SPECIAL ARTICLE

Comparing Local and Regional Variation in Health Care Spending

Yuting Zhang, Ph.D., Seo Hyon Baik, Ph.D., A. Mark Fendrick, M.D., and Katherine Baicker, Ph.D.

From the Department of Health Policy and Management, Graduate School of Public Health, University of Pittsburgh, Pittsburgh (Y.Z., S.H.B.); the Department of Internal Medicine, School of Medicine, and the Department of Health Management and Policy, School of Public Health, University of Michigan, Ann Arbor (A.M.F.); and the Department of Health Policy and Management, Harvard School of Public Health, Boston (K.B.). Address reprint requests to Dr. Zhang at the Department of Health Policy and Management, University of Pittsburgh, 130 De Soto St., Crabtree Hall A664, Pittsburgh, PA 15261, or at ytzhang@pitt.edu.

N Engl J Med 2012;367:1724-31.
DOI: 10.1056/NEJMsa1203980

ABSTRACT

BACKGROUND

Wide geographic variation in health care spending has generated both concern about inefficiency and policy debate about geographic-based payment reform. Evidence regarding variation has focused on hospital referral regions (HRRs), which incorporate numerous local hospital service areas (HSAs). If there is substantial variation across local areas within HRRs, then policies focusing on HRRs may be poorly targeted.

METHODS

Using prescription drug and medical claims data from a 5% random sample of Medicare beneficiaries from 2006 through 2009, we compared variation in health care spending and utilization among 306 HRRs and 3436 HSAs. We adjusted for beneficiary-level demographic characteristics, insurance status, and clinical characteristics.

RESULTS

There was substantial local variation in health care (drug and nondrug) utilization and spending. Furthermore, many of the low-spending HSAs were located in high-spending HRRs, and many of the high-spending HSAs were in low-spending HRRs. For drug spending, only 50.7% of the HSAs located within the borders of the highest-spending quintile of HRRs were in the highest-spending quintile of HSAs; conversely, only 51.5% of the highest-spending HSAs were located within the borders of the highest-spending HRRs. Similar patterns were observed for nondrug spending.

CONCLUSIONS

The effectiveness of payment reforms in reducing overutilization while maintaining access to high-quality care depends on the effectiveness of targeting. Our analysis suggests that HRR-based policies may be too crudely targeted to promote the best use of health care resources. (Funded by the Institute of Medicine and others.)

A substantial body of evidence has emerged highlighting wide geographic variation in health care spending that is not driven by patient characteristics and is not associated with the quality of care or patient outcomes.[1-7] In light of this evidence, many policy proposals suggest targeting high-spending areas for lower Medicare payments or other coverage constraints, with a focus on areas such as the hospital referral regions (HRRs) described in the Dartmouth Atlas of Health Care.[1] These policies are predicated on the idea that there is a system-level component driving some areas to have high utilization and others low. The effectiveness of these policies in reducing overutilization while maintaining access to high-quality care depends on the effectiveness of targeting: if there is substantial variation across local areas within HRRs, then focusing on high-cost HRRs may leave many high-spending locales untouched while inadvertently penalizing some low-spending locales.

We compared variation in medical spending and prescribing patterns at the broader market level — the 306 HRRs that have been defined in the United States — with variation at the local level — the 3436 hospital service areas (HSAs) that compose the HRRs.[1] HRRs are the areas served by large tertiary care hospitals, where patients are referred for major cardiovascular surgical procedures and for neurosurgery. HSAs, located within HRRs, are local areas where the residents primarily use the hospitals located within the area. Thus, HSAs better capture the local health care markets, where Medicare beneficiaries receive most of their care.

We compared HSAs and HRRs with respect to variation in spending and use of prescription drugs and medical care, and we examined the degree to which high-spending HSAs are clustered together within high-spending HRRs. This analysis can help evaluate the effectiveness of policies targeted at different levels of aggregation. If the intention of such policies is to capture variation in local markets, it is important to understand the local-market heterogeneity that larger units may mask.

