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

Abe Dunn, Adam Hale Shapiro, and Eli Liebman

September 27, 2013

Abstract

This appendix covers the following topics: (1) additional technical discussion of the data and methodology; (2) alternative indexes and decompositions; (3) service-specific indexes; (4) population based estimates (continued); (5) medical-care indexes and quality; and (6) list of additional robustness checks.

1 Moving from the MarketScan Data to the Analysis: A Technical Discussion

1.1 Measuring the Quantity of Service by Service Type (continued)

There are a few methodological points that are important to consider. First, a small fraction of the claim lines (less than 5% of the claims observations for non-facility claim lines) are missing procedure codes. For these procedures we take the average price of the missing procedure codes for that service and disease. The results presented here do not change when alternative methods for calculating utilization are used. For instance, we obtain similar results when we drop claim lines that are missing procedure codes. Also, some claim lines have negative billing amounts that represent corrections to bills. These negative amounts often represent corrections and are left in the analysis so that the corrected billing amounts are applied in our analysis.[1] We treated zero, positive, and negative claim line amounts similarly for all service categories:

1. If the payment amount was positive, then we allocate the quantity specified above to that claim.

2. If the payment to the provider (i.e., the allowed reimbursement amount) was zero, then the associated quantity was set to zero. Essentially, the claim line is dropped from out analysis.

3. If the payment amount was less than zero, this indicates a claims reversal, which makes it important to incorporate in the analysis. In this case, we set the quantity amount equal to the amount described above, multiplied by negative 1, which makes the corresponding proper adjustment to the total quantity.

Since our unit of analysis can be as detailed as 11-digit NDC codes or CPT codes and modifiers, another concern is that some services may appear infrequently. To address this concern we use the maximum amount of potential data to construct the quantity of service measure. The population used to construct the RVUs was all enrollees in our data that were not enrolled in a capitated plan for any month in a given year from 2003 to 2007. This broader population had roughly 12 million enrollees in 2003 and 28 million enrollees in 2007. Using a broader population allows us to get more information on prices of various services that were used in our studied population.

1.2 Weights

We use the distribution from the MEPS U.S. commercially insured, under-65 population to fix our population for each MSA. We used the MEPS Household Component online query tool (available at http://meps.ahrq.gov/mepsweb/data_stats/MEPSnetHC.jsp) to determine the age and sex distribution for this population. We use the following age categories to construct our weights: 0, 1-4, 5-17, 18-24, 25-34, 35-44, 45-54, 55-64. Generally, we found that our results weren't sensitive to changes in the age groupings.

Our sample is calculated each year by selecting all enrollees who were not enrolled in a capitated plan in that year, who have prescription drug coverage and who are enrolled

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

throughout the entire year. For this analysis, we also require that we observe enrollment 6 months prior to the observation year and 6 months after the observation year. We do this in order to observe a more complete medical history for each enrollee in the sample.[2] Since newborns cannot be enrolled throughout the year, if one of their parents is in our sample they remain in our sample. Using this sample we calculate the numbers of enrollees in each age and sex category for each MSA. Then for each MSA we define the person weight as the MEPS U.S. population counts of age and sex divided by our sample count of age and sex in each MSA. The weight for each enrollee is applied to his/her expenditures and episodes prior to aggregating.

1.3 Application of the ETG and MEG groupers

The ETG Symmetry grouper was run on the full sample of MarketScan data for all available years. This was done to attain the maximum accuracy on all available claims. We focus on completed episodes, rather than the cost of the disease per year. For example, we look at the expenditures on a completed episode of a pregnancy, rather than the amount spent on pregnancy for a calendar year.[3] We find that these two ways of analyzing disease expenditures are quite similar in practice when episodes are aggregated over a large population.

Later in this appendix we apply the MEG grouper from Truven Health, which is another disease episode grouper similar to the ETG Symmetry grouper. Following the recommendations of Truven Health, we ran the grouper on the data one year at a time. That is, for each enrollee, we run only a single calendar year of claims through the grouper.[4]

2 Alternative Indexes

To provide a robustness checks, this section presents results from three alternative indexes.

2.1 Alternative Grouping Methodology

The paper primarily focuses on a single methodology of grouping medical claims, the Symmetry ETG grouper from Optum. We selected the ETG grouper for several reason.

[2] We find similar results when we do not have the requirement that we observe an additional 6 months pre- and post-.

[3] The second approach is applied in Aizcorbe and Nestoriak (2011).

[4] For more information see McCurdy et al. (2008) and Rosen et al. (2012), which provide a comparisions of the ETG and MEG groupers.

First, the grouping methodology had been applied previously in other related works (e.g., Aizcorbe and Nestoriak (2011) and Dunn et al. (2012)). Second, we found the ETG software relatively easy to work with, which includes a more user-friendly interface. Third, the severity adjustment applied by the ETG grouper takes into account both the age and the sex of the patient, which are not considered by the MEG grouper. Finally, the ETG grouper allows for four categories of severity adjustment, compared to just two categories for the MEG grouper. Although these factors led us to apply the ETG grouper in the main text, we do not necessarily view the ETG grouper as the preferred methodology. In this section, we present several of the results of the paper, but applying the MEG grouper rather than the ETG grouper.

Table A2 reports the same summary statistics for the MEG grouper as shown in Table 2 of the paper. The table shows the different disease categories, the spending for each disease category, the number of episodes, and dollars of spending per episode. Many of the episode categories are quite similar to those using the ETG grouper. Angina pectoris is very similar to heart disease and is one of the top expenditure categories for both the MEG and ETG. Similarly, preganancy ranks high on both lists. One notable difference is that a huge share of expenditure is allocated to preventative health services by the MEG grouper. It appears that the ETG grouper allocates many of the expenditures to specific disease categories, which are defined as "preventative" by the MEG grouper.

Table A2. Total Average Annual Expenditures and Average Number of Episodes - Weighted to U.S. Totals for Commercial Insurance - MEG Grouper

	Description	Severity	Total Dollars (Billions)	Episodes	Dollars Per Episode	Fraction of Spending
1	Encounter for Preventive Health Services	0	$22.16	62,886,711	$352.43	5.1%
2	Delivery	1	$19.32	1,963,548	$9,841.75	4.4%
3	Angina Pectoris, Chronic Maintenance	1	$10.99	2,340,805	$4,692.97	2.5%
	Angina Pectoris, Chronic Maintenance	2	$4.48	306,548	$14,627.63	1.0%
4	Essential Hypertension, Chronic Maintenance	1	$11.84	14,277,981	$829.21	2.7%
	Essential Hypertension, Chronic Maintenance	2	$1.61	1,331,923	$1,210.63	0.4%
5	Other Arthropathies, Bone and Joint Disorder	1	$10.66	17,271,162	$617.38	2.4%
	Other Arthropathies, Bone and Joint Disorder	2	$0.60	421,003	$1,426.33	0.1%
6	Diabetes Mellitus Type 2 and Hyperglycemia	1	$8.67	5,079,122	$1,707.70	2.0%
	Diabetes Mellitus Type 2 and Hyperglycemia	2	$2.18	677,335	$3,219.11	0.5%
7	Osteoarthritis, Except Spine	1	$10.56	3,823,485	$2,762.24	2.4%
8	Depression	1	$4.47	5,189,194	$861.65	1.0%
	Depression	2	$5.98	3,302,912	$1,809.81	1.4%
9	Neoplasm, Malignant: Breast, Female	1	$9.91	863,936	$11,466.16	2.3%
10	Osteoarthritis, Lumbar Spine	1	$8.81	1,766,812	$4,987.01	2.0%
11	Complications of Surgical and Medical Care	1	$3.75	634,398	$5,905.61	0.9%
	Complications of Surgical and Medical Care	2	$3.71	356,035	$10,421.42	0.8%
12	Other Spinal and Back Disorders: Low Back	1	$7.40	11,406,095	$648.78	1.7%
13	Sinusitis	1	$7.34	18,842,653	$389.59	1.7%
14	Asthma, chronic maintenance	1	$4.66	3,849,913	$1,209.51	1.1%
	Asthma, chronic maintenance	2	$2.19	1,011,111	$2,166.86	0.5%
15	Other Inflammations and Infections of Skin	1	$6.84	25,378,174	$269.40	1.6%
	Other		$269.39	325,623,442	$827.30	61.6%
	Total		$437.52	508,604,300	$860.24	100.0%

Table A4.1 is similar to Table 4 in the paper, but using MEG grouper categories. Table A4.1 shows that the results from the MEG grouper look similar to those from the ETG grouper along many dimensions. For instance, similar to results from the ETG grouper, looking at delivery we see that utilization differences are small, although price differences are large. For depression, we observe that price differences are relatively small, although the utilization differences are large. The weighted average variation measures, as shown at the bottom of Table A4.1, are also similar in magnitude to those based on the ETG grouper reported in Table 4.

