Profiling Efficiency and Quality of Physician Organizations in Medicare

Gregory C. Pope, M.S. and John Kautter, Ph.D.

This article presents a methodology for profiling the cost efficiency and quality of care of physician organizations (POs). The method is implemented for the Boston metropolitan area using 2002 Medicare claims. After adjustments for case mix and other factors, 4 of 30 organizations are identified with different than average efficiency. Twenty-one of 30 organizations are identified with a different composite quality of care than average. Without changes in PO behavior, the gains from redirecting patients from lower to higher efficiency and quality providers are likely to be limited.

INTRODUCTION

Measuring, or profiling, physician efficiency and quality of care is central to several current initiatives in health care. Pay-for-performance and pay-for-quality programs rely on identifying and rewarding physician efficiency and quality performance (Rosenthal and Epstein, 2006). In provider tiering, insurers rank providers by efficiency and quality (Robinson, 2003). Consumer-directed health plans presume that consumers will have access to information on provider efficiency and quality to help them wisely spend health care dollars (U.S. Government Accountability Office, 2006). In addition, physician profiling is widely used by insurers to give physicians feedback on their performance, and to

select physicians for insurers' networks (Sandy, 1999). The Medicare Payment Advisory Commission (MedPAC) has called for physician profiling to be implemented by the Medicare Program (Medicare Payment Advisory Commission, 2005).

Several studies of physician profiling have appeared in the literature (Cave, 1995; Tucker et al., 1996; Hofer et al., 1999; Thomas, Grazier, and Ward, 2004; Medicare Payment Advisory Commission, 2006; Thomas, 2006), but much remains to be learned about the feasibility, validity, reliability, and usefulness of provider profiling. We conduct an exploratory study of physician profiling using comprehensive claims data on Medicare fee-for-service (FFS) beneficiaries from a single market area, defined as the Boston metropolitan statistical area (MSA). Our empirical results are specific to the sample market area and time period studied, however, our methods are generalizable to other markets and time periods. We selected the Boston MSA because it has a sufficiently large number of Medicare beneficiaries and POs to support our feasibility analysis, and because of our familiarity with this market, which helps us interpret and judge the face validity of our profiling results.

This study takes a population-based approach to profiling. Profiles are based on care provided to patients during a calendar year, not during individual episodes of care[1]. Both cost efficiency and process quality indicators (QIs) obtainable from claims are profiled. Because

[1] Refer to Leapfrog Group & Bridges to Excellence (2004) for a discussion of population-based versus episode profiling.

individual physician profiles are unreliable (Hofer et al., 1999), we profile POs.

METHODS

Data

In this study, we use 100 percent 2002 Medicare FFS Parts A and B claims and enrollment data for 350,000 Medicare beneficiaries residing in the Boston MSA. This MSA consists of seven counties—five in Massachusetts and two in New Hampshire. Our analysis sample includes all Medicare beneficiaries residing in the Boston MSA in 2002 who had at least one office or other outpatient evaluation and management visit, no months of Medicare private plan enrollment, who were continuously enrolled in both Parts A and B Medicare throughout 2002 (for decedents, through date of death in 2002), and who had Medicare as their primary insurance coverage. These restrictions create a sample of beneficiaries who can be assigned based on their office visits to a specific provider and who have complete Medicare Parts A and B claims so that expenditures and QIs are comparable across beneficiaries and organizations.

Area beneficiaries without any evaluation and management office visits in 2002 are not included in our analysis. Months of hospice enrollment are excluded because the curative phase of medical care has ended and standard QIs are less relevant. Also, we excluded beneficiaries entitled to Medicare by end stage renal disease from the efficiency profiling analysis because our case-mix adjustment model was calibrated only for aged and disabled beneficiaries.

Medicare expenditures are defined as Medicare payments to medical providers in 2002 for Medicare-covered services, excluding hospice. Expenditures are annualized and then weighted by the fraction of months in 2002 that a beneficiary is alive and eligible for Medicare. Per person expenditures were capped at \$100,000 to reduce the influence of outliers.