METHODS

DATA SOURCE AND STUDY SAMPLE

We obtained data on enrollment, pharmacy claims, and medical claims from 2006 through 2009 from the Centers for Medicare and Medicaid Services (CMS) for a 5% random sample of Medicare beneficiaries. For each year from 2007 through 2009, we identified all beneficiaries who had been enrolled for at least 1 month in Part A, Part B, and a stand-alone Part D prescription-drug plan (PDP), because CMS has both medical and drug data only for those enrolled in PDPs. (Data from 2006 were used for the calculation of prospective risk scores, as described below.) The resulting sample consisted of 1,013,477 beneficiaries in 2007, as well as 1,024,183 in 2008 and 1,022,662 in 2009, for a total of 3,060,322 beneficiary-year observations. On the basis of the beneficiary's ZIP Code of residence, we assigned each beneficiary to 1 of the 3436 HSAs, and hence to 1 of the 306 HRRs[1]; HSAs are nested within HRRs.

We conducted sensitivity analyses for two subpopulations: beneficiaries who were 65 years of age or older and those who were enrolled for the full year or until they died, to ensure that our results were not substantially affected by a small proportion of disabled beneficiaries or by those who switched plans during the year (see Table S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org). Coefficients of variation and ratios of the 75th to the 25th percentile were virtually unchanged.

The study was approved by the institutional review board at the University of Pittsburgh. Consent from participants was not required because we used deidentified secondary data.

The study outcomes were use of and spending for medical services and prescription drugs. All outcome measures were calculated in per-person-per-year units (with the spending by part-year enrollees annualized). For medication use, we defined two outcomes: total gross drug spending, including Part D plan payment before rebates, beneficiary out-of-pocket spending, and subsidy amount; and the number of monthly prescription drugs (1 if the supply was for ≤30 days, or the number of days divided by 30 if the supply was for >30 days). For medical services, we defined four outcomes: total nondrug medical spending, the number of inpatient admissions, the number of outpatient office visits, and the number of visits to the emergency department. Total nondrug medical spending included Medicare and beneficiary payments for all medical services (including inpatient care, outpatient care, physician visits, home health care, hospice care, skilled nursing home care, and medical devices) and was adjusted for local price-level differences with the use of county-level factor prices provided by the Medicare Payment Advisory Commission.[7,8] We did not adjust drug spending for regional price differences because the variation in drug prices among regions was negligible.[4]

ADJUSTMENT VARIABLES

To account for differences in population characteristics across regions, we adjusted for three major categories of beneficiary-level variables: demographic characteristics, income and insurance status, and clinical characteristics. Demographic characteristics included age in 5-year increments (from ≤34 years of age, 35 to 39, 40 to 44, and so on through 90 to 94, and ≥95), sex, and race or ethnic group (non-Hispanic white, non-Hispanic black, Hispanic, Asian, and other). Part D data have an enhanced race code from the Research Triangle Institute that is verified by first and last name algorithms, with much-improved sensitivity (>77%; κ coefficient, 0.79).[9]

We adjusted for individual-level insurance status and a proxy for income. We used variables indicating Medicaid coverage (available to beneficiaries with an income that is less than approximately 75% of the federal poverty level, but with some state variation) and Part D federal low-income subsidies (which vary on the basis of cutoff points in relation to the federal poverty level) to create income categories (<75% of the federal poverty level, 75 to 135%, >135 to 150%, and >150%). We also included two indicators for supplemental drug coverage, using Part D data: generic drug coverage in the "doughnut hole" gap, and coverage of both generic and brand-name drugs in the gap. We also controlled for share of the year that the beneficiary was enrolled. In addition, we adjusted for an indicator of disability for beneficiaries who were eligible for Medicare because of disability. Last, we controlled for income on the basis of the ZIP Code (logarithm of the median household income within the ZIP Code in which the beneficiary lived) and educational level (percentages of residents in the ZIP Code who had not completed high school, had completed high school only, and had attended or completed college).