Table A4.1 Sources of Price Variation Across MSAs by Disease - $MCE_{d,r}$, $SPI_{d,r}$ and $SUI_{d,r}$ - MEG

	Description	Severity	COV of $MCE_{d,r}$	COV of $SPI_{d,r}$	COV of $SUI_{d,r}$
1	Encounter for Preventive Health Services	0	0.15	0.10	0.07
2	Delivery	1	0.19	0.19	0.05
3	Angina Pectoris, Chronic Maintenance	1	0.23	0.18	0.18
	Angina Pectoris, Chronic Maintenance	2	0.31	0.28	0.19
4	Essential Hypertension, Chronic Maintenance	1	0.08	0.08	0.08
	Essential Hypertension, Chronic Maintenance	2	0.41	0.15	0.17
5	Other Arthropathies, Bone and Joint Disorder	1	0.10	0.09	0.10
	Other Arthropathies, Bone and Joint Disorder	2	0.31	0.21	0.28
6	Diabetes Mellitus Type 2 and Hyperglycemia	1	0.10	0.06	0.08
	Diabetes Mellitus Type 2 and Hyperglycemia	2	0.15	0.11	0.11
7	Osteoarthritis, Except Spine	1	0.22	0.15	0.16
8	Depression	1	0.27	0.08	0.22
	Depression	2	0.20	0.10	0.15
9	Neoplasm, Malignant: Breast, Female	1	0.18	0.14	0.14
10	Osteoarthritis, Lumbar Spine	1	0.25	0.17	0.14
11	Complications of Surgical and Medical Care	1	0.30	0.29	0.15
	Complications of Surgical and Medical Care	2	0.33	0.26	0.14
12	Other Spinal and Back Disorders: Low Back	1	0.18	0.11	0.14
13	Sinusitis	1	0.14	0.08	0.13
14	Asthma, chronic maintenance	1	0.11	0.07	0.11
	Asthma, chronic maintenance	2	0.21	0.21	0.20
15	Other Inflammations and Infections of Skin	1	0.10	0.07	0.09
	Weighted Average (Full Sample - 10,000 Episodes in the Data)		0.230	0.180	0.167
	Weighted Average (Only Diseases with 50,000 Episodes in the Data)		0.192	0.152	0.141

Table A6.1 corresponds to Table 6 of the paper, but applies the MEG grouper instead of the ETG grouper. The ranking of cities by MCE, SPI and SUI are quite similar. The correlation between the MEG-MCE and ETG-MCE is 0.96; the correlation between the MEG-SPI and ETG-SPI is 0.98; and the correlation between the MEG-SUI and ETG-SUI is 0.91. While the correlations are high, it is also worth noting that these two distinct methodologies arrive at different rankings. For example, Gary, Indiana is ranked 7th by the MEG-MCE, but is ranked 13th by the ETG-MCE.

Table A6.1 MSA Medical-Care Price Indexes and Variation in Indexes - MCE_r, SPI_r and SUI_r - MEG

MSA Name	Rank MCE_r	MCE_r	Rank SPI_r	SPI_r	Rank SUI_r	SUI_r
Salinas, CA	1	1.294	1	1.452	83	0.897
Milwaukee-Waukesha-West Allis, WI	2	1.279	4	1.257	38	1.004
Oakland-Fremont-Hayward, CA	3	1.204	3	1.285	73	0.933
Peoria, IL	4	1.191	9	1.11	6	1.085
Minneapolis-St. Paul-Bloomington, MN-WI	5	1.183	11	1.103	9	1.069
MSA in the Midwest	6	1.174	6	1.128	22	1.032
Fort Worth-Arlington, TX	7	1.167	10	1.107	15	1.042
MSA in the West	8	1.160	2	1.291	82	0.898
MSA in the Midwest	9	1.158	16	1.074	8	1.07
Dallas-Plano-Irving, TX	10	1.144	7	1.121	29	1.018
Indianapolis, IN	11	1.137	13	1.095	27	1.02
Houston-Sugar Land-Baytown, TX	12	1.133	20	1.064	12	1.054
Gary, IN	13	1.114	55	0.97	3	1.129
Miami-Miami Beach-Kendall, FL	14	1.095	22	1.053	25	1.025
Denver-Aurora, CO	15	1.087	33	1.021	13	1.051
Hartford-West Hartford-East Hartford, CT	71	0.911	53	0.976	71	0.937
Cleveland-Elyria-Mentor, OH	72	0.911	78	0.882	21	1.034
Lansing-East Lansing, MI	73	0.910	74	0.908	30	1.017
Augusta-Richmond County, GA-SC	74	0.898	69	0.931	62	0.961
Nassau-Suffolk, NY	75	0.894	26	1.035	85	0.876
Akron, OH	76	0.881	83	0.849	16	1.039
MSA in the South	77	0.878	66	0.941	72	0.934
Warren-Farmington Hills-Troy, MI	78	0.876	77	0.892	46	0.99
MSA in the Midwest	79	0.875	72	0.911	64	0.961
Pittsburgh, PA	80	0.875	84	0.83	10	1.069
Detroit-Livonia-Dearborn, MI	81	0.870	80	0.877	37	1.005
Kingsport-Bristol-Bristol, TN-VA	82	0.866	82	0.865	36	1.006
MSA in the South	83	0.864	75	0.901	65	0.956
Providence-New Bedford-Fall River, RI-MA	84	0.832	79	0.877	67	0.949
Youngstown-Warren-Boardman, OH-PA	85	0.792	85	0.799	43	0.994
mean		1.01		1.00		1.00
sd		0.10		0.06		0.07
COV		0.10		0.06		0.07
p10		0.89		0.92		0.91
p90		1.11		1.07		1.09
N		85		85		85

2.2 Alternative Base - Paasche Index

Another dimension in which the indexes may differ is the weighting of the price and utilization components. Rather than use the Laspeyres index, we examine estimates from the alternative Paasche index. In this case, the service price index (SPI) is then calculated as:

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

which measures the price difference in region r compared to the base region, using the utilization mix in region r. Similarly, the service utilization index (SUI) may be defined as:

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

which measures the utilization in region r compared to the base region, using the prices in region r. The variation using the Paasche index is shown in Table A4.2. One can see that many of the same qualitative findings hold. The overall measures of variation, shown along the bottom of the table, are quite similar to those of the main estimates in the paper. In addition, the qualitative findings for specific diseases are also quite similar. For instance, we see pregnancy with high service price variation and low service utilization variation, and we observe the reverse for depression (i.e., high utilization variation and low service price variation).

