,225
Average	3,313	3,302
Standard Deviation	410	404
Median	3,269	3,259
Ratio of 25th percentile to the 75th percentile	117.2%	116.7%
Ratio of 10th percentile to the 90th percentile	134.7%	134.3%
Correlation	-	>0.99

Appendix Table 2 compares summary statistics across the 306 HRRs for input price adjusted medical spending between our primary method of apportioning claims without procedure codes and a method that does not apportion claims. Models adjusted for age, sex, and health status.

Appendix Table 3. Sensitivity of HRR medical aggregate quantity estimates to not distributing spending on claims that are missing procedure codes

	With Spending Distributed	Without Spending Distributed
Sample Size	112,903,225	112,903,225
Average	3,311	3,129
Standard Deviation	294	289
Median	3,327	3,137
Ratio of 25th percentile to the 75th percentile	111.5%	110.5%
Ratio of 10th percentile to the 90th percentile	125.1%	124.4%
Correlation	-	0.97

Appendix Table 3 presents summary statistics of the aggregate quantity measure across the 306 HRRs comparing the primary method of apportioning claims without procedure codes and a method that does not apportion claims. Models adjusted for age, sex, and health status.

The method of distributing spending makes almost no difference in our estimates of input price adjusted medical spending. The two methods are correlated greater than 0.99 for input price adjusted spending. As expected, the aggregate quantity measure has a lower average spending estimate because the apportioning procedure increase the average spending associated with each procedure. However, the correlation between the HRR level quantity estimates is 0.97, indicating that our conclusions are not sensitive to this method.

Sensitivity Analysis 2: Dropping Enrollees with Capitated Claims

For claims that are paid under a capitated contract, we impute spending based on the non-capitated equivalent national price multiplied by the area-level price index. While only around 5 percent of enrollees have any capitation, the penetration of capitation varies substantially across HRRs, with a handful of markets having a quarter or more of enrollees with at least one capitated claim in a year. To test the sensitivity of this imputation approach, we computed a model that dropped these enrollees with capitation. Appendix Table 4 displays the comparison of input price adjusted medical spending.

Appendix Table 4. Sensitivity of input price adjusted medical spending estimates to capitation strategy

	Imputation Method	Omission Method
Sample Size	112,903,225	107,836,573
Average	3,313	3,279
Standard Deviation	410	391
Median	3,269	3,243
Ratio of 75th percentile to the 25th percentile	117.2%	115.6%
Ratio of 90th percentile to the 10th percentile	134.7%	134.0%
Correlation	-	0.96

Appendix Table 4 compares summary statistics across the 306 HRRs for input price adjusted medical spending between our primary method of imputing capitated values and the alternative of omitting enrollees with capitated claims. Models adjusted for age, sex, and health status.

Using the alternative approach of dropping enrollees with capitated claims does not significantly impact the magnitude of variation in input price adjusted medical spending. In addition, the correlation between the two approaches for input price adjusted medical spending is 0.96, indicating that the strategy does alter the conclusions of the study. However, as a handful of markets have high proportions of capitated enrollees, the impact of this approach may significantly impact their spending relative to other areas. For the 25 HRRs that have 20% or more of their enrollees with capitated claims, 7 change quartile rank.

Sensitivity Analysis 3: Omitting Enrollees without Drug Data

Around 20 percent of enrollees do not have drug data available. We do not observe in the data whether these individuals lack drug coverage, or if they have drug coverage but that the data is not provided to MarketScan. Our primary approach is to estimate area-level drug spending in a separate model that includes only those with coverage, and to add the results to those estimated on medical spending. This implicitly assumes that those without drug coverage data have similar drug spending as those with drug coverage data. To test the similarity of these two groups, we compare input price adjusted medical spending between those that have drug coverage data and those that do not with the full sample. The results of this sensitivity are presented in Appendix Table 5.

Appendix Table 5. Sensitivity of input price adjusted HRR medical spending estimates to excluding individuals without drug coverage data.

	Sample excluding enrollees without drug coverage	Sample including only enrollees without drug coverage	Full Sample
Sample Size	86,788,820	26,114,405	112,903,225
Average	3,392	2,984	3,313
Standard Deviation	416	403	410
Median	3,359	2,955	3,269
Ratio of 75th percentile to the 25th percentile	116.9%	121.2%	117.2%
Ratio of 90th percentile to the 10th percentile	133.0%	141.4%	134.7%
Correlation with the full sample	0.76	0.98	-

Appendix Table 5 compares summary statistics across the 306 HRRs for input price adjusted medical spending between those with drug coverage data and those lacking drug data. Models adjusted for age, sex, and health status.

Average input price adjusted medical spending among enrollees with drug coverage, holding age, sex, and health status constant, is around $400 dollars higher per year than for those without coverage. The amount of variation among HRRs is similar between the two groups, and the correlation between HRR estimates in between the sample excluding those without drug coverage is 0.98. It is likely that some unobserved factor is responsible for the difference in input price adjusted spending, although the high correlation between groups suggests that the conclusion of our results is not sensitive to including enrollees that lack drug coverage data.