Clinical characteristics included risk scores, indicators for institutionalization (defined as 90 days of care in a nursing home), and death during the year. We calculated the two prospective risk scores using the diagnoses and spending in the prior year: the CMS Hierarchical Condition Category (CMS-HCC) scores for nondrug medical care services, and the analogous scores for prescription drugs (CMS-RxHCC).[10] Risk scores are a proxy for health status, with higher scores indicating more severe illness and higher expected utilization of health care services (in our study sample, the CMS-RxHCC score ranged from 0 to 6.6 and the CMS-HCC score ranged from 0.1 to 12.4). We used prior-year instead of current-year diagnosis and spending to calculate the risk score (except for the 4% of beneficiaries in the sample who were new enrollees, for whom we used concurrent risk scores based on age and sex). Even with the use of the prior-year risk score, physician coding may be endogenous; that is, physicians in higher-spending regions may code patients as more severely ill than physicians in lower-spending regions code similar patients.[11] The results of a sensitivity analysis that excluded risk scores were quite similar (Tables S1, S2, and S3 and Fig. S1 in the Supplementary Appendix).

STATISTICAL ANALYSIS

We used these data to generate an adjusted average value for each outcome in each HSA or HRR. We pooled three years of data (2007, 2008, and 2009) and conducted an individual-level linear regression for each outcome.[12] Each regression included HSA or HRR indicator variables, year indicators, and the adjustment variables described above. Regressions were weighted by the percent of the year enrolled, so those data for beneficiaries who were enrolled for only part of the year would contribute less to the model. We then calculated the predicted value for each HSA or HRR using the estimating equation evaluated at national averages for the covariates, thus capturing variation at the HSA or HRR level that was purged of variation in population characteristics that we were able to control for with our observed covariates.

We used these adjusted HSA or HRR outcomes to perform two sets of analyses. First, we determined the degree of variation among HSAs and among HRRs, calculating statistics such as the ratio of the 75th to the 25th percentile and the coefficient of variation. The coefficient of variation, defined as the ratio of the standard deviation to the mean, is a normalized measure of dispersion. We included HSAs with 50 or more enrollees (only 2908 of the 3436 HSAs met this requirement) to avoid the introduction of noise driven by small cell sizes, and we used the same sample to conduct the HRR-level analysis.

Second, we evaluated the degree to which HSAs with similar spending levels clustered together within HRRs. Although there would be a correlation between spending at the HSA level and spending at the HRR level — HRRs are just an aggregation of HSAs — we assessed the degree of dispersion of HSA spending within and between HRRs. Specifically, we divided HRRs into quintiles on the basis of their adjusted spending and also divided HSAs into quintiles on the basis of their adjusted spending. We then tabulated the share of high-spending and low-spending HSAs located within the borders of high-spending and low-spending HRRs to gauge the variation in spending by HSAs within and between HRRs.

RESULTS

REGIONAL VARIATIONS IN ADJUSTED OUTCOMES

Table 1 presents the variation in drug and nondrug medical spending, counts of monthly prescriptions filled, inpatient admissions, outpatient office visits, and visits to the emergency department per person per year among different regions. Beneficiaries in the median HSA filled an average of 53 monthly prescription drugs per year, or 4.4 prescriptions per month, corresponding to \$2,912 in annual gross drug spending. Medical spending and drug spending were similarly variable, with coefficients of variation of 0.15. The ratio of drug spending at the 75th percentile to that at the 25th percentile was 1.21, whereas the corresponding ratio for the number of filled prescriptions was only 1.13. This suggests that the variation in the mix of drugs prescribed was larger than the variation in the number of drugs prescribed. Of course, part of this variation may be due to the severity of illness or other unmeasured patient characteristics.

For the analogous variation across HRRs, the pattern of variation across categories was quite similar, although the overall degree of variation was somewhat lower.

VARIATION ACROSS HSAs WITHIN HRRs

We gauged the degree to which HSAs with high spending were concentrated together in HRRs with high spending. A formal test of whether there was variation across HSAs nested within HRRs can reject the null hypothesis of no systematic HSA variation within HRRs (joint $F=4.86$, $P<0.01$). About 41% of the variation in adjusted HSA drug spending was between HRRs, and 59% was within HRRs. About 43% of the variation in adjusted HSA nondrug medical spending was between HRRs, with 57% within HRRs. (These results were virtually unchanged when we excluded the top and bottom 1% of the HSA values.)