Table A4.2 Sources of Price Variation Across MSAs by Disease - $MCE_{d,r}$, $SPI_{d,r}$ and $SUI_{d,r}$ - Paasche Index

	Description	Severity	COV of $MCE_{d,r}$	COV of $SPI_{d,r}$	COV of $SUI_{d,r}$
1	Pregnancy, with delivery 1	1	0.18	0.17	0.04
	Pregnancy, with delivery 2	2	0.20	0.19	0.05
2	Joint degeneration, localized - back 1	1	0.18	0.14	0.14
	Joint degeneration, localized - back 2	2	0.28	0.19	0.18
	Joint degeneration, localized - back 3	3	0.29	0.22	0.18
3	Ischemic heart disease 1	1	0.22	0.17	0.17
	Ischemic heart disease 2	2	0.22	0.22	0.15
	Ischemic heart disease 3	3	0.33	0.24	0.17
4	Hypertension 1	1	0.10	0.09	0.11
	Hypertension 2	2	0.11	0.10	0.12
	Hypertension 3	3	0.14	0.13	0.12
	Hypertension 4	4	0.35	0.33	0.23
5	Diabetes 1	1	0.11	0.06	0.10
	Diabetes 2	2	0.18	0.10	0.14
	Diabetes 3	3	0.20	0.12	0.16
	Diabetes 4	4	0.21	0.17	0.14
6	Routine exam 1	1	0.15	0.12	0.06
7	Mood disorder, depressed 1	1	0.20	0.07	0.17
	Mood disorder, depressed 2	2	0.23	0.10	0.18
	Mood disorder, depressed 3	3	0.24	0.17	0.18
8	Malignant neoplasm of breast 1	1	0.20	0.15	0.18
	Malignant neoplasm of breast 2	2	0.27	0.19	0.20
9	Hyperlipidemia, other 1	1	0.10	0.08	0.09
10	Joint degeneration, localized - neck 1	1	0.19	0.13	0.17
	Joint degeneration, localized - neck 2	2	0.31	0.18	0.27
	Joint degeneration, localized - neck 3	3	0.30	0.24	0.21
11	Chronic sinusitis 1	1	0.17	0.08	0.13
	Chronic sinusitis 2	2	0.20	0.10	0.17
	Chronic sinusitis 3	3	0.25	0.18	0.18
12	Joint degeneration, localized - knee & l	1	0.25	0.17	0.17
	Joint degeneration, localized - knee & l	2	0.31	0.18	0.24
	Joint degeneration, localized - knee & l	3	0.29	0.22	0.28
13	Asthma 1	1	0.11	0.06	0.11
	Asthma 2	2	0.11	0.07	0.10
	Asthma 3	3	0.31	0.20	0.19
	Asthma 4	4	0.24	0.22	0.23
14	Joint derangement - knee & lower leg 1	1	0.26	0.18	0.20
	Joint derangement - knee & lower leg 2	2	0.21	0.22	0.16
15	Inflammation of esophagus 1	1	0.12	0.10	0.09
	Inflammation of esophagus 2	2	0.14	0.11	0.11
	Inflammation of esophagus 3	3	0.33	0.23	0.24
	Weighted Average (Full Sample - 10,000 Episodes in the Data)		0.23	0.18	0.17
	Weighted Average (Only Diseases with 50,000 Episodes in the Data)		0.19	0.16	0.14

The ranking of the Paasche MSA indexes is shown in Table A6.2. The Paasche indexes are actually quite similar to the Laspeyres-type SPI and SUI indexes used in the paper. The correlation between the Paasche-SPI and the Laspeyres-SPI is 0.98; and the correlation between the Paasche-SUI and Laspeyres-SUI is 0.95. Although some of the specific MSA rankings differ, the results are quite similar.

Table A6.2 MSA Medical-Care Price Indexes and Variation in Indexes - MCE_r, SPI_r and SUI_r - Paasche Index

MSA Name	Rank MCE_r	MCE_r	Rank SPI_r	SPI_r	Rank SUI_r	SUI_r
Milwaukee-Waukesha-West Allis, WI	1	1.288	2	1.271	26	1.025
Salinas, CA	2	1.245	1	1.404	80	0.899
MSA in the Midwest	3	1.240	5	1.161	5	1.080
Oakland-Fremont-Hayward, CA	4	1.219	3	1.261	51	0.980
Minneapolis-St. Paul-Bloomington, MN-WI	5	1.170	8	1.123	16	1.049
MSA in the Midwest	6	1.165	14	1.101	11	1.066
Fort Worth-Arlington, TX	7	1.143	10	1.115	27	1.025
Indianapolis, IN	8	1.138	7	1.125	22	1.029
Gary, IN	9	1.136	40	1.010	1	1.138
MSA in the West	10	1.133	4	1.261	79	0.902
Dallas-Plano-Irving, TX	11	1.127	9	1.123	39	0.999
Peoria, IL	12	1.121	13	1.101	35	1.004
Houston-Sugar Land-Baytown, TX	13	1.120	17	1.076	17	1.047
Miami-Miami Beach-Kendall, FL	14	1.102	21	1.061	21	1.033
Denver-Aurora, CO	15	1.102	28	1.033	7	1.075
Las Vegas-Paradise, NV	71	0.902	52	0.977	83	0.882
Nassau-Suffolk, NY	72	0.902	37	1.018	82	0.883
Louisville, KY-IN	73	0.901	56	0.963	65	0.952
MSA in the South	74	0.899	25	1.037	85	0.860
MSA in the South	75	0.898	61	0.949	64	0.952
Memphis, TN-MS-AR	76	0.891	44	0.997	84	0.878
Providence-New Bedford-Fall River, RI-MA	77	0.884	83	0.864	36	1.004
Kingsport-Bristol-Bristol, TN-VA	78	0.881	81	0.880	30	1.013
Warren-Farmington Hills-Troy, MI	79	0.872	79	0.881	52	0.978
MSA in the South	80	0.871	66	0.940	76	0.914
MSA in the Midwest	81	0.865	69	0.926	67	0.940
Augusta-Richmond County, GA-SC	82	0.862	68	0.933	73	0.931
Detroit-Livonia-Dearborn, MI	83	0.861	82	0.879	58	0.968
MSA in the South	84	0.844	74	0.905	71	0.933
Youngstown-Warren-Boardman, OH-PA	85	0.793	85	0.820	54	0.976
mean		1.000		1.009		0.992
sd		0.098		0.100		0.060
COV		0.098		0.099		0.060
p10		0.884		0.892		0.911
p90		1.136		1.120		1.069
N		85		85		85

In this paper we chose to present an additive decomposition, which leaves a cross-term. Of course, numerous alternative indexes and decompositions may be applied. For instance, one can also see that $SPI^{Paasche} \cdot SUI^{Laspeyres} = MCE$ or $SPI^{Laspeyres} \cdot SUI^{Paasche} = MCE$ is another potential decomposition that is exact and excludes a cross-term. One advantage of presenting the $SPI^{Laspeyres}$ and $SUI^{Laspeyres}$ is that the Paasche indexes may be derived by readers (e.g. $SPI^{Paasche}. = \frac{MCE}{SUI^{Laspeyres}}$). Although we leave a cross-term, it should be noted that the cross-term is fairly small. This can be seen by examining Table A6.3, which replicates Table 6 in the paper but includes the cross-term.

Table A6.3 MSA Medical-Care Price Indexes and Variation in Indexes - MCE_r, SPI_r and SUI_r

MSA Name	Rank MCE_r	MCE_r	Rank SPI_r	SPI_r	Rank SUI_r	SUI_r	Cross Term
Milwaukee-Waukesha-West Allis, WI	1	1.288	3	1.254	33	1.011	0.023
Salinas, CA	2	1.245	1	1.391	84	0.888	-0.035
MSA in the Midwest	3	1.240	6	1.149	10	1.069	0.022
Oakland-Fremont-Hayward, CA	4	1.219	4	1.251	57	0.971	-0.003
Minneapolis-St. Paul-Bloomington, MN-WI	5	1.170	8	1.122	16	1.047	0.002
MSA in the Midwest	6	1.165	15	1.093	13	1.058	0.014
Fort Worth-Arlington, TX	7	1.143	10	1.114	28	1.022	0.007
Indianapolis, IN	8	1.138	12	1.105	34	1.010	0.022
Gary, IN	9	1.136	44	0.997	4	1.124	0.015
MSA in the West	10	1.133	2	1.262	80	0.902	-0.031
Dallas-Plano-Irving, TX	11	1.127	7	1.129	36	1.004	-0.006
Peoria, IL	12	1.121	9	1.120	25	1.024	-0.023
Houston-Sugar Land-Baytown, TX	13	1.120	20	1.068	17	1.040	0.012
Miami-Miami Beach-Kendall, FL	14	1.102	22	1.063	20	1.036	0.003
Denver-Aurora, CO	15	1.102	33	1.026	11	1.065	0.011
Las Vegas-Paradise, NV	71	0.902	34	1.026	77	0.924	-0.047
Nassau-Suffolk, NY	72	0.902	30	1.031	83	0.890	-0.020
Louisville, KY-IN	73	0.901	65	0.947	68	0.936	0.017
MSA in the South	74	0.899	24	1.051	85	0.870	-0.023
MSA in the South	75	0.898	64	0.948	65	0.949	0.001
Memphis, TN-MS-AR	76	0.891	37	1.018	82	0.895	-0.021
Providence-New Bedford-Fall River, RI-MA	77	0.884	82	0.878	27	1.023	-0.018
Kingsport-Bristol-Bristol, TN-VA	78	0.881	81	0.878	31	1.014	-0.011
Warren-Farmington Hills-Troy, MI	79	0.872	78	0.896	39	0.998	-0.022
MSA in the South	80	0.871	60	0.953	73	0.927	-0.008
MSA in the Midwest	81	0.865	70	0.925	67	0.937	0.003
Augusta-Richmond County, GA-SC	82	0.862	71	0.925	76	0.924	0.013
Detroit-Livonia-Dearborn, MI	83	0.861	79	0.893	48	0.985	-0.017
MSA in the South	84	0.844	77	0.906	70	0.933	0.006
Youngstown-Warren-Boardman, OH-PA	85	0.793	85	0.821	54	0.977	-0.005
mean		1.000		1.018		0.996	-0.008
sd		0.098		0.097		0.064	0.022
COV		0.098		0.095		0.064	-
p10		0.887		0.915		0.920	0.011
p90		1.142		1.124		1.073	-0.031
N		85		85		85	85

