

© Health Research and Educational Trust  
DOI: 10.1111/1475-6773.12501  
RESEARCH BRIEF

## Geographic Variation in Quality of Care for Commercially Insured Patients

*Michael Richard McKellar, Mary Beth Landrum, Teresa B. Gibson, Bruce E. Landon, A. Mark Fendrick, and Michael E. Chernew*

---

**Background.** Extensive evidence documents geographic variation in spending, but limited research assesses geographic variation in quality, particularly among commercially insured enrollees.

**Objective.** To measure geographic variation in quality measures, correlation among measures, and correlation between measures and spending for commercially insured enrollees.

**Data Source.** Administrative claims from the 2007–2009 Truven MarketScan database.

**Methods.** We calculated variation in, and correlations among, 10 quality measures across 306 Hospital Referral Regions (HRRs), adjusting for beneficiary traits and sample size differences. Further, we created a quality index and correlated it with spending.

**Results.** The coefficient of variation of HRR-level performance ranged from 0.04 to 0.38. Correlations among quality measures generally ranged from 0.2 to 0.5. Quality was modestly positively related to spending.

**Conclusion.** Quality varied across HRRs and there was only a modest geographic “quality footprint.”

**Key Words.** Geographic variation, quality, spending, markets

---

Decades of research documents variation in health care spending across geographic areas even after controlling for patient characteristics (Fisher et al. 2003a,b; Zuckerman et al. 2010; Medicare Payment Advisory Commission 2011; Institute of Medicine [IOM] 2013). This research has contributed substantially to the policy debate. For example, based in part on estimates from geographic variation, Peter Orszag, the former director of the Congressional Budget Office, testified that “nearly 30 percent of Medicare’s costs could be saved without negatively affecting health outcomes,” and this work supported the passage of the Affordable Care Act (Chandra 2009; Pear 2009; Rosenthal 2012).

Just as variation in spending may highlight opportunities to save money, variation in quality may highlight opportunities to improve care. The quality variation literature generally focuses on Medicare or the total population (Fisher et al. 2003a; Baicker and Chandra 2004a; Baicker, Buckles, and Chandra 2006; Jones et al. 2012; Radley et al. 2012; Tsai et al. 2013). Considerable (though not all) research suggests that variation in Medicare may not generalize to the commercial population (Chernew et al. 2010; Franzini, Mikhail, and Skinner 2010; Franzini et al. 2011, 2015). Therefore, it may also be the case that quality patterns in Medicare may not be generalizable to the commercial population.

Understanding variation in quality provided to the commercial population is important for a number of reasons. First, given the reliance of the Affordable Care Act on private insurers to cover large segments of the population and recognition that quality is not uniformly high, we believe it is important to examine variation in the quality of care provided to the millions of commercially insured individuals in order to provide insight regarding the potential magnitude of quality deficits. Second, health plans are commonly held accountable for their performance on quality metrics. Yet these metrics likely reflect the performance of the providers in their network and therefore may reflect the markets they serve. Measures of health plan quality performance are typically not adjusted for the geographic distribution of their enrollees. The extent of geographic variation in quality may speak to the importance of this lack of geographic adjustment on a plan's measured performance. Third, the relationship between quality and spending reflects potential inefficiency in the health care system. If higher commercial spending (due to either inefficient utilization or high prices) is not correlated with quality, it suggests inefficiency in the commercial segment. We know from other work that such inefficiency exists in the Medicare market (Fisher et al. 2003b; Baicker and Chandra 2004b; Congressional Budget Office 2008; Landrum et al. 2008; Hassett, Neville, and Weeks 2014), but given the reliance of health reform on commercial markets, comparable analysis in the commercial market is important.