There was substantial variation among HSAs within HRRs. For example, we found Manhattan to be one of the HRRs with the highest adjusted drug spending and Albuquerque, New Mexico, to be one of the lowest — but there was substantial dispersion in spending across the HSAs within those HRRs: the lowest-spending HSA in Manhattan had lower spending than about 25% of the HSAs within Albuquerque. For comparison, approximately 66% of the variation in HSA-level income was between HRRs, whereas 34% was within HRRs; approximately 55% of the variation in HSA-level education was between HRRs, and 45% was within; approximately 70% of the variation in HSA-level proportion of whites was between HRRs, and 30% was within; and approximately 66% of the variation in HSA-level CMS-HCC risk score was between HRRs, and 34% was within. Analysis at the HRR level thus masks more local variation in health care spending within the HRR than it does variation in several (although not all) covariates.

Figures 1A and 1B show the degree of clustering of high-spending HSAs within high-spending HRRs (see Table S2 in the Supplementary Appendix for a complete set of conditional probabilities). For HRRs in each quintile of adjusted HRR spending, Figure 1A shows the share of the HSAs within that HRR that were high-spending or low-spending HSAs. For example, for adjusted drug spending, 50.7% of the HSAs located within the highest-spending HRR quintile were in the highest-spending quintile of HSAs; 50.3% of the HSAs in the lowest-spending HRR quintile were in the lowest-spending HSA quintile. Similar patterns were observed for adjusted medical spending: 57.4% of the HSAs in the highest-spending HRR quintile were in the highest-spending quintile of HSAs; 54.0% of the HSAs in the lowest-spending HRR quintile were in the lowest-spending HSA quintile.

Table 1. Variation in Adjusted Medicare Drug and Nondrug Medical Spending, Number of Monthly Prescriptions Filled, Inpatient Admissions, Outpatient Office Visits, and Emergency Department Visits, All in Per-Person-Per-Year Units, in Hospital Service Areas and Hospital Referral Regions.*

Variable	Minimum	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile	Maximum	Mean	Ratio of 75th Percentile to 25th Percentile	Coefficient of Variation†
Hospital service areas										
Drug spending (U.S. $)‡	1,001	2,398	2,645	2,912	3,198	3,445	5,261	2,931±433	1.21	0.15
Monthly prescriptions filled (no.)§	28.26	46.86	49.48	52.60	55.69	58.78	75.39	52.72±4.89	1.13	0.09
Medical spending (U.S. $)¶	6,487	9,858	10,791	11,812	12,964	14,144	23,888	11,961±1,798	1.20	0.15
Inpatient admissions (no.)	0.15	0.34	0.38	0.43	0.49	0.55	0.88	0.44±0.09	1.28	0.20
Outpatient office visits (no.)	2.02	5.25	5.92	6.59	7.30	8.03	12.86	6.62±1.18	1.23	0.18
Emergency department visits (no.)	0.00	0.55	0.62	0.71	0.80	0.91	2.40	0.72±0.16	1.28	0.22
Total spending (U.S. $)	8,871	12,614	13,601	14,767	15,988	17,245	28,178	14,892±1,931	1.18	0.13
Hospital referral regions										
Drug spending (U.S. $)‡	2,347	2,665	2,787	2,952	3,136	3,275	4,211	2,972±268	1.13	0.09
Monthly prescriptions filled (no.)§	44.37	48.12	50.45	52.42	54.18	55.99	64.04	52.32±3.05	1.07	0.06
Medical spending (U.S. $)¶	8,123	10,403	11,081	11,766	12,624	13,580	19,478	11,912±1,369	1.14	0.11
Inpatient admissions (no.)	0.30	0.36	0.39	0.43	0.47	0.50	0.61	0.43±0.05	1.20	0.12
Outpatient office visits (no.)	3.64	5.62	6.11	6.62	7.10	7.78	10.74	6.66±0.91	1.16	0.14
Emergency department visits (no.)	0.35	0.60	0.65	0.70	0.75	0.80	1.01	0.70±0.08	1.16	0.12
Total spending (U.S. $)	10,656	13,255	13,849	14,782	15,680	16,586	23,689	14,884±1,477	1.13	0.10

* Plus–minus values are means ±SD. Data are for beneficiaries in the 2908 hospital service areas that have more than 50 beneficiaries (of a total of 3436 hospital service areas). To account for the differences in population mix across regions, we adjusted for three major categories of beneficiary-level variables: demographic characteristics, insurance status, and clinical characteristics in the individual-level regressions for each outcome. All outcomes shown are after adjustment for the differences in population mix.