2.3 Non-Index Decomposition

Our paper focuses on an index based approach, but an alternative to this approach is to conduct a non-index based decomposition. A couple of advantages of this approach is that it is also additive and it is exact with no interactive terms. This is similar to the work of Bundorf et al. (2009). Using the notation from the paper, the exact decomposition is:

$$c_{d,r} - c_{d,B} = q_{d,B}(p_{d,r} - p_{d,B}) + p_{d,r}(q_{d,r} - q_{d,B})$$

Where $c_{d,r} - c_{d,B}$ is the dollar amount difference in the episode price; $q_{d,B}(p_{d,r} - p_{d,B})$ is the component of that difference attributable to service price; and $p_{d,r}(q_{d,r} - q_{d,B})$ is the component of that difference attributable to service utilization. The value $c_{d,B} = q_{d,B} \cdot p_{d,B}$ is \$815. Table A6.3 shows the components of these decompositions for each city and the ranking.[5]

[5] For this decomposition, we focus on only the aggregate MSA indexes. The COV is not reported for this decomposition because the mean has already been subtracted.

Table A6.4 MSA Medical-Care Episode Differences and the Price and Utilization Components - Non-Index Based Decomposition

MSA Name	Rank	Average Episode Difference	Rank	Price Contribution	Rank	Utilization Contribution
Milwaukee-Waukesha-West Allis, WI	1	$245	3	$216	20	$29
Salinas, CA	2	$211	1	$332	84	-$120
MSA in the Midwest	3	$205	6	$129	4	$76
Oakland-Fremont-Hayward, CA	4	$189	4	$215	55	-$26
Minneapolis-St. Paul-Bloomington, MN-WI	5	$149	8	$108	15	$41
MSA in the Midwest	6	$143	15	$84	7	$59
Fort Worth-Arlington, TX	7	$126	10	$101	24	$25
Indianapolis, IN	8	$122	12	$95	22	$28
Gary, IN	9	$120	44	$5	1	$115
MSA in the West	10	$118	2	$224	82	-$106
Dallas-Plano-Irving, TX	11	$112	7	$114	38	-$2
Peoria, IL	12	$108	9	$106	34	$2
Houston-Sugar Land-Baytown, TX	13	$107	20	$64	14	$43
Denver-Aurora, CO	14	$92	34	$29	6	$63
Miami-Miami Beach-Kendall, FL	15	$92	22	$60	19	$32
Las Vegas-Paradise, NV	71	-$72	33	$29	80	-$102
Louisville, KY-IN	72	-$74	65	-$36	63	-$38
Nassau-Suffolk, NY	73	-$74	30	$33	83	-$108
MSA in the South	74	-$75	24	$50	85	-$125
MSA in the South	75	-$76	64	-$35	64	-$41
Memphis, TN-MS-AR	76	-$82	38	$22	81	-$104
Providence-New Bedford-Fall River, RI-MA	77	-$89	82	-$93	31	$4
Kingsport-Bristol-Bristol, TN-VA	78	-$90	81	-$93	33	$3
Warren-Farmington Hills-Troy, MI	79	-$98	78	-$78	51	-$20
MSA in the South	80	-$99	61	-$31	74	-$68
MSA in the Midwest	81	-$103	71	-$54	66	-$49
Augusta-Richmond County, GA-SC	82	-$105	70	-$54	68	-$52
Detroit-Livonia-Dearborn, MI	83	-$107	79	-$81	56	-$26
MSA in the South	84	-$121	77	-$70	67	-$51
Youngstown-Warren-Boardman, OH-PA	85	-$162	85	-$140	52	-$22
sd		$81		$79		$52
p10		-$89		-$70		-$84
p90		$120		$106		$55
N		85		85		85

Although the decomposition method is quite different, one can see that the qualitative results remain very similar to those reported in the paper. The correlations for these three indexes are 1.00 for the episode based index, 0.99 for the service price components, and 0.92 for the service utilization component.

3 Service-Specific Indexes

This section reports in greater detail the service-specific indexes reported in Table 7 in the paper. To construct $SPI_{r,s}$ the price of each service type s for treating disease d, $p_{d,r,s}$, is weighted by the expenditure share of that service type across diseases. For example, let the inpatient hospital expenditure share for disease d be denoted $\theta_{d,Inpatient}$ where $\sum \theta_{d,Inpatient} = 1$. Then the price index for the service category would be: $SPI_{r,s} = \sum_d p_{d,r,s} \cdot \theta_{d,Inpatient}$. Similar calculations are made for the $MCE_{r,s}$ and $SUI_{r,s}$. In contrast to the overall index that is weighted by the total expenditure share for each disease, this index is weighted by the expenditure share of a particular service.

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

	MSA	SPI_r	Inpatient Hospital	Outpatient Hospital	Office - General MD	Office MD - Speciality	Other	Pharmacy
1	Salinas, CA	1.391	2.213	1.520	0.943	0.979	1.239	1.030
2	MSA in the West	1.262	1.480	1.396	1.137	1.040	1.311	1.047
3	Milwaukee-Waukesha-West Allis, WI	1.254	1.271	1.334	1.380	1.342	1.132	1.035
4	Oakland-Fremont-Hayward, CA	1.251	1.501	1.269	1.121	1.067	1.430	1.036
5	San Diego-Carlsbad-San Marcos, CA	1.158	1.296	1.234	0.968	0.974	1.262	1.040
6	MSA in the Midwest	1.149	1.124	1.198	1.151	1.152	1.113	1.069
7	Dallas-Plano-Irving, TX	1.129	1.065	1.287	1.072	1.070	1.120	1.017
8	Minneapolis-St. Paul-Bloomington, MN-WI	1.122	1.120	1.073	1.296	1.257	1.061	0.981
9	Peoria, IL	1.120	0.938	1.268	1.152	1.281	1.232	0.854
10	Fort Worth-Arlington, TX	1.114	1.102	1.246	1.049	1.062	1.069	1.003
76	Cleveland-Elyria-Mentor, OH	0.912	0.784	0.830	0.928	0.918	0.866	1.073
77	MSA in the South	0.906	0.919	0.853	0.870	0.807	0.709	1.097
78	Warren-Farmington Hills-Troy, MI	0.896	0.739	0.710	0.983	0.998	0.877	1.096
79	Detroit-Livonia-Dearborn, MI	0.893	0.797	0.706	0.982	0.988	0.860	1.050
80	MSA in the Midwest	0.882	0.751	0.765	0.880	0.880	0.784	1.133
81	Kingsport-Bristol-Bristol, TN-VA	0.878	0.799	0.790	1.008	1.007	0.840	0.872
82	Providence-New Bedford-Fall River, RI-MA	0.878	0.906	0.634	0.988	0.972	0.902	0.954
83	Akron, OH	0.877	0.775	0.781	0.884	0.867	0.798	1.075
84	Pittsburgh, PA	0.854	0.764	0.666	0.920	0.910	0.811	1.050
85	Youngstown-Warren-Boardman, OH-PA	0.821	0.660	0.732	0.826	0.813	0.663	1.101
	COV	0.095	0.196	0.205	0.118	0.116	0.157	0.068