---

Address correspondence to Michael Richard McKellar, M.H.S.A., Department of Health Care Policy, Harvard Medical School, 2514 Mathews Ave Unit 5, Redondo Beach, CA 90278; e-mail: rmckell@umich.edu. Mary Beth Landrum, Ph.D., Teresa B. Gibson, Ph.D., Bruce E. Landon, M.D., M.B.A., M.S., and Michael E. Chernew, Ph.D., are with the Department of Health Care Policy, Harvard Medical School, Boston, MA. A. Mark Fendrick, M.D., is with the North Campus Research Complex, Ann Arbor, MI.

In this research brief, we study geographic variation in 10 commonly used process and outcome quality measures and examine the tendency for areas with high quality scores in one measure to have high quality scores in another. In order to highlight potential inefficiencies in the commercial sector, we also explore the correlation between total health care spending and quality performance in 10 measures across these geographic areas.

## METHODS

### *Data/Sample*

We analyzed administrative claims data from over 41 million employees and their dependents between age 18 and 65 enrolled at some point in the period between 2007 and 2009 using the Truven Health MarketScan Commercial Claims and Encounters Database. The data are comprised of commercial hospital, physician, and drug claims with their associated spending amounts, as well as procedure and diagnosis codes. These data encompass over 150 large private employers and health plans and include a range of utilization and demographic information.

### *Variables*

We constructed 10 quality metrics that can be measured using data from administrative claims. These encompass outcome measures as well as process measures. The measures were selected because they are frequently used in quality measurement and are often tied to pay-for-performance programs (Virnig et al. 2002; Zhang, Baicker, and Newhouse 2010). The outcome measures include 30-day readmission following an inpatient admission as well as two prevention quality indicators (PQIs) (e.g., ambulatory care sensitive hospitalizations for acute and chronic conditions) endorsed by the Agency for Health care Research and Quality (Davies et al. 2001; National Quality Measures Clearinghouse 2013). For PQI conditions, which are relatively rare within the commercial and under-65 population, composite measures were created and designed to include enrollees admitted for any of the indicated acute and chronic conditions.

We also constructed six Healthcare Effectiveness Data and Information Set (HEDIS) process measures including (1) mammography screening within the last 2 years among women ages 42–65; (2) disease-modifying antirheumatic drugs (DMARDs) treatment for rheumatoid arthritis; (3) use of

bronchodilator within 30 days of chronic obstructive pulmonary disease (COPD) diagnosis; (4) major depression prescription treatment management and adherence; (5) annual hemoglobin A1c testing among patients with diabetes; and (6) avoidance of imaging for patients with lower back pain (National Committee for Quality Assurance [NCQA] 2012). Lastly, an indicator was constructed for the appropriate use of antibiotic prescriptions for bacterial pneumonia based on the Physician Consortium for Performance Improvement (PCPI) guidelines (American Medical Association [AMA] 2012). Continuous enrollment throughout measure-specific periods was required to ensure completeness of recorded measures. For instance, to be included in the mammography cohort, an enrollee was required to have at least 2 years of continuous enrollment and was only eligible once within the 3-year window. For measures based on follow-up to a specific event, such as a readmission, diagnosis of back pain, and diagnosis of COPD, continuous enrollment was only required during the relevant follow-up period.

Spending included actual paid amounts for services and was measured by aggregating all hospital, physician, and drug claims. We obtained Hospital Referral Regions (HRR) level risk-adjusted spending, including adjustments for input price, age, sex, and health status, following methods described in the Institute of Medicine (IOM) report on geographic variation (IOM 2013). Our final spending measure is an HRR-level average, calculated per beneficiary per year, with adjustments for enrollees with partial-year enrollment.

We adjust for variation in input prices using the Hospital Wage Index for inpatient facility claims and the Geographic Practice Cost Indices for outpatient and physician claims (Melnick and Keeler 2007). Health status was evaluated based on the enrollees' prior-year diagnostic-cost-group (DxCg) risk score. The DxCg risk score, commonly used by private payers as a risk-adjustment tool, calculates enrollee health status using demographic characteristics, claims, and enrollment information as well as diagnoses (DxCg RiskSmart Stand Alone Software 2013). This system is similar to the Hierarchical Condition Categories used by Medicare (Pope et al. 2004).