† The coefficient of variation, defined as the ratio of the standard deviation to the mean, is a normalized measure of dispersion.

‡ Total drug spending included Part D plan payment before rebates, beneficiary out-of-pocket spending, and subsidy amount. We did not adjust drug spending for regional price differences because the variation in drug prices among regions was negligible.

§ The number of monthly prescription drugs was 1 if the supply was for 30 days or fewer, and it was the number of the days divided by 30 if the supply was for more than 30 days.

¶ Total medical spending included Medicare and beneficiary payment for all medical services (including inpatient admissions, outpatient office visits, physician visits, home health care, hospice care, skilled nursing home care, and medical devices) and was adjusted for local price-level differences with the use of county-level factor prices provided by the Medicare Payment Advisory Commission.

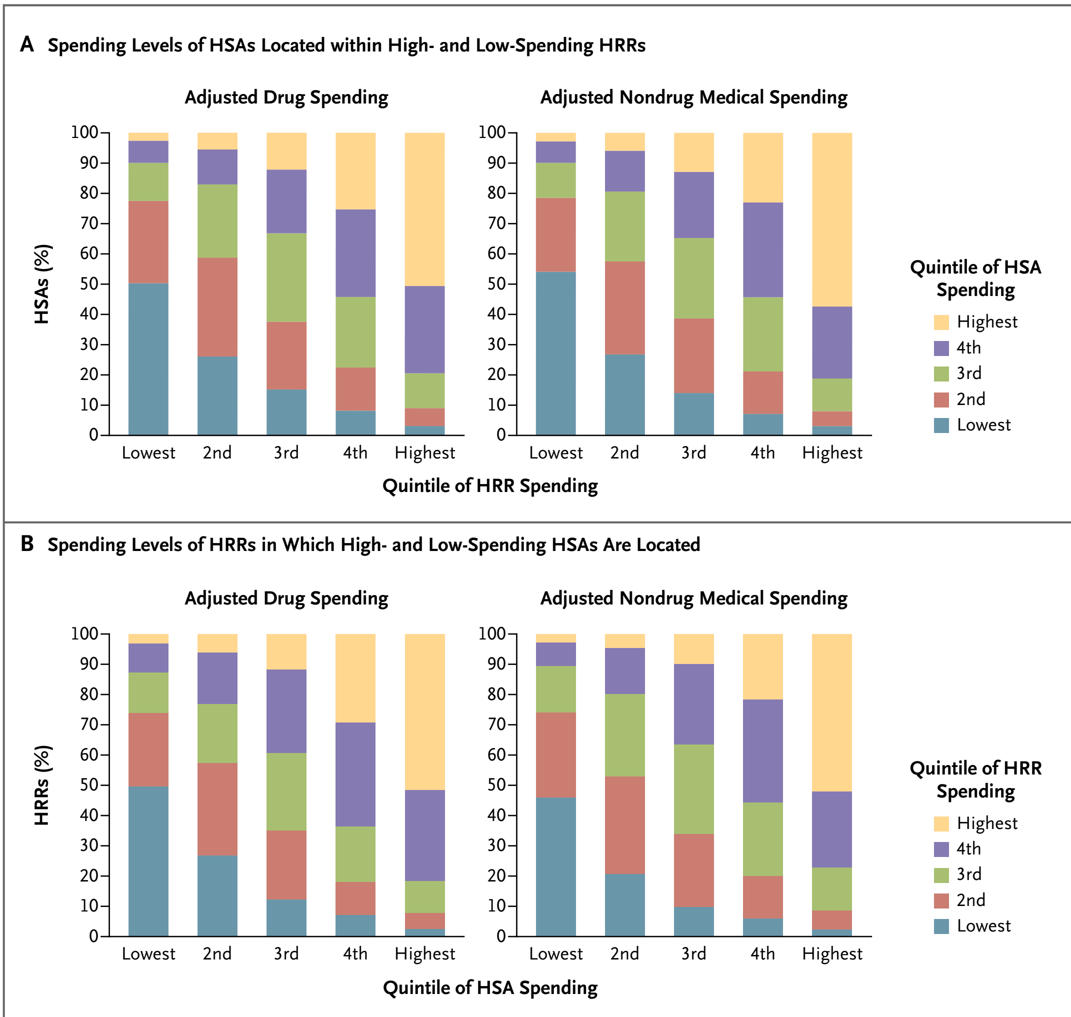

Figure 1. Quintile Analysis of Adjusted Drug Spending and Nondrug Medical Spending among Hospital Referral Regions (HRRs) and Hospital Service Areas (HSAs).

For each quintile of HRR spending, Panel A shows the distribution of spending quintiles of HSAs located within the HRRs. For example, with respect to adjusted drug spending, 50.7% of the HSAs located within HRRs in the highest-spending quintile of HRRs are in the highest-spending quintile of HSAs, and 50.3% of the HSAs located within HRRs in the lowest-spending quintile of HRRs are in the lowest-spending quintile of HSAs. For each quintile of HSA spending, Panel B shows the distribution of HRR spending quintiles. For adjusted drug spending, 51.5% of the HSAs in the highest-spending quintile are located within HRRs in the highest-spending quintile of HRRs, and 49.6% of the lowest-spending HSAs are located within HRRs in the lowest-spending quintile of HRRs.

For HSAs in each quintile of HSA spending, Figure 1B shows the distribution of HSAs according to the quintile of HRR spending. For adjusted drug spending, 51.5% of the HSAs in the highest-spending quintile were located within the highest-spending quintile of HRRs; 49.6% of the HSAs in the lowest-spending quintile were located within the lowest-spending quintile of HRRs. For adjusted medical spending, 52.0% of the highest-spending HSAs were located within the highest-spending HRRs, and 46.0% of the lowest-spending HSAs were located within the lowest-spending HRRs.