Identifying POs and Networks

POs are identified by their employer identification number (EIN). An EIN, also known as a Federal tax identification number, is a nine-digit number that the U.S. Internal Revenue Service (IRS) assigns to business entities. The IRS uses this number to identify taxpayers that are required to file various business tax returns. POs identified by EINs include solo practices, partnerships, traditional integrated physician group practices, physician/hospital organizations, hospital medical staffs or affiliated physicians, independent practice associations, management services organizations, medical foundations, and other organizational forms. Organizations identified by EINs can include multiple practice locations under common ownership or control. We did not contact the organizations to verify their structure, but we used a datafile available from CMS to crosswalk organizations' EINs to their names. The large organizations identified by EIN in our data had face validity according to the authors' knowledge of the local Boston market.

We define physician networks as groups of affiliated POs. We study mutually exclusive provider-sponsored networks such as integrated delivery systems, not highly overlapping insurer-sponsored networks. Physician networks vary in their degree of clinical and financial integration, standards for network membership, the amount of performance feedback they provide to network clinicians, the degree of practice standardization they attempt to impose on network physicians, and the extent of provider monitoring they engage in.

Physician networks are not identifiable in Medicare claims. We collected the names of POs affiliated with networks through publicly available information, including internet Web sites, newspapers, trade journals, and other media. EINs were used to identify the POs that are part of networks. While our method of identifying networks is not definitive, we believe it provides a largely accurate picture of the major physician networks in the Boston MSA.

Assigning Patients

Since our physician profiling simulation is conducted on the Medicare FFS population, beneficiaries are not enrolled in the profiled POs. However, for the purposes of profiling, it is necessary to attribute the services received by beneficiaries to specific POs. We assign beneficiaries to POs and networks that account for the largest share of their office and other outpatient evaluation and management visits (as measured by Medicare allowed charges). This algorithm assigns each Medicare patient to one and only one PO or network.

Our algorithm assigns beneficiaries to organizations and networks that provide a large enough share of their evaluation and management services to be held accountable for the efficiency and quality of their care. Most beneficiaries receive the majority of their outpatient evaluation and management care from a single PO, which is therefore in a position to coordinate their care. The average proportion of outpatient evaluation and management services that a Boston beneficiary receives within their assigned organization is 74 percent. Slightly under one-half of beneficiaries (44 percent) receive 80 to 100 percent, about one-half receive 40 to 79 percent, and fewer than 10 percent of beneficiaries receive less than 40 percent of services from their assigned organization. A previous study found that

POs believe they have primary responsibility for the health care of patients to whom they have provided the plurality of outpatient evaluation and management services (McCall, Pope, and Adamache, 1998).

Efficiency Profiling

Using the beneficiaries assigned to each PO, we define an efficiency index for the organization as:

$$\text{Efficiency Index} = \frac{\text{(Actual Per Capita Expenditures)}}{\text{(Predicted Per Capita Expenditures)}}.$$

If actual equals predicted, the efficiency index equals 1.00, meaning the observed expenditures of beneficiaries assigned to the PO equal the expenditures expected for these beneficiaries. In this case, the PO is neither efficient nor inefficient relative to expectations. If the efficiency index is less than 1.00, actual expenditures are less than predicted. The PO is more efficient than predicted. Conversely, if the index is greater than 1.00, the PO is less efficient than predicted. This is the standard statistic used in efficiency profiling exercises, that is often referred to as "observed/expected" (Thomas, Grazier, and Ward, 2004).

Predicted expenditures in our efficiency index are based on average expenditures in the Boston MSA, either unadjusted or adjusted for various factors. Hence, efficiency is measured relative to the average, not relative to the most efficient practices. An organization may be more or less efficient than average, that is, have an efficiency index above or below 1.00. An implicit assumption of the efficiency index is that expenditure variation that is not predicted is the result of variations in efficiency, not other unmeasured factors.