For enrollees with missing data from the prior year (29.7 percent of enrollee/years), DxCg imputation was performed using the average scores among enrollees with available information adjusted based on age and sex. Earlier work for the IOM found estimates of geographic variation in spending results insensitive to inclusion of case mix measures (Harvard University 2012), suggesting case mix would not have a large effect on quality measures related to spending (e.g., readmissions). We used the HRRs to define geographic regions.

*Statistical Methods*

Our approach follows that outlined by the IOM committee study on Geographic Variation in Health Care Spending and Promotion of High-Value Care (IOM 2013). Specifically, we use a two-stage approach to measure variation. In the first stage, we estimated average effects for each HRR. For PCPI and HEDIS measures, which are restricted to certain populations, we computed average performance without risk-adjustment by calculating the proportion of enrollees who met the criteria over the total number of enrollees who qualified for that measure (e.g., women between 40 and 65 years that obtained a mammogram/women between 40 and 65 years). For the PQI composites and 30-day readmissions, which are not restricted to specific populations, we obtained risk-adjusted averages using logistic regression models controlling for age (treated as 5-year categorical bands), sex, age/sex interaction, and enrollee's prior year DxCG score. The residuals from the logistic regression models were then averaged at each HRR for PQIs and 30-day readmissions to obtain HRR-level adjusted quality measures. This method is almost identical to using HRR-level fixed effects, but it is much less computationally burdensome.

We used a second stage of analysis because the observed HRR-level effects reflect both true variation and sampling variability, and therefore variation in the mean HRR-level performance would yield an overestimate of the true variation. The noise increases for smaller HRRs and measures applied to smaller populations. To address this problem, we fit hierarchical or random effects models that included HRR-specific random effects to estimate the variance in performance across HRRs. Because all of the quality indicators are binary, we fit logistic regression models and then converted variance estimates from the log-odds to probability scale (Goldstein, Browne, and Rasbash 2002).

Similarly, we estimated pairwise correlations of the HRR-level quality indicators to one another using a multivariate hierarchical model that included correlated random intercepts for each measure. To adjust for multiple comparisons, a Bonferroni correction was made to obtain an overall type I error of 0.05. Individual tests were determined significant at the  $\alpha = .001$  level.

We also performed a factor analysis to summarize the 10 quality measures and identify potential latent factors using the estimated correlation matrix with reliability-adjusted correlations. Reliabilities were calculated using the variance estimates of the random effects models and the average

sample sizes across HRRs for each measure. Factors were generated using the principal factor method, and the number of factors kept was determined based on the eigenvalues  $\geq 1$  criteria. Standardized factor scores were generated for each HRR and were correlated with risk-adjusted total health care spending. Finally, we ran a simple linear model regressing factor scores on total spending.

## RESULTS

Over 35.5 million enrollees met the criteria for at least one of the 10 measures. The mean performance differed substantially from measure to measure. Across the 306 HRRs, the mean performance on two process measures (DMARD treatment for rheumatoid arthritis and yearly hemoglobin A1c test for patient with diabetes) was over 85 percent. However, four of the seven process measures (HEDIS and PCPI) had mean performances below 75 percent (Table 1).

Similarly, the variation in performance across HRRs depended on the measure. For instance, acute and chronic PQIs varied considerably across HRRs, with coefficients of variation (CV) of 0.38 and 0.34, respectively. However, there was less variation in the HEDIS and PCPI measures, with CVs ranging from 0.04 (DMARD treatment for rheumatoid arthritis) to 0.09 (depression treatment and avoidance of lower back imaging for patients with lower back pain). The CV was 0.08 for 30-day readmission.

Table 2 describes the pairwise correlations between each quality measure. Of the 45 pairwise correlations, 25 had statistically significant positive correlations with correlations ranging between 0.21 (i.e., hemoglobin A1c testing for enrollees with diabetes and antibiotic treatment for bacterial pneumonia) and 0.85 (i.e., acute and chronic PQI composite measures). Unexpectedly, increased hemoglobin A1c testing was correlated with higher rates of avoidable PQI acute admissions ( $\rho = -.19, p < .001$ ).