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

	MSA	MCE_r	Inpatient Hospital	Outpatient Hospital	Office - General MD	Office MD - Speciality	Other	Pharmacy
1	Milwaukee-Waukesha-West Allis, WI	1.288	1.451	1.679	1.338	1.153	0.965	0.980
2	Salinas, CA	1.245	2.118	1.276	1.021	0.899	1.041	0.908
3	MSA in the Midwest	1.240	1.243	1.455	1.056	1.330	1.063	1.113
4	Oakland-Fremont-Hayward, CA	1.219	1.344	1.265	1.166	1.133	1.532	0.975
5	Minneapolis-St. Paul-Bloomington, MN-WI	1.170	1.407	1.054	1.491	1.140	1.108	1.013
6	MSA in the Midwest	1.165	1.142	1.365	0.994	1.267	1.122	0.987
7	Fort Worth-Arlington, TX	1.143	1.229	1.126	1.110	1.167	1.169	1.069
8	Indianapolis, IN	1.138	1.051	1.627	0.881	0.815	1.011	1.094
9	Gary, IN	1.136	0.997	1.636	0.749	1.078	0.951	1.010
10	MSA in the West	1.133	1.340	1.196	1.129	1.022	1.060	1.010
76	Memphis, TN-MS-AR	0.891	0.740	0.690	0.951	1.082	1.286	0.856
77	Providence-New Bedford-Fall River, RI-MA	0.884	0.881	0.907	0.906	0.801	0.838	0.936
78	Kingsport-Bristol-Bristol, TN-VA	0.881	0.731	0.742	0.937	1.069	0.867	1.009
79	Warren-Farmington Hills-Troy, MI	0.872	0.729	0.901	0.997	0.875	0.702	1.017
80	MSA in the South	0.871	0.768	0.646	0.755	1.114	0.856	1.088
81	MSA in the Midwest	0.865	0.829	0.776	1.164	0.824	0.724	0.990
82	Augusta-Richmond County, GA-SC	0.862	0.858	0.914	0.654	0.945	0.956	0.786
83	Detroit-Livonia-Dearborn, MI	0.861	0.776	0.940	0.943	0.785	0.747	0.936
84	MSA in the South	0.844	0.921	0.679	0.705	0.913	0.698	1.047
85	Youngstown-Warren-Boardman, OH-PA	0.793	0.665	0.852	0.696	0.695	0.836	0.925
	COV	0.098	0.203	0.247	0.246	0.185	0.211	0.077

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

	MSA	SUI_r	Inpatient Hospital	Outpatient Hospital	Office - General MD	Office MD - Speciality	Other	Pharmacy
1	Cambridge-Newton-Framingham, MA	1.182	0.878	2.307	0.615	0.762	1.019	0.872
2	Boston-Quincy, MA	1.173	0.924	2.247	0.597	0.723	1.053	0.875
3	MSA in the Midwest	1.132	1.184	1.514	0.906	0.774	1.173	1.022
4	Gary, IN	1.124	1.104	1.527	0.820	1.086	0.905	0.998
5	MSA in the South	1.122	1.224	1.475	1.044	0.706	0.940	1.112
6	MSA in the Northeast	1.102	0.978	1.327	0.879	0.944	1.285	1.086
7	Pittsburgh, PA	1.092	1.082	1.545	0.722	0.973	0.998	0.923
8	Kalamazoo-Portage, MI	1.078	1.115	1.519	1.237	0.680	0.674	1.042
9	St. Louis, MO-IL	1.071	1.160	1.137	0.833	0.940	1.223	1.051
10	MSA in the Midwest	1.069	1.095	1.158	0.920	1.145	0.956	1.041
76	Augusta-Richmond County, GA-SC	0.924	0.951	0.950	0.687	0.949	0.898	0.983
77	Las Vegas-Paradise, NV	0.924	0.990	0.442	1.299	1.174	1.123	0.954
78	Atlanta-Sandy Springs-Marietta, GA	0.912	0.857	0.697	0.836	1.138	1.027	1.016
79	New York-White Plains-Wayne, NY-NJ	0.902	0.946	0.598	0.830	1.421	1.111	0.752
80	MSA in the West	0.902	0.887	0.816	0.990	0.979	0.856	0.957
81	Des Moines, IA	0.898	0.933	0.558	0.908	1.054	1.115	1.013
82	Memphis, TN-MS-AR	0.895	0.856	0.578	0.936	1.068	1.326	0.903
83	Nassau-Suffolk, NY	0.890	0.911	0.494	0.843	1.485	1.027	0.834
84	Salinas, CA	0.888	0.951	0.805	1.083	0.917	0.817	0.883
85	MSA in the South	0.870	0.932	0.646	0.906	0.960	0.758	1.057
	COV	0.064	0.090	0.335	0.216	0.188	0.162	0.084

4 Population-based Estimates (continued)

In the paper, we focus on the amount of variation, conditional on a diagnosis, because we view this as an accurate, but also more conservative measure of variation than the population-based measure. In particular, we are concerned about potentially overstating variation across regions if we ignore diagnosis. In this section, we analyze issues related to population-based and episode-based estimates in greater detail. We begin by repeating Table 8, but for the full 15 diseases that are included in Table 4. One can see that for most each of the diseases listed in Table A9.1, the population-based estimates show larger variation in expenditures, as measured by the $DECI_{d,r}$, than the expenditure variation that controls for the treated prevalence (i.e., the $MCE_{d,r}$). The average variation for the "typical" disease is shown on the bottom of the Table.

Table 9.1 Sources of Price Variation Across MSAs by Disease - $MCE_{d,r}$, $SPI_{d,r}$, $SUI_{d,r}$, $PREV_{d,r}$, and $DECI_{d,r}$