We calculate efficiency indexes unadjusted and adjusted for various cost factors. For the unadjusted index, predicted expenditures are the Boston MSA average

Table 1

Power of Statistical Tests of Physician Organization Efficiency, by Number of Assigned Patients and Efficiency, in the Boston Metropolitan Statistical Area: 2002

Percent Difference of Physician Organization from Average Efficiency (Better or Worse)	Number of Assigned Patients						
	100	500	1,000	2,000	5,000	10,000	20,000
	Percent						
5	1	2	4	7	19	43	80
10	2	7	15	33	80	99	100
15	3	17	38	74	99	100	100
20	5	33	68	96	100	100	100
25	8	54	89	100	100	100	100
30	12	74	98	100	100	100	100
50	43	100	100	100	100	100	100

NOTES: Two-sided test of the difference of the observed/expected expenditure ratio from 1. Significance level of the test is 1 percent. Standard deviation of the population is 2.08 (measured in Boston 2002 Medicare data and adjusted for case mix, geographic location, and hospital add-on payments). Power is the probability of rejecting the null hypothesis of average efficiency when it is false.

SOURCE: Pope, G.C. and Kautter, J., RTI International, Waltham, MA.

and the efficiency index simply indexes assigned beneficiary per capita expenditures relative to this average. For the case-mix adjusted index, expenditures are predicted using a concurrent version of the CMS hierarchical condition categories risk-adjustment model, or CMS-HCC model (Pope et al., 2004). The CMS-HCC model is the basis of risk adjustment of Medicare capitation payments to private health plans and predicts per capita expenditures using assigned beneficiary diagnoses recorded on Medicare claims and demographic characteristics from Medicare enrollment files. The concurrent version of the model uses 2002 diagnoses to predict 2002 expenditures, rather than the 2001 diagnoses used in the prospective version of the model. The concurrent version may be thought of as a case mix rather than a risk adjuster.

An efficiency index adjusted for geography is calculated by taking account of per capita expenditures in the county of residence of each assigned beneficiary when predicting their expenditures. County per capita expenditures reflect differences in intensity of care and in Medicare prices by county. POs drawing higher proportions of their patients from high-cost counties will therefore have higher predicted expenditures. Finally, we calculate an efficiency index excluding Medicare indirect medical education and disproportionate share payments to hospitals. Physician groups or networks whose assigned beneficiaries are disproportionately admitted to hospitals receiving these add-on payments may appear to be inefficient; excluding these subsidies may provide a more accurate measure of resource costs or quantity of services.

The difference of the efficiency index from 1.00 is tested for statistical significance to determine the likelihood that the observed deviation is due to random fluctuations in expenditures. Because we conduct a large number of statistical tests of significance, we use a 1-percent significance level in our statistical testing rather than the more usual 5 percent significance level. We conduct two-tailed tests of the statistical significance of the efficiency ratio, i.e., whether it is significantly greater than 1.00 or significantly less than 1.00.

Power Analysis of Efficiency Testing

The power of the statistical testing of the efficiency index is the probability that the statistical test will conclude that an organization's efficiency is different from average when it is in fact different from average. Eighty percent is the conventional standard for adequate power. Table 1 shows

the power of the statistical test of the efficiency index as a function of specified true differences of a PO from average efficiency and its number of assigned beneficiaries. With 100 patients, there is inadequate power to detect even a 50-percent deviation in efficiency. With 500 patients, about a 30-percent deviation can be adequately detected. With 2,000 patients, about a 15-percent difference can be detected, with 5,000 patients about a 10-percent difference, and with 20,000 patients, about a 5-percent difference.

The conclusion from Table 1 is that statistical testing of efficiency indexes can reliably detect only quite large deviations in efficiency among practices with small to moderate numbers of assigned patients (e.g., 30 percent or greater deviations from average among practices with 500 or fewer assigned patients). Applying efficiency profiling to small to moderate-sized practices may be unfair or require aggregation to larger profiling units. For example, according to Table 1, only 15 percent of practices with 1,000 assigned patients whose efficiency deviates from average by 10 percent will be detected by statistical testing of the efficiency index. Profiling will unfairly single out some practices as inefficient, while not identifying other equally inefficient practices.