One factor was retained from the factor analysis, and this factor accounted for 29.6 percent of the total variation across the 10 measures. The Cronbach's alpha (a measure of how well a set of variables measures a single, one-dimensional latent aspect such as quality) estimate of 0.72 suggested coherence across the quality measures. Factor loadings (Table 3) indicated that the chronic and acute PQI composite measures were strongly correlated with the factor (loadings of 0.74 and 0.81, respectively). Hemoglobin A1c testing and 30-day readmissions showed weak correlations with the factor

Table 1: Variation in Quality for the Commercially Insured, 2007–2009

	<i>Mean Compliance</i>	<i>Std. Deviation</i>	<i>Coef. of Variation</i>	<i>Mean Qualifying Population</i>	<i>Minimum Qualifying Population</i>	<i>Maximum Qualifying Population</i>
Mammography with last 2 years, ages 42–65 (HEDIS)	0.65	0.05	0.08	46,372	1,570	485,747
DMARD treatment for rheum. arthritis (HEDIS)	0.89	0.04	0.04	360	8	3,647
Bronchodilator within 30 days of COPD diag. (HEDIS)	0.80	0.06	0.07	96	2	727
Major depression prescription drug treatment and adherence (HEDIS)	0.64	0.05	0.09	616	26	6,278
Appropriate antibiotic prescribed for bacterial pneumonia (PCPI)	0.69	0.05	0.08	1,326	66	14,065
Yearly hemoglobin a1C test for patient with diabetes (HEDIS)	0.85	0.06	0.08	11,505	420	135,811
Avoidance of lower back imaging for patient with lower back pain (HEDIS)	0.73	0.07	0.09	10,262	410	92,594
Acute prevention quality indicator, per 10,000 (AHRQ)	0.19	0.07	0.38	116,092	4,595	1,232,236
Chronic prevention quality indicator, per 10,000 (AHRQ)	0.27	0.09	0.34	116,092	4,595	1,232,236
30-day readmission rate	0.08	0.01	0.08	8,020	268	86,412

*Notes.* Table displays summary statistics of quality measures in the commercial market across 306 HRRs. Readmissions and PQIs are adjusted for age, sex, and health status. PQIs are presented per 10,000 instead of per 100 because they are relatively rare. High scores for readmissions and PQIs represent poor quality.

AHRQ, Agency for Healthcare Research and Quality; DMARDs, disease-modifying antirheumatic drugs; HEDIS, health care effectiveness data and information set; PCPI, Physician Consortium for Performance Improvement.

(loadings of 0.10 and 0.06, respectively). Finally, there were statistically significant, though modest, positive correlations between standardized factor scores and area-level spending  $\rho = .21, p < .001$  (Figure 1). The estimates imply a

Table 2: Correlation in Quality Measures Across HRRs, 2007–2009

	Mammography	DMARD Treatment for Rheum. Arthritis	Bronco COPD	Depress Tx	Pneu Antibx	Hemo A1C Test	Lower Back Imaging	30-Day Readmission	Chronic PQJ	Acute PQJ
Mammography	1.00									
DMARD Treatment for Rheum. Arthritis	0.29*	1.00								
Bronco COPD	0.30*	0.49*	1.00							
Depress Tx	0.42*	0.26*	0.30*	1.00						
Pneu Antibx	0.38*	0.61*	0.71*	0.53*	1.00					
Hemo A1C Test	-0.01	0.20	0.14	-0.03	0.21*	1.00				
Lower back imaging	0.25*	0.07	0.23	0.52*	0.32*	0.11	1.00			
30-day readmission	-0.09	-0.03	0.07	0.22	0.04	0.04	<0.01	1.00		
Chronic PQJ	0.29*	0.15	0.22	0.55*	0.41*	-0.05	0.62*	0.03	1.00	
Acute PQJ	0.24*	0.29*	0.22	0.45*	0.48*	-0.19*	0.67*	0.24*	0.85*	1.00

Notes: Table displays pairwise correlations of 10 quality measures across the 306 HRRs. PQJ composites and 30-day readmissions are adjusted for age, sex, age/sex interaction, and health status. Each measure was coded such that a higher score was associated with better performance for each of the indicators (high scores for readmissions and PQJs represent good quality). Bonferroni correction adjusted for multiple comparisons for the 45 pairwise comparisons (individual tests at  $\alpha = .001$ ).