	Description	Severity	COV of $MCE_{d,r}$	s.e.	COV of $SPI_{d,r}$	s.e.	COV of $SUI_{d,r}$	s.e.	COV of $PREV_{d,r}$	s.e.	COV of $DECI_{d,r}$	s.e.
1	Pregnancy, with delivery	1	0.18	(0.009)	0.17	(0.004)	0.04	(0.005)	0.17	(0.001)	0.17	(0.009)
	Pregnancy, with delivery	2	0.20	(0.015)	0.19	(0.008)	0.05	(0.006)	0.16	(0.001)	0.18	(0.011)
2	Joint degeneration, localized - back	1	0.18	(0.007)	0.13	(0.004)	0.15	(0.006)	0.18	(0.001)	0.18	(0.006)
	Joint degeneration, localized - back	2	0.28	(0.024)	0.18	(0.017)	0.18	(0.012)	0.22	(0.002)	0.26	(0.018)
	Joint degeneration, localized - back	3	0.29	(0.022)	0.20	(0.014)	0.17	(0.015)	0.31	(0.002)	0.39	(0.021)
3	Ischemic heart disease	1	0.22	(0.010)	0.17	(0.009)	0.17	(0.007)	0.30	(0.001)	0.29	(0.009)
	Ischemic heart disease	2	0.22	(0.016)	0.22	(0.021)	0.15	(0.013)	0.31	(0.002)	0.35	(0.016)
	Ischemic heart disease	3	0.33	(0.055)	0.24	(0.030)	0.17	(0.018)	0.31	(0.003)	0.33	(0.033)
4	Hypertension	1	0.10	(0.002)	0.09	(0.002)	0.11	(0.002)	0.16	(0.001)	0.24	(0.002)
	Hypertension	2	0.11	(0.007)	0.10	(0.008)	0.12	(0.004)	0.22	(0.001)	0.28	(0.007)
	Hypertension	3	0.14	(0.029)	0.13	(0.035)	0.12	(0.012)	0.30	(0.001)	0.34	(0.011)
	Hypertension	4	0.35	(0.099)	0.33	(0.056)	0.24	(0.038)	0.36	(0.002)	0.45	(0.039)
5	Diabetes	1	0.11	(0.006)	0.06	(0.006)	0.10	(0.003)	0.18	(0.001)	0.19	(0.004)
	Diabetes	2	0.18	(0.026)	0.13	(0.043)	0.15	(0.018)	0.24	(0.002)	0.26	(0.014)
	Diabetes	3	0.20	(0.015)	0.11	(0.015)	0.16	(0.009)	0.19	(0.002)	0.25	(0.014)
	Diabetes	4	0.21	(0.023)	0.17	(0.021)	0.14	(0.015)	0.19	(0.002)	0.22	(0.022)
6	Routine exam	1	0.15	(0.001)	0.12	(0.001)	0.06	(0.001)	0.24	(0.000)	0.26	(0.001)
7	Mood disorder, depressed	1	0.20	(0.004)	0.06	(0.004)	0.17	(0.003)	0.22	(0.001)	0.29	(0.004)
	Mood disorder, depressed	2	0.23	(0.015)	0.11	(0.022)	0.18	(0.009)	0.24	(0.002)	0.30	(0.012)
	Mood disorder, depressed	3	0.24	(0.024)	0.20	(0.042)	0.18	(0.012)	0.26	(0.004)	0.33	(0.020)
8	Malignant neoplasm of breast	1	0.20	(0.027)	0.14	(0.009)	0.16	(0.022)	0.14	(0.002)	0.20	(0.028)
	Malignant neoplasm of breast	2	0.27	(0.026)	0.18	(0.013)	0.20	(0.014)	0.20	(0.003)	0.26	(0.028)
9	Hyperlipidemia, other	1	0.10	(0.002)	0.08	(0.003)	0.10	(0.002)	0.15	(0.001)	0.17	(0.003)
10	Joint degeneration, localized - neck	1	0.19	(0.007)	0.11	(0.005)	0.18	(0.007)	0.26	(0.001)	0.23	(0.005)
	Joint degeneration, localized - neck	2	0.31	(0.031)	0.19	(0.035)	0.27	(0.025)	0.30	(0.004)	0.37	(0.024)
	Joint degeneration, localized - neck	3	0.30	(0.029)	0.23	(0.016)	0.22	(0.015)	0.31	(0.003)	0.35	(0.020)
11	Chronic sinusitis	1	0.17	(0.003)	0.08	(0.002)	0.13	(0.003)	0.30	(0.001)	0.26	(0.004)
	Chronic sinusitis	2	0.20	(0.027)	0.09	(0.005)	0.16	(0.017)	0.29	(0.002)	0.23	(0.014)
	Chronic sinusitis	3	0.25	(0.015)	0.17	(0.010)	0.18	(0.009)	0.39	(0.002)	0.37	(0.014)
12	Joint degeneration, localized - knee & lower leg	1	0.25	(0.020)	0.16	(0.008)	0.17	(0.009)	0.22	(0.002)	0.28	(0.011)
	Joint degeneration, localized - knee & lower leg	2	0.31	(0.028)	0.18	(0.017)	0.23	(0.021)	0.30	(0.004)	0.38	(0.022)
	Joint degeneration, localized - knee & lower leg	3	0.29	(0.023)	0.24	(0.057)	0.28	(0.019)	0.31	(0.005)	0.47	(0.025)
13	Asthma	1	0.11	(0.006)	0.06	(0.004)	0.10	(0.007)	0.17	(0.001)	0.21	(0.005)
	Asthma	2	0.11	(0.006)	0.07	(0.003)	0.10	(0.006)	0.16	(0.001)	0.16	(0.006)
	Asthma	3	0.31	(0.189)	0.17	(0.035)	0.16	(0.037)	0.24	(0.002)	0.25	(0.075)
	Asthma	4	0.24	(0.054)	0.28	(0.154)	0.25	(0.036)	0.24	(0.003)	0.37	(0.027)
14	Joint derangement - knee & lower leg	1	0.26	(0.017)	0.17	(0.009)	0.21	(0.020)	0.25	(0.003)	0.26	(0.018)
	Joint derangement - knee & lower leg	2	0.21	(0.008)	0.21	(0.005)	0.18	(0.011)	0.14	(0.002)	0.20	(0.008)
15	Inflammation of esophagus	1	0.12	(0.004)	0.10	(0.008)	0.09	(0.003)	0.17	(0.001)	0.20	(0.005)
	Inflammation of esophagus	2	0.14	(0.015)	0.12	(0.017)	0.11	(0.013)	0.28	(0.002)	0.29	(0.015)
	Inflammation of esophagus	3	0.33	(0.077)	0.20	(0.042)	0.22	(0.046)	0.34	(0.004)	0.36	(0.040)
	Weighted Average (Full Sample - 10,000 Episodes in the Data)		0.223		0.161		0.167		0.246		0.304	
	Weighted Average (Only Diseases with 50,000 Episodes in the Data)		0.178		0.131		0.140		0.234		0.270	

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

To place our measures of variation in context, we compare our variation statistics to variation statistics calculated from Dartmouth Atlas data that we obtained from the website (http://www.dartmouthatlas.org/). One notable difference between our analysis, and that of the Dartmouth researchers, is that our geographic units are distinct. Dartmouth researchers study geographic units called Hospital Referral Regions (HRRs) that are based on the flows of Medicare patients to providers. We were unable to attain HRR for enrollees in our data, but we believe that the MSA is roughly comparable to the HRR unit of analysis. Another difference is that we focus on only 85 MSAs with a large number of enrollees. To make the Dartmouth Atlas estimates comparable to ours, we not only report the estimates from the full sample of HRRs, but we also report estimates from the 85 HRRs with the greatest number of Medicare enrollees. Focusing on the top 85 areas, we find that the average COV is 0.26 from the selected Dartmouth studies, which compares to an average of 0.30 for our $DECI_{d,r}$ variation and 0.27 if we look at only those diseases for which we observe 50,000 or more episodes in our data.[6] Therefore, it appears that the

[6] Note that there are many additional differences in these statistics. Dartmouth studies are often focusing on just a single procedure, while we look at a utilization measure. In contrast, our measure includes multiple types of services, such as inpatient and outpatient services.

level of variation is quite similar. Also note that the variation in the MCE, which controls for treated prevalence, is lower than the population-based estimates.

An important point to note is that variation increases if one expands the geographic areas that one looks at. Examining the Darmouth studies, we find that the average COV increases by about 30 percent if one expands the number of cities, from the 85 most populous to include all HRRs.

Table A9.2 Coefficient of Variation for Surgical Procedures - Dartmouth Group

	Sample of HRRs	
	All	Top 85
Electrocardiographic Monitoring per 1,000 Medicare Enrollees (1996)	0.597	0.449
Non-Stress Nuclear Studies per 1,000 Medicare Enrollees (1996)	0.572	0.467
Spinal Fusion per 1,000 Medicare Enrollees (1996-97)	0.459	0.366
Sholder arthroplasty (2005-06)	0.433	0.289
Lumbar Discectomy per 1,000 Medicare Enrollees (1996-97)	0.427	0.371
Imaging Stress Testing per 1,000 Medicare Enrollees (1996)	0.399	0.326
Non-Imaging Stress Testing per 1,000 Medicare Enrollees (1996)	0.397	0.324
Percutaneous Coronary Intervention per 1,000 Medicare Enrollees (1996)	0.371	0.253
Unallocated Cardiac Catheterization Laboratories per 100,000 Residents (1996)	0.358	0.306
Total carotid artery revascularization (2007)	0.356	0.179
Decompression for Lumbar Stenosis per 1,000 Medicare Enrollees (1996-97)	0.353	0.280
Initial Vascular Access Surgery per 1,000 Medicare Enrollees (1996-97)	0.343	0.287
Lung Cancer Surgery per 1,000 Medicare Enrollees (1996-97)	0.341	0.194
Hip Arthroplasty (2005-06)	0.325	0.212
Allocated Cardiac Catheterization Laboratories per 100,000 Residents (1996)	0.319	0.271
Echocardiography per 1,000 Medicare Enrollees (1996)	0.318	0.305
Spine Surgery per 1,000 Medicare Enrollees (1996-97)	0.307	0.260
Total Stress Testing per 1,000 Medicare Enrollees (1996)	0.293	0.223
Elective AAA Repair per 1,000 Medicare Enrollees (1996-97)	0.274	0.169
Coronary Angiography per 1,000 Medicare Enrollees (1996)	0.265	0.189
Knee Arthroplasty (2005-06)	0.265	0.218
Pacemaker Insertion per 1,000 Medicare Enrollees (1996)	0.254	0.200
CT/MRI per 1,000 Medicare Enrollees (1996-97)	0.244	0.188
Aortic Valve Replacement per 1,000 Medicare Enrollees (1996-97)	0.223	0.193
Coronary Artery Bypass Grafting per 1,000 Medicare Enrollees (1996)	0.210	0.172
Total AAA Repair per 1,000 Medicare Enrollees (1996-97)	0.190	0.156
Admissions for Acute Myocardial Infarction per 1,000 Medicare Enrollees (1996)	0.173	0.152
Average Across Studies	0.336	0.259

Notes. This table was constructed by drawing on data from various reports from the Dartmouth Atlas website where downloadable data is available (http://www.dartmouthatlas.org/publications/reports.aspx).