In sum, there is substantial misalignment of high-spending HSAs and HRRs. Many low-spending HSAs are located within the borders of high-spending HRRs, and many high-spending HSAs are located within the borders of low-spending HRRs.

Discussion

Much policy attention has been drawn to the large and persistent geographic variation in health care spending — for good reason. The presence of such variation (in the absence of commensurate variation in patient needs or even in health outcomes) suggests that high-intensity practice patterns in some areas signal an inefficient use of resources. This has led to discussion of policy levers that would rein in spending in high-utilization, high-cost areas, such as the lowering of Medicare payments to providers in those areas. Such policy levers are meant to focus on a level of aggregation that captures the local health care delivery system — too low a level (e.g., a lever aimed at individual physicians) could miss the system-level factors that account for high utilization in some areas, whereas too high a level (a lever aimed at multiple systems and markets) could reward or punish utilization that is well beyond the control of providers and could mask substantial heterogeneity.

Analyses have primarily been focused on variation among HRRs — areas that are defined on the basis of large tertiary care facilities and that incorporate numerous HSAs. There are advantages to looking at such large areas (e.g., they may be large enough to capture relatively homogeneous patient pools), but the disadvantage of such an exclusive focus is that it can mask substantial heterogeneity at the more local level. This is particularly important in considering the effects of policy levers that aim to influence local practice patterns for primary care or reduce avoidable hospitalizations, for example.

We examined the degree of heterogeneity within HRRs. We found that there was substantial local variation in utilization and spending for pharmaceutical and medical services and that there was substantial dispersion with respect to the level of local spending within HRRs.