DMARDs, disease-modifying antirheumatic drugs; PQJs, prevention quality indicators.  
\* $p < .001$ .

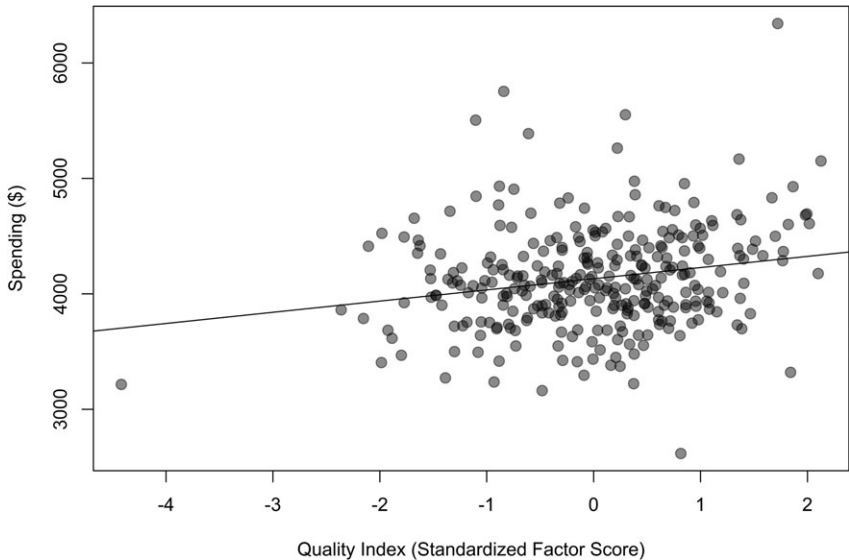


Table 3: Factor Analysis and Correlation with Total Spending, 2007–2009

	One-Factor Loadings	Correlation of Factor with Total Spending
Mammography	0.405	0.21*
DMARD treatment for rheum. arthritis	0.359	
Bronco COPD	0.430	
Depress Tx	0.631	
Pneu Antibx	0.670	
Hemo A1C	0.104	
Lower back imaging	0.640	
30-day readmission	0.062	
Chronic PQI	0.744	
Acute PQI	0.809	

Notes. Table displays the factor loadings of the 10 quality measures across the 306 HRRs. Loadings represent the correlations between each quality measure with the underlying factor. 29.6% of the variability of the 10 measures can be explained by the one-factor analysis. Standardized factors were correlated with total spending. DMARDs, disease-modifying antirheumatic drugs; PQIs, prevention quality indicators. \* $p < .001$ .

Figure 1: Assessment of Health Care Spending on Quality Index (Standardized Factor Score) across HRRs, 2007–2009



10 percent increase in spending was associated with a performance improvement of about 0.18 standard deviations.

## DISCUSSION

Consistent with evidence from Medicare, we find significant variation in quality for the commercially insured. Moreover, we find that areas have a weak “quality footprint”; areas that perform better than average on a given measure often perform better on others, although the correlation was generally modest. Our results show considerably greater geographic variation in the 30-day readmissions measure and the chronic and acute PQI measures than in the HEDIS and PCPI process measures. Given our study design, any explanation of this pattern would be speculative. It may be that a wider array of forces influences these broader outcomes, creating greater variation. In fact, evidence suggests that variation in socioeconomic factors may explain some variation in readmission rates (Kind et al. 2014; Barnett, Hsu, and McWilliams 2015; Sheingold, Zuckerman, and Shartzler 2016).