Sensitivity Analysis 4: Differences between Data Providers

MarketScan data is provided by both employers and insurance health plans. There may be unobserved differences in the two populations that is not accounted for by age, sex, or health status. We find that enrollees in the health plan sample spend around $350 less than those in the employer samples, although variation is similar and the two samples are correlated 0.71.

Appendix Table 6. Differences between input price adjusted medical spending in the employer and health plan samples.

	Employer	Health Plan
Sample Size	63,734,486	49,168,739
Average	3,475	3,138
Standard Deviation	467	398
Coefficient of Variation	0.13	0.13
Ratio of 75th percentile to the	118.3%	118.0%

25th percentile		
Ratio of 90th percentile to the 10th percentile	138.5%	137.7%
Correlation	-	0.71

Appendix Table 6 compares the extent of variation in data provided by employers and data provided by health plans. Models adjusted for age, sex, and health status.

Sensitivity Analysis 5: Shrinking Estimates

In order to account for the inherent increased variability of small samples, we shrink estimates to the national mean according to their sample size and standard deviation. Appendix Table 7 shows that this procedure does not affect results or conclusions, due to the relatively large sample in each HRR.

Appendix Table 7. Differences between HRR estimates that are shrunk according to sample size and standard deviation and estimates that are not shrunk

	Not Shrunk	Shrunk
Average	3,315	3,313
Standard Deviation	418	410
Median	0.13	0.12
Ratio of 75th percentile to the 25th percentile	117.3%	117.2%
Ratio of 90th percentile to the 10th percentile	135.6%	134.7%
Correlation	-	>0.99

Appendix Table 7 compares results between shrinking area level estimates of input price adjusted medical spending and not shrinking area level estimates. Models adjusted for age, sex, and health status.

Supplemental Analysis 1: HCUP

State-level hospital admission rates from the MarketScan Database were compared to discharge rates for the population of privately insured individuals from State Inpatient Databases (SID) collected by the Health care Cost and Utilization Project (HCUP). HCUP SID data include all-payer, discharge-level information for most states.

A comparison between MarketScan and HCUP helps to assess the representativeness of the MarketScan database, although the comparison to HCUP is not expected to be exact. MarketScan is a convenience sample of health care experience for individuals with insurance coverage provided by large firms and medium-sized insurance companies. HCUP, on the other hand, includes the health care experience of individuals across the continuum of private insurance including large group plans, small group plans, and individual plans.

Twenty-nine states in 2009 are the focus of comparison: Arizona, Arkansas, California, Colorado, Florida, Hawaii, Kansas, Kentucky, Massachusetts, Michigan, Minnesota, Missouri, Nebraska, Nevada, New Jersey, New Mexico, New York, North Carolina, Oklahoma, Oregon, Rhode Island, Tennessee, Texas, Utah, Vermont, Washington, West Virginia, Wisconsin, and Wyoming. These states met the MarketScan data release provisions for release of descriptive statistics in geographic areas and discharge information was also reported in publicly-available HCUP data (HCUPnet). Survey estimates from the American

Community Survey by insurance type (i.e., private insurance) served as the denominator of the HCUP discharge rates.

To compare MarketScan admission rates to discharge rates reported in HCUP, MarketScan admission rates were standardized to the HCUP population in the same state using direct standardization. Specifically, two standardized rates were calculated, one based on 3 age strata (under 18, 18-44 and 45-64) and one based gender strata (see equation (1)). Discharge rates were not available by both age and gender in HCUPnet for privately insured individuals so these stratifiers were analyzed separately. Infants were excluded from the discharge or admission counts as births are typically recorded as an admission in MarketScan (many times the baby is grouped with the mother as the admission occurred at the same time) and as a discharge in HCUP (which often reports a discharge separate from the mother for each infant). As a result, counts of infant admissions are much larger in HCUP than in MarketScan.

$$(1)\ DSR = \sum_{i=1}^{k}(P_i \ * \ r_i)$$

Where DSR = directly standardized rate where *i* is the k age or gender strata, P_i = the proportion of the standard population (from Census), and r_i = the crude admission rate

Results

MarketScan enrollees comprise 19.7% (23,291,119) of the 118,263,000 privately-insured individuals in these 29 states. Overall, the HCUP discharge rate is 52.4 per 1,000 and the MarketScan standardized admission rate is 49.85 per 1,000 (95% confidence interval 49.8, 49.9). For 25 states, the HCUP discharge rate exceeds the MarketScan rate, in 3 states the HCUP rate falls into the 95% confidence interval of the MarketScan rate, and in 1 state (Wyoming) the HCUP rate is below the MarketScan rate. As stated previously, the comparison is not expected to be exact because HCUP includes all privately insured individuals and MarketScan includes the experience of individuals with health care coverage provided by large and medium sized firms.