To avoid these problems, we restrict our efficiency profiling to individual POs with at least 2,000 assigned patients. The specific sample size of 2,000 is somewhat arbitrary, but strikes a balance between profiling a reasonably large number of individual organizations and achieving adequate power for our statistical testing of efficiency deviations. Table 1 indicates that with 2,000 beneficiaries our statistical tests have adequate power at conventional levels (80 percent) to detect about a 15-percent deviation in efficiency.

Quality Profiling

As shown in Table 2, we calculate seven claims-based quality measures, including those for: (1) diabetes mellitus, (2) heart failure, (3) coronary artery disease, and

Table 2

Measures for Quality Profiling of Physician Organizations

Quality Indicator	Measure
Diabetes Mellitus	HbA1c Management: Percentage of diabetic patients with one or more HbA1c tests.
	Lipid Measurement: Percentage of diabetic patients with at least one low-density lipoprotein (LDL) cholesterol test.
	Urine Protein Testing: Percentage of diabetic patients with at least one test for microalbumin during the measurement year; or who had evidence of medical attention for existing nephropathy (diagnosis of nephropathy or documentation of microalbuminuria or albuminuria).
	Eye Exam: Percentage of diabetic patients who received a dilated eye exam or seven standard stereoscopic photos with interpretation by an optometrist or ophthalmologist or imaging validated to match diagnosis from these photos during the reporting year, or during the prior year if the patient is at low risk for retinopathy. A patient is considered low risk if the following criterion is met: has no evidence of retinopathy in the prior year. (This measure is adapted for claims data measurement.)
Heart Failure	Left Ventricular Ejection Fraction Testing: Percentage of patients hospitalized with a principal diagnosis of heart failure during the current year who had left ventricular ejection fraction testing during the current year.
Coronary Artery Disease	Lipid Profile: Percentage of coronary artery disease patients receiving at least one lipid profile during the reporting year.
Preventive Care	Breast Cancer Screening: Percentage of female beneficiaries age 50-69 who had a mammogram during the measurement year or the year prior to the measurement year.

NOTE: Two of the measures—eye exam and mammography—are specified as 2-year measures, but we used the single year of claims data available to us to calculate them, implying that our estimates of these measures are biased downward (lower bounds).

SOURCE: Centers for Medicare & Medicaid Services: Quality Measurement and Health Assessment Group: Data from the Doctors' Office Quality Project, 2005.

Table 3

Market Share of Physician Organizations, by Number of Assigned Medicare Fee-for-Service Beneficiaries Residing in the Boston Metropolitan Statistical Area: 2002

Number of Assigned Beneficiaries	Number of Organizations	Market Share[1]	Cumulative Market Share
		Percent	
10,000 or More	1	5.0	5.0
5,000-9,999	4	9.6	14.6
2,500-4,999	13	13.2	27.8
1,000-2,499	31	14.9	42.7
500-999	53	9.9	52.6
250-499	94	9.0	61.6
100-249	431	19.2	80.8
1-99	7,725	19.2	100.0

[1]Market share is share of total number of assigned beneficiaries.

SOURCE: Pope, G.C. and Kautter, J., RTI International, Waltham, MA.

(4) preventive care. The seven quality measures are the subset of those developed by CMS (2005) for the Doctors Office Quality (DOQ) project that can be calculated from administrative claims data. They have been well established and validated through the extensive review process conducted as part of the DOQ project. The DOQ measures are focused on care provided in ambulatory settings. Two of the measures—eye exam and mammography—are specified as 2-year measures, but we used the single year of claims data available to us to calculate them, implying that our estimates of these measures are biased downward (lower bounds).

We use the selected quality measures to conduct physician profiling on quality of care for Medicare FFS beneficiaries residing in the Boston MSA in 2002 and assigned to a PO. We profile POs relative to the Boston MSA average. For each of the individual QI rates, we perform a two-tailed statistical test at the 1 percent significance level of the difference between the PO rate and the Boston MSA rate.