Similar effects are observed in our data. Rather than focusing on just the top 85 cities with more than 20,000 enrollees per year, we focus on those MSAs where we observe an average of 2,000 enrollees per year. The estimates are reported in Table A9.3. Similar to the analysis with the Dartmouth Atlas data, we observe a notable increase in variation, as measured by the COV, when we expand the number of geographic locations.

Table A9.3 Sources of Price Variation Across MSAs by Disease - $MCE_{d,r}$, $SPI_{d,r}$, $SUI_{d,r}$, $PREV_{d,r}$, and $DECI_{d,r}$ -- 2000 or More Enrollees in MSA

	Description	Severity	COV of $MCE_{d,r}$	COV of $SPI_{d,r}$	COV of $SUI_{d,r}$	COV of $PREV_{d,r}$	COV of $DECI_{d,r}$
1	Ischemic heart disease	1	0.30	0.22	0.22	0.33	0.38
	Ischemic heart disease	2	0.42	0.28	0.27	0.37	0.44
	Ischemic heart disease	3	0.53	0.38	0.32	0.36	0.44
2	Joint degeneration, localized - back	1	0.23	0.16	0.19	0.22	0.26
	Joint degeneration, localized - back	2	0.42	0.25	0.29	0.29	0.36
	Joint degeneration, localized - back	3	0.43	0.42	0.33	0.41	0.52
3	Pregnancy, with delivery	1	0.18	0.17	0.07	0.22	0.22
	Pregnancy, with delivery	2	0.26	0.22	0.11	0.28	0.28
4	Hypertension	1	0.13	0.12	0.13	0.20	0.29
	Hypertension	2	0.18	0.16	0.15	0.29	0.37
	Hypertension	3	0.37	0.23	0.20	0.37	0.45
	Hypertension	4	1.32	0.50	0.40	0.48	0.64
5	Diabetes	1	0.16	0.08	0.13	0.23	0.24
	Diabetes	2	0.32	0.15	0.27	0.36	0.40
	Diabetes	3	0.38	0.17	0.26	0.29	0.34
	Diabetes	4	0.38	0.27	0.28	0.31	0.37
6	Routine exam	1	0.16	0.16	0.07	0.26	0.27
7	Mood disorder, depressed	1	0.21	0.09	0.18	0.25	0.30
	Mood disorder, depressed	2	0.32	0.26	0.23	0.29	0.37
	Mood disorder, depressed	3	0.49	0.38	0.38	0.34	0.53
8	Hyperlipidemia, other	1	0.13	0.11	0.11	0.20	0.22
9	Joint degeneration, localized - neck	1	0.26	0.15	0.24	0.33	0.35
	Joint degeneration, localized - neck	2	0.57	0.26	0.49	0.42	0.56
	Joint degeneration, localized - neck	3	0.52	0.37	0.40	0.44	0.57
10	Malignant neoplasm of breast	1	0.39	0.20	0.31	0.20	0.36
	Malignant neoplasm of breast	2	0.55	0.28	0.48	0.33	0.51
11	Joint degeneration, localized - knee & lower leg	1	0.35	0.20	0.27	0.24	0.33
	Joint degeneration, localized - knee & lower leg	2	0.54	0.31	0.47	0.37	0.57
	Joint degeneration, localized - knee & lower leg	3	0.84	0.40	0.56	0.44	0.65
12	Chronic sinusitis	1	0.22	0.11	0.17	0.40	0.37
	Chronic sinusitis	2	0.36	0.19	0.26	0.36	0.37
	Chronic sinusitis	3	0.56	0.26	0.28	0.57	0.59
13	Joint derangement - knee & lower leg	1	0.45	0.30	0.48	0.35	0.51
	Joint derangement - knee & lower leg	2	0.25	0.25	0.21	0.22	0.28
14	Asthma	1	0.20	0.10	0.15	0.25	0.29
	Asthma	2	0.20	0.10	0.16	0.24	0.26
	Asthma	3	2.81	4.51	0.42	0.42	0.47
	Asthma	4	0.61	0.40	0.43	0.41	0.58
15	Inflammation of esophagus	1	0.19	0.13	0.15	0.23	0.28
	Inflammation of esophagus	2	0.37	0.18	0.23	0.38	0.40
	Inflammation of esophagus	3	0.56	0.25	0.64	0.48	0.60
	Weighted Average (Full Sample - 10,000 Episodes in the Data)		0.415	0.258	0.291	0.320	0.485
	Weighted Average (Only Diseases with 50,000 Episodes in the Data)		0.287	0.194	0.211	0.284	0.376

It is important to emphaisize that there are important differences in our statistics and those reported in the Dartmouth studies. Specifically, our analysis focuses on overall utilization differences for treatment, rather than the level of variation at the procedure level. So, for example, if a simple mastectomy is replaced by a radical mastectomy, then looking at a utilization measure for those two measures together, would show less utilization differences across markets, than looking at a only differences in the rate of radical mastectomy alone. Depression is another example. If one substitutes talk therapy for drug therapy, then the amount of variation measured by our utilization measure will be lower than what would be found by looking at a single measure (e.g., a talk therapy visit).

4.1 Prevalence and MSA Factors

The key difference between the episode-based measure and the population-based measure is the treated prevalence for each diseases. Therefore, it may be useful to examine the source of the variation for treated prevalence. Table A5.1 presents analysis that looks at how much variation in treated prevalence for each disease may be explained by MSA fixed effects, MSA-MPC fixed effects, and MSA-ETG Category fixed effects (identical to the analysis in Table 5). We find that, similar to the utilization measure, MSA fixed effects explain relatively little of the variation in treated prevalence across areas, with an R^2 of 0.19. Including MSA-Major Disease Category fixed effects more than doubles the R^2 measure (and adjusted R-squared measure), indicating that factors that affect the prevalence of a disease in an area tend to be specific to the type of disease.

Table A5.3 Decomposing the Sources of Treated Prevalence Variation

Log(Treated Prevalence)

	R^2	Adj R^2	MSE	N
MSA-FE	0.191	0.175	0.247	11305
MSA-MPC-FE	0.478	0.400	0.210	11305
MSA-Disease-FE	0.822	0.692	0.151	11305

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

4.2 Regional Diagnostic Practices

This paper focuses on analyzing the variation in spending by disease-episode, which controls for health status by categorizing patients into disease bins. In this respect, our work is similar to other work that analyzes variation in expenditures by controlling for disease, such as, Zhang et al. (2010) and MedPac (2003, 2009). A disadvantage of this approach is that the diagnosis decisions may vary geographically, which is suggested by recent research of Song et al. (2010). The Song et al. study finds that an individual moving to a high-spending area from a low-spending area receives a greater number of diagnosis than she would have received if she had not moved. In this case, it is possible that a patient given a diagnosis code indicating a minor illness in one MSA may be given a diagnosis code indicating a more severe illness in another MSA. This could cause an episode-based measure to understate the cost of treatment in high-spending areas. This possibility is consistent with the findings of Welch et al. (2011) which shows substantial variation in geographic frequency of diagnosis, but little variation in population-based mortality rates across high and low diagnosis areas. In more general terms, if less severe cases are more likely to be treated in some regions than others, then areas with higher treated prevalence may have lower than average spending for the same disease category.