Our work has several limitations. First, quality measures based on administrative data are imperfect, as they cover limited clinical areas and are unable to account for enrollees who leave the database, possibly missing nonrandom attrition. Second, we do not observe specific providers or plans so we cannot attribute variation to specific organizations. Third, we use MarketScan, which is a large convenience sample of commercially insured individuals and may differ from the entire commercial population. Variation in patient and market characteristics that are not captured in the MarketScan data may contribute to the variation we report in this study. Additionally, the geographic coverage of MarketScan includes data from only one-quarter to one-fifth of all U.S. counties. However, approximately 70 percent of the U.S. population live in these counties, which are broadly representative of the country as a whole (Baker, Bundorf, and Kessler 2014). Our study has several strengths including a large geographically diverse population, capturing over 10 percent of the U.S. population across many health plans and employers. Additionally, it evaluates the commercially insured, a population that has not been thoroughly studied in the past. By focusing on disease cohorts, we reduce (but do not eliminate) concern about unmeasured health status. Finally, by using hierarchical models, we address concerns that sample size limitations could inflate measured variation.

Geographic variation in practice patterns continues to attract considerable interest. Much of the attention has focused on spending and much has

focused on the Medicare population. This analysis extends that work, looking at variation in quality provided to a commercially insured population. We find evidence of meaningful variation and a modest geographic footprint. This finding, if generalizable across populations, has ramifications for pay-for-performance programs. For example, assuming that much of the geographic variation is due to variation in provider performance, plans in lower performing geographies will have to devote more resources to obtain higher quality indicator scores than those in high-performing areas.

Overall, geographic variation research has been influential because it highlights potential inefficiencies in the health care sector and may illustrate the importance of geographic adjustments when evaluating health plan quality. Quality is an important aspect of this analysis, and quality variation in the commercial sector has been understudied. Like the other work in this literature, our paper only provides a broad sense of practice pattern differences. It can motivate attention but does not point to particular solutions. Other research will be needed to understand this variation and develop strategies to improve results in lower performing providers, and hence lower performing areas.

## ACKNOWLEDGMENTS

*Joint Acknowledgment/Disclosure Statement:* We are grateful to the Institute of Medicine and the Commonwealth Fund for funding the underlying research of this paper. Teresa Gibson holds a dual appointment with Truven Health Analytics, the organization that creates the MarketScan Databases. Drs. Chernen, Landrum, and Landon are employed by Harvard Medical School's Department of Healthcare Policy, and Dr. Fendrick by the University of Michigan Medical School. Mr. McKellar was employed by Harvard Medical School's Department of Healthcare Policy while the research was conducted.

*Disclosures:* None.

*Disclaimers:* None.

## REFERENCES

- American Medical Association [AMA]. 2012. "PCPI Approved Quality Measures" [accessed on August 1, 2013]. Available at <http://www.ama-assn.org/ama/pub/physician-resources/physician-consortium-performance-improvement/pcpi-measures.page>