In addition, we develop a composite quality score. While composite scoring has not been widely used in profiling health care services, research indicates aggregated measures may improve understanding of often complex profiling indicators by combining measures of many dimensions of care into a single score (Landrum, Bronskill, and Normand, 2000). We use a straightforward method to develop a composite quality score (Centers for Medicare & Medicaid Services, 2004). The numerators of all individual quality measures are summed to determine a composite numerator. The denominators of all individual quality measures are also summed to produce a composite denominator. The final composite score is produced by dividing the composite numerator by the composite denominator.

RESULTS

Market Structure

Table 3 provides information on the market structure of POs that serve Medicare FFS beneficiaries residing in the Boston MSA. More than 8,000 POs are identified by their EINs on claims of beneficiaries residing in the Boston MSA. But only 627 organizations have 100 or more assigned beneficiaries. The size distribution of Boston area POs is highly skewed, with only a few large organizations and many small organizations with a limited number of assigned patients. The very largest organizations control only a moderate portion

of the total MSA market, with a combined market share of only about 15 percent for the five organizations with 5,000 or more assigned beneficiaries. At the bottom end, about 40 percent of the market is comprised of practices with less than 250 assigned beneficiaries each.

Only 30 POs have 2,000 or more assigned beneficiaries. Two thousand assigned patients is the minimum that we establish for individual profiling of POs. These organizations jointly account for about one-third (35 percent) of the market. Thus, two-thirds of beneficiaries in the Boston area are assigned to smaller practices for which efficiency profiling is not very reliable on an individual practice basis. We group these smaller practices by organization size (defined as number of assigned Medicare patients) to examine the relationship of size to efficiency and quality. The 30 POs that we individually profile include large independent group practices and hospital or health system affiliated POs. Almost all of the organizations are multispecialty, including primary care physicians, but one is a group of oncologists specializing in cancer care.

We also analyzed the market structure of physician networks. The largest physician network has 59,082 assigned beneficiaries. Three other physician networks were identified, each having a market share of less than 4 percent. Each network has more than the minimum 2,000 beneficiaries we are requiring for efficiency profiling. However, only one-quarter of Boston beneficiaries are assigned to one of these four networks.

Efficiency Profiling

Table 4 summarizes the results of efficiency profiling for large POs with 2,000

Table 4

**Efficiency Indexes of Physician Organizations and Networks:
Summary Statistics for the Boston Metropolitan Statistical Area (MSA), 2002**

| Organization and Network | Efficiency Indexes | | | |
| | | Adjustments | | |
	Unadjusted	Case Mix	Case Mix and County	Case Mix, County, and Hospital Add-Ons[1]
Boston MSA Average	1.00	1.00	1.00	1.00
Large Organizations (_N_=30)[2]				
Mean	1.09	1.02	1.02	1.01
High	2.26	1.13	1.12	1.11
Low	0.69	0.81	0.88	0.90
Range	1.57	0.32	0.24	0.21
Percent Reduction in Range[3]	—	80	85	87
Standard Deviation	0.326	0.082	0.061	0.051
Number of Indexes ≠ 1.00[4]	18	9	4	4
Networks (_N_=4)				
Mean	1.11	1.05	1.04	1.03
High	1.31	1.10	1.08	1.06
Low	0.90	1.01	1.00	1.01
Range	0.41	0.09	0.08	0.05
Percent Reduction in Range[3]	—	78	80	88
Standard Deviation	0.173	0.044	0.041	0.022
Number of Indexes ≠ 1.00[4]	4	2	1	1

[1] Hospital add-ons are indirect medical education and disproportionate share payments.

[2] Organizations with 2,000 or more assigned Medicare patients.

[3] Relative to the unadjusted index.

[4] Number of efficiency indexes statistically significantly different from 1.00 at the 1 percent level of significance.

SOURCE: Pope, G.C. and Kautter, J., RTI International, Waltham, MA.

or more assigned beneficiaries, and for physician networks. All efficiency indexes are normalized such that they equal 1.00 for the Boston MSA as a whole. The mean unadjusted efficiency indexes for large organizations and for networks are greater than 1.00, indicating per capita patient costs above the area average. The range of the unadjusted efficiency index across individually profiled large organizations is 0.69 to 2.26, a 3 to 1 ratio. Across the more aggregated networks, the range is smaller, but still substantial, 0.90 to 1.31.