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

$$\log(SUI_{d,r}) = \alpha \log(PREV_{d,r}) + \varepsilon_{d,r} \quad (1)$$

If $\alpha = 0$ this would indicate that treated prevalence is not systematically affecting measured utilization levels across markets. Table A9.4 presents estimates of the regression. The relationship is negative and statistically significant, indicating some possible systematic relationship; however, the magnitude is economically small, with a coefficient of -0.09. That is, a treated prevalence rate that is 10 percent higer is associated with a 0.9 percent smaller utilization level. We also include additional specifications that control for potential measurement issues. We apply instruments, control for the service price, and apply MSA fixed effects. Including service price as a control variable, model (2), shows a negative relationship between service price and utilization, but has practically no effect on the prevalence coefficient. Including MSA fixed effects, model (3), has very little effect on either the price or treated prevalence estimates. As a final check of potential measurement issues, we include instruments for treated prevalence and service prices. As an instrument for price, we use service prices for other diseases, since the quality of treatment for different diseases is likely distinct, but there are likely common cost factors that are correlated with price (e.g. nurses' wages). Similarly, as an instrument for treated prevalence, we use the treated prevalence for other diseases in the MSA. The results that apply instrumental variables appear in model (4), which shows a negative but insignificant relationship between treated prevalence and utilization.

Table A9.4 Regressions of log($SUI_{d,r}$) on the Log of Treated Prevalence

	(1)	(2)	(3)	(4)
log($PREV_{d,r}$)	-0.0964***	-0.103***	-0.0824***	-0.0310
	(-8.95)	(-9.63)	(-9.77)	(-1.11)
log($SPI_{d,r}$)		-0.118***	-0.0666***	-0.329***
		(-5.70)	(-5.06)	(-5.77)
MSA Fixed Effects	No	No	Yes	Yes
Price Control	No	Yes	Yes	Yes
Instrumental Variable	No	No	No	Yes
R^2	0.0313	0.0410	0.0314	-0.0168
N	26520	26520	26520	26350

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

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

For each of the four specifications, the magnitude of the coefficient on prevalence is quite small and suggests that regional diagnosis differences may not be problematic. This can also be revealed through the very high correlation (0.99) between the residuals of specification (1) and the $SUI_{d,r}$. Based on these simple correlations, these estimates indicate that controlling for prevalence makes little difference to the underlying indexes. However, more researcher is needed to better understand the differences in diagnostic practices across markets for the commercial medical-care market.

5 Medical-Care Expenditures and Quality Measures

There is considerable variation in spending, service prices and service utilization, but it is unclear what this variation means for consumer surplus, since greater spending may be associated with high-quality care. Although this relationship between spending and quality does not appear to be present in Medicare markets,[7] less research has been conducted in the commercial sector. Here we use a set of procedural quality measures constructed from the MarketScan data to examine whether there is an association between quality and the MCE, SUI and SPI. The quality measures are constructed using Healthcare Effectiveness Data and Information Set (HEDIS) guidelines from the National Committee for Quality Assurance (NCQA). Additional details regarding the construction of the quality measures are included in the following subsection.

The top portion of Table A10 shows the correlation between the log of specific quality measures and the log of the $MCE_{d,r}$, $SPI_{d,r}$, and $SUI_{d,r}$ for the corresponding diseases. For instance, we compare the quality measure indicating beta-blocker treatment with indexes where d = hyptertension and d = ischemic heart disease. Looking at correlations between disease-specific quality measures and their corresponding disease indexes, there does not appear to be a consistent pattern. While the correlation for diabetes testing for pediatric patients and cholesterol testing for patients with ischemic vascular disease show some positive and significant relationships, we see that many of the other correlations are insignificant or significant and negative.

In the bottom portion of Table A10 we include additional quality measures that assess preventative measures. Since preventative measures are not disease-specific we compare this measure to the aggregate MCE_r, SPI_r and SUI_r measures. There is a strong positive relationship between the SUI_r and the preventative quality measure. For completeness,

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

we also compare the MSA-level indexes to a composite of the four non-preventative measures listed in the top portion of the table. For the composite of non-preventative quality measures, there appears to be a significant correlation with SPI_r. This latter correlation may be indicative of a range of demographic and physician factors that may be correlated with both prices and quality of treatment at the MSA level.

Table A10. Correlation between Quality and Episode Indexes: MCE, SPI and SUI

	Log(MCE)	Log(SPI)	Log(SUI)
Log Quality Measure - Persistence of Beta-Blocker Treatment after a Heart Attack			
Hypertension	-0.1989*	0.1222	-0.2254**
p-value	0.0681	0.2654	0.038
Ischemic Heart Disease	-0.1041	-0.1331	0.0492
p-value	0.3432	0.2245	0.6545
Log Quality Measure - HbA1c Test for Pediatric Patients			
Diabetes	0.2194**	0.1298	0.2045*
p-value	0.0436	0.2364	0.0605
Log Quality Measure - Complete Lipid Profile for Patients 18 years and older with Ischemic Vascular Disease			
Ischemic Heart Disease	0.1595	0.3402***	-0.2252**
p-value	0.1447	0.0014	0.0382
Log Quality Measure - Those with Back Pain Not Reporting an MRI within first 6 Months			
Joint Degeneration - Back	0.0902	0.1778	-0.0172
p-value	0.4119	0.1036	0.8755
Log Preventative Measures - Composite (Summation of Two Indexes of Preventative Care)			
Aggregate Indexes	0.1612	-0.1156	0.4496***
p-value	0.1404	0.292	0.000
Log Non-Prev. Quality Measures - Composite (Summation of the Four Quality Indexes Listed Above)			
Aggregate Indexes	0.3308***	0.3563***	0.0118
p-value	0.002	0.0008	0.9147
Log Quality Measure - Composite (Summation of the Four Quality Indexes and the Two Preventative Care Indexes)			
Aggregate Indexes	0.2334**	0.0288	0.3745***
p-value	0.0315	0.7937	0.0004

* 90% signficance level ** 95% signficance level *** 99% significance level

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

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

In this study we find mixed evidence regarding measures of quality in each of the three indexes. Although measures of quality and the associated spending on related diseases is unclear, we find a strong positive relationship with aggregate measures of spending and composite measures of quality. The positive correlation between the overall utilization measure and quality measures of preventative care is particularly strong, but the underlying cause of this relationship is unclear.

5.1 Quality Measures

We construct six quality measures from the claims data using methods outlined by National Committee for Quality Assurance (NCQA). We focus on quality measures that may be constructed from administrative claims data. These quality measures are described in greater detail in NCQA Measure Technical Specifications.[8]

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

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

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

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

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

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

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

6 Additional Robustness Checks

To check the robustness of the results presented in this paper, we generated the tables presented here under a number of alternative assumptions (Tables 2 through 7). The following is a list of additional robustness checks.

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

2. Changing the included diseases - For selecting diseases to be included in the sample we alter the threshold for the number of episodes observed in the data. Recall that the threshold applied here was 10,000. Similar results are attained if the threshold falls to 5,000 (accounting for 86 percent of spending) or is increased to 50,000 (accounting for 50 percent of spending). The problem with a lower threshold is that there will not be a sufficient number of episodes to attain an accurate estimate. In contrast, the problem with too high a threshold is that it accounts for a more limited fraction of overall spending.

3. Unweighted analysis - We conducted the same analysis but removed the population weights.

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

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

6. No severity adjustment - We aggregated diseases to the ETG category level and left out differences in severity.

7. Aggregate across diseases - For this analysis we aggregated diseases to the Major Practice Category (MPC) level.

References

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

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

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

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

[5] Dunn, Abe and Shapiro, Adam, (2012), "Physician Market Power and Medical-Care Expenditures", Bureau of Economic Analysis, *Working Paper.*

[6] MaCurdy, Thomas, Jason Kerwin, Jonathan Gibbs, Eugene Lin, Carolyn Cotterman, Margaret O'Brien-Strain, Nick Theobald, and Frederick Thomas, (2008), "Evaluating the Functionality of the Symmetry ETG and Medstat MEG Software in Forming Episodes of Care Using Medicare Data", Acumen, LLC.

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

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

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

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

[11] Rosen, Allison, Eli Liebman, Ana Aizcorbe, and David Cutler, (2012), "Comparing Commercial Systems for Characterizing Episodes of Care", Bureau of Ecnomic Analysis, *Working Paper.*

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

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

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

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