- Baicker, K., K. S. Buckles, and A. Chandra. 2006. "Geographic Variation in the Appropriate Use of Cesarean Delivery." *Health Affairs* 25 (5): w355–67.
- Baicker, K., and A. Chandra. 2004a. "The Productivity of Physician Specialization: Evidence from the Medicare Program." *American Economic Review* 94 (2): 357–61.
- . 2004b. "Medicare Spending, the Physician Workforce, and Beneficiaries' Quality of Care." *Health Affairs* Jan-Jun (Suppl): W4-184-97.
- Baker, L. C., M. K. Bundorf, and D. P. Kessler. 2014. "Vertical Integration: Hospital Ownership of Physician Practices Is Associated with Higher Prices and Spending." *Health Affairs* 33 (5): 756–63.
- Barnett, M. L., J. Hsu, and J. M. McWilliams. 2015. "Patient Characteristics and Differences in Hospital Readmission Rates." *Journal of the American Medical Association Internal Medicine* 175 (11): 1803–12.
- Chandra, A. 2009. "Geography and the Keys to Health Care Reform." *Health Affairs*. [accessed on April 12, 2016]. Available at <http://healthaffairs.org/blog/2009/06/17/geography-and-the-keys-to-health-care-reform/>
- Chernew, M. E., L. M. Sabik, A. Chandra, T. B. Gibson, and J. P. Newhouse. 2010. "Geographic Correlation between Large-Firm Commercial Spending and Medicare Spending." *The American Journal of Managed Care* 16 (2): 131–8.
- Congressional Budget Office. 2008. *Geographic Variation in Health Care Spending*. Congress of the United States Congressional Budget Office [accessed on August 1, 2014]. Available at <https://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/89xx/doc8972/02-15-geoghealth.pdf>
- Davies, S. M., B. Duncan, J. J. Geppert, M. K. Gould, C. Haberland, P. A. Heidenreich, M. B. McClellan, K. M. McDonald, P. Romano, and M. Schleinitz. 2001. *Guide to Prevention Quality Indicators*. Rockville, MD: Agency for Healthcare Research and Quality.
- DxCG RiskSmart Stand Alone Software. 2013. *DxCG Budgeting & Underwriting Bundle for the Commercial*. Waltham, MA: Medicaid and Medicare Populations.
- Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and E. L. Pinder. 2003a. "The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care." *Annals of Internal Medicine* 138 (4): 273–87.
- . 2003b. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care." *Annals of Internal Medicine* 138 (4): 288–98.
- Franzini, L., O. Mikhail, and J. Skinner. 2010. "McAllen and El Paso Revisited: Medicare Variations Not Always Reflected in the Under-Sixty-Five Population." *Health Affairs* 29 (12): 2302–10.
- Franzini, L., O. I. Mikhail, M. Zezza, I. Chan, S. Shen, and J. D. Smith. 2011. "Comparing Variation in Medicare and Private Insurance Spending in Texas." *American Journal of Managed Care* 17 (12): E488–95.
- Franzini, L., S. Taychakhoonavudh, R. Parikh, and C. White. 2015. "Medicare and Private Spending Trends from 2008 to 2012 Diverge in Texas." *Medical Care Research and Review: MCRR* 72 (1): 96–112.