Correlation coefficients between age strata-specific rates are summarized in Table 8.

Appendix Table 8. Correlation coefficients, Age strata, MarketScan and HCUP State Inpatient Databases (n=29)

Age Group	Correlation
Age <18	0.6805
Age 18-44	0.6789
Age 45-64	0.7884

All $p<0.001$

Appendix table 8 displays the correlation between MarketScan admissions and HCUP admissions by state for three age groups.

Correlation coefficients between state-level, MarketScan directly standardized rates are summarized in Table X2. The sex-adjusted correlation coefficient is much lower than the age adjusted correlation, as HCUP includes individuals over age 65 and MarketScan does not. Also, infant admissions are included in the sex-adjusted rates.

Appendix Table 9. Correlation coefficients, MarketScan directly standardized rates and HCUP State Inpatient Databases (n=29)

	Correlation
Age adjusted	0.7326
Sex adjusted	0.5810

All p<0.001

Appendix table 8 displays the correlation between age and sex adjusted MarketScan admissions and adjusted HCUP admissions by state.

Bibliography

1. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care. *Annals of Internal Medicine.* 2003;138(4):288-299.

2. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care. *Annals of Internal Medicine.* 2003;138(4):273-288.

3. MedPAC. Medicare Payment Advisory Commission Report to the Congress: Regional Variation in Medicare Service Use. 2011; http://www.medpac.gov/documents/Jan11_RegionalVariation_report.pdf.

4. Wennberg J, Gittelsohn. Small area variations in health care delivery. *Science.* 1973;182(117):1102-1108.

5. McLaughlin CG, Normolle DP, Wolfe RA, McMahon LF, Jr, , Griffith JR. Small-area variation in hospital discharge rates. Do socioeconomic variables matter? *Medical Care.* 1989;27(5):507-521.

6. Baker LC, Fisher ES, Wennberg JE. Variations in hospital resource use for Medicare and privately insured populations in California. *Health Affairs (Millwood).* 2008;27(2):w123-134.

7. Wennberg JE, Fisher ES, Stukel TA, Sharp SM. Use of Medicare claims data to monitor provider-specific performance among patients with severe chronic illness. *Health Affairs (Millwood).* 2004;Suppl(Web Exclusives):VAR5-18.

8. Hirth RA, Turenne MN, Woods JD, et al. Predictors of Type of Vasuclar Access in Hemodialysis Patients. *Journal of the American Medical Association.* 1996;276(16):1303-1308.

9. Baicker K, Buckles KS, Chandra A. Geographic variation in the appropriate use of cesarean delivery. *Health Aff (Millwood).* Sep-Oct 2006;25(5):w355-367.

10. Baicker K, Chandra A. Medicare Spending, The Physician Workforce, and Beneficiaries' Quality of Care. *Health Affairs.* 2004;23:184-197.

11. Centers for Medicare & Medicaid Services. National Health Expenditure Data. 2011; http://www1.cms.gov/NationalHealthExpendData/07_NHEA_Related_Studies.asp#TopOfPagedownloads/tables.pdf. Accessed April 25, 2010.

12. Chernew ME, Sabik LM, Chandra A, Gibson TB, Newhouse JP. Geographic Correlation Between Large-Frim Commercial Spending and Medicare Spending. *The American Journal of Managed Care.* 2010;16(2):131-138.

13. Wennberg JE, Fisher ES, Goodman DC, Skinner JS. *Tracking the care of patients with severe chronic illness. The Dartmouth Atlas of Health Care 2008.* Lebanon, NH: The Dartmouth Institute for Health Policy and Clinical Practice;2008.

14. Wennberg JE, Fisher ES, Stukel TA, Skinner JS, Sharp SM, Bronner KK. Use of hospitals, physician visits, and hospice care during last six months of life among cohorts loyal to highly respected hospitals in the United States. *Bmj.* Mar 13 2004;328(7440):607.

15. Dimick JB, Staiger DO, Birkmeyer JD. Ranking hospitals on surgical mortality: the importance of reliability adjustment. *Health Services Research.* 2010;45(6):1614-1629.

16. Agency for Healthcare Research & Quality. Quality indicator user guide: prevention quality indicators (PQI) composite measures. 2012; http://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/V44/Composite_User_Technical_Specification_PQI%20V4.4.pdf. Accessed 4/20/2012.

17. Starfield B, Shi L, Macinko J. Contribution of Primary Care to Health Systems and Health. *The Milbank Quarterly.* 2005;83(3):457-502.

18. Kaiser Family Foundation, Health Research and Educational Trust (HRET). Employer Health Benefits: 2007 Annual Survey. 2007; http://www.kff.org/insurance/7672/upload/76723.pdf. Accessed May, 2011.