These findings are, of course, subject to several limitations. First, our analysis is based on the Medicare population. Patterns among the commercially insured may differ. Medicare does, however, account for 20% of all national health care spending as of 2010,[13] and many of the policy levers that have been under discussion apply to Medicare payment rates. Second, our risk-adjustment variables are imperfect, and we did not capture patient preferences.[14] To the extent that these preferences vary across local areas, they could drive some of the observed patterns of heterogeneity. It is somewhat reassuring, however, that the patterns of unadjusted outcomes were quite similar to the patterns of adjusted outcomes. Third, causal connections are inherently difficult to draw from ecologic data. Although the variation described here (and elsewhere in the literature) is strongly suggestive of the inefficient use of resources, it is difficult to use these data to forecast the effect of different policy levers on spending patterns.

Nevertheless, these findings do have policy implications. Policies that aim to reduce spending in high-cost areas by targeting high-spending HRRs may fail with respect to both sensitivity and specificity: about half the HSAs located in the highest-spending HRRs are not the highest-spending HSAs, and about half the HSAs located in the lowest-spending HRRs are not the lowest-spending HSAs. That said, higher-spending HSAs are generally in higher-spending HRRs. Whether this degree of concordance is sufficient to achieving policy goals depends on the tolerance of policymakers for the ramifications of imperfect targeting.

These findings do not tell us what the "right" level of aggregation for policy is. There is clearly variation in spending even within HSAs — should policy focus on an even more local level? The movement toward accountable care organizations aims to tie payments to the care delivered by provider groups that are large enough to pool risk and average out individual-level variation in needs and idiosyncratic outcomes but small enough to hold the group accountable for the use of resources. In the absence of formal accountable care organizations, payments tied to the practice patterns in local areas are intended to accomplish similar goals. This analysis suggests that policies focused exclusively on the HRR may be too blunt to promote the best use of health care resources.

Supported by grants (to Dr. Zhang) from the Institute of Medicine (HHSP22320042509XI), National Institute of Mental Health (RC1 MH088510), and Agency for Healthcare Research and Quality (R01 HS018657).

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

REFERENCES

1. The Dartmouth atlas of health care. Lebanon, NH: The Dartmouth Atlas Project, 2010.
2. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending. Part 2: health outcomes and satisfaction with care. Ann Intern Med 2003;138:288-98.
3. *Idem.* The implications of regional variations in Medicare spending. Part 1: the content, quality, and accessibility of care. Ann Intern Med 2003;138:273-87.
4. Zhang Y, Baicker K, Newhouse JP. Geographic variation in Medicare drug spending. N Engl J Med 2010;363:405-9.
5. *Idem*. Geographic variation in the quality of prescribing. N Engl J Med 2010;363:1985-8.
6. Donohue JM, Morden NE, Gellad WF, et al. Sources of regional variation in Medicare Part D drug spending. N Engl J Med 2012;366:530-8.
7. Report to the Congress: regional variation in Medicare service use. Washington, DC: Medicare Payment Advisory Commission, 2011 (http://www.medpac.gov/documents/Jan11_RegionalVariation_report.pdf).
8. Report to the Congress: measuring regional variation in service use. Washington, DC: Medicare Payment Advisory Commission, 2009 (http://www.medpac.gov/documents/Dec09_RegionalVariation_report.pdf).
9. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries: final report. Research Triangle Park, NC: RTI International, 2008 (http://www.ahrq.gov/qual/medicareindicators/medicareindicators.pdf).
10. Risk adjustment: 2011 RxHCC model software. Baltimore: Centers for Medicare and Medicaid Services, 2011 (https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk_adjustment.html).
11. Song Y, Skinner J, Bynum J, Sutherland J, Wennberg JE, Fisher ES. Regional variations in diagnostic practices. N Engl J Med 2010;363:45-53. [Erratum, N Engl J Med 2010;363:198.]
12. Buntin MB, Zaslavsky AM. Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures. J Health Econ 2004;23:525-42.
13. Martin AB, Lassman D, Washington B, Catlin A. Growth in US health spending remained slow in 2010: health share of gross domestic product was unchanged from 2009. Health Aff (Millwood) 2012;31:208-19.
14. Doyle JJ. Returns to local-area health care spending: using health shocks to patients far from home. Am Econ J Appl Econ 2011;3:221-43.