- Goldstein, H., W. Browne, and J. Rasbash. 2002. "Partitioning Variation in Generalized Linear Multilevel Models." *Understanding Statistics* 1: 223–32.
- Harvard University. 2012. *Geographic Variation in Health Care Spending, Utilization, and Quality among the Privately Insured*. Washington, DC: Institute of Medicine.
- Hassett, M. J., B. A. Neville, and J. C. Weeks. 2014. "The Relationship between Quality, Spending and Outcomes among Women with Breast Cancer." *Journal of the National Cancer Institute* 106 (10): 11–21.
- Institute of Medicine [IOM]. 2013. *Variation in Health Care Spending: Target Decision Making, Not Geography*. Washington, DC: Institute of Medicine.
- Jones, S. W., M. R. Patel, D. Dai, S. Subherwal, J. Stafford, S. Calhoun, and E. D. Peterson. 2012. "Temporal Trends and Geographic Variation of Lower-Extremity Amputation in Patients with Peripheral Artery Disease: Results from U.S. Medicare 2000-2008." *Journal of the American College of Cardiology* 60 (21): 2230–6.
- Kind, A. J. H., S. Jencks, J. Brock, M. Yu, C. Bartels, W. Ehlenbach, C. Greenberg, and M. Smith. 2014. "Neighborhood Socioeconomic Disadvantage and 30-Day Rehospitalization: A Retrospective Cohort Study." *Annals of Internal Medicine* 161 (11): 765–74.
- Landrum, M. B., E. Meara, A. Chandra, E. Guadagnoli, and N. L. Keating. 2008. "Is Spending More Always Wasteful? The Appropriateness of Care and Outcomes among Colorectal Cancer Patients." *Health Affairs* 27 (1): 159–68.
- Medicare Payment Advisory Commission. 2011. *Report to the Congress: Regional Variation in Medicare Service Use*. Washington, DC: Medicare Payment Advisory Commission [accessed on December 1, 2014]. Available at [http://www.medpac.gov/documents/reports/Jan11\\_RegionalVariation\\_report.pdf?sfvrsn=0](http://www.medpac.gov/documents/reports/Jan11_RegionalVariation_report.pdf?sfvrsn=0)
- Melnick, G., and E. Keeler. 2007. "The Effects of Multi-Hospital Systems on Hospital Prices." *Journal of Health Economics* 26 (2): 400–13.
- National Committee for Quality Assurance [NCQA]. 2012. *Quality Compass*. National Committee for Quality Assurance [accessed on December 1, 2014]. Available at <http://www.ncqa.org/HEDISQualityMeasurement/QualityMeasurementProducts/QualityCompass.aspx>
- National Quality Measures Clearinghouse. 2013. *National Quality Forum-Endorsed Measures*. Agency for Healthcare Research and Quality [accessed on December 1, 2014]. Available at <http://www.qualitymeasures.ahrq.gov/browse/nqf-endorsed.aspx>
- Pear, R. 2009. Health Care Spending Disparities Stir a Fight. *New York Times* [accessed on December 1, 2014]. Available at [http://www.nytimes.com/2009/06/09/us/politics/09health.html?\\_r=1&](http://www.nytimes.com/2009/06/09/us/politics/09health.html?_r=1&)
- Pope, G. C., J. Kautter, R. P. Ellis, A. S. Ash, J. Z. Ayanian, L. I. Lezzoni, M. J. Ingber, J. M. Levy, and J. Robst. 2004. "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model." *Health Care Financing Review* 25 (4): 119–41.
- Radley, D., S. How, A.-K. Fryer, D. McCarthy, and C. Schoen. 2012. *Rising to the Challenge: Results from a Scorecard on Local Health System Performance, 2012*. The Commonwealth Fund [accessed on August 1, 2014]. Available at <http://>

[www.commonwealthfund.org/~media/files/publications/fund-report/2012/mar/local-scorecard/1578\\_commission\\_rising\\_to\\_challenge\\_local\\_scorecard\\_2012\\_finalv2.pdf](http://www.commonwealthfund.org/~media/files/publications/fund-report/2012/mar/local-scorecard/1578_commission_rising_to_challenge_local_scorecard_2012_finalv2.pdf)

- Rosenthal, T. 2012. "Geographic Variation in Health Care." *Annual Review of Medicine* 63: 493–509.
- Sheingold, S. H., R. Zuckerman, and A. Shartzter. 2016. "Understanding Medicare Hospital Readmission Rates and Differing Penalties between Safety-Net and Other Hospitals." *Health Affairs* 35 (1): 124–31.
- Tsai, T. C., K. E. Joynt, E. J. Orav, A. A. Gawande, and A. K. Jha. 2013. "Variation in Surgical-Readmission Rates and Quality of Hospital Care." *New England Journal of Medicine* 369 (12): 1134–42.
- Virnig, B. A., N. Lurie, Z. Huang, D. Musgrave, A. M. McBean, and B. Dowd. 2002. "Racial Variation in Quality of Care among Medicare+Choice Enrollees." *Health Affairs* 21 (6): 224–30.
- Zhang, Y., K. Baicker, and J. P. Newhouse. 2010. "Geographic Variation in the Quality of Prescribing." *New England Journal of Medicine* 363 (21): 1985–8.
- Zuckerman, S., T. Waidmann, R. Berenson, and J. Hadley. 2010. "Clarifying Sources of Geographic Differences in Medicare Spending." *New England Journal of Medicine* 363 (1): 54–62.

## SUPPORTING INFORMATION

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

Appendix SA1: Author Matrix.