Adjusted for case-mix differences the variation in efficiency is dramatically compressed. The case-mix adjusted range in the efficiency index across large organizations is 0.81 to 1.13, less than a 50-percent variation. The range in adjusted network indexes is reduced to 1.01 to 1.10. Adjusting for case mix brings the mean index for large organizations and for networks much closer to 1.00, indicating that much of their unadjusted excess costs are due to a more expensive case mix of patients.

Geographic (county of beneficiary residence) and hospital payment add-on (removing indirect medical education and disproportionate share payments) adjustments further compress the measured range in efficiency, although not as dramatically as case-mix adjustment. With all three adjustments—case mix, county, and hospital add-ons—the range in organization efficiency indexes is 0.90 to 1.11. This represents a reduction of 87 percent in the range in the unadjusted efficiency index, and a similar reduction in its standard deviation.

The majority (18 of 30) of large organizations' unadjusted efficiency indexes are statistically different from the Boston average (1.00). Only 4 of the 30 fully adjusted indexes are statistically different from the area average. One of the four indexes, with a value of 1.11, belongs to the cancer care

organization that has an unusually sick patient mix that might not be fully adjusted for by our case-mix adjuster. Excluding this organization, the only two organizations identified as statistically inefficient have actual costs that exceed predicted costs by only 4 and 6 percent. The single organization identified as statistically efficient has costs 10 percent less than predicted.

Similar observations apply to network efficiency. After all adjustments, measured efficiency is statistically different from the area average for only one network. This network's costs are 6 percent higher than predicted. Another network's costs are 4 percent higher than predicted, but this index is not statistically different from 1.00. It seems difficult to publicly single out the network with an index of 1.06 as the sole inefficient network.

After all adjustments, the mean efficiency index of large organizations (1.01) is nearly equal to the metropolitan area average (1.00), and the mean efficiency index of networks (1.03) is only slightly worse than average (1.00). When we calculated mean fully adjusted efficiency indexes for patients assigned to eight size ranges of organizations (from less than 100 assigned Medicare patients to 10,000 or more), we found little variation in efficiency by organization size. Mean efficiency varied only from 0.97 (for organizations with 100 to 249 assigned patients) to 1.04 (for organizations with 10,000 or more assigned patients— only 1 organization), and only these two extremes were statistically significantly different from the area average of 1.00.

Quality Profiling

Table 5 provides results for quality profiling of large POs and networks. For the quality measures used in this study, the Boston MSA averages range from 58 percent (eye exam and breast cancer screening) to 88

Quality Indicator Rates[1] of Physician Organizations and Networks:
Summary Statistics for the Boston Metropolitan Statistical Area (MSA), 2002

Organization and Network	Quality Indicator							
	Diabetes				Coronary Artery Disease Lipid Profile	Congestive Heart Failure Left Ventricular Function Testing	Breast Cancer Screening	Composite[2]
	HbA1c Management	Lipid Management	Urine Protein Testing	Eye Exam				
Boston MSA Average	87	69	69	58	64	88	58	67
Large Organizations (N=30)[3]								
Mean	89	69	70	58	61	89	59	67
High	96	93	86	75	79	98	69	79
Low	78	40	55	48	36	77	39	57
Range	18	53	31	27	43	21	30	22
Standard Deviation	4.6	12.6	7.3	5.5	10.7	5.0	6.5	5.6
Number of Indicators ≠ Boston Average[4]	16	18	10	6	21	5	14	21
Networks (N=4)								
Mean	87	68	70	56	61	89	59	66
High	89	69	72	60	66	93	62	68
Low	83	66	67	54	51	87	56	63
Range	6	3	5	6	15	6	6	5
Standard Deviation	2.9	1.3	2.2	2.6	6.9	2.9	2.5	2.2
Number of Indicators ≠ Boston Average[4]	2	1	1	0	2	1	1	2

[1] Percentage of eligible patients receiving service.

[2] Sum of number of patients in numerator of individual indicators divided by sum of number of patients in denominator.

[3] Organizations with 2,000 or more assigned Medicare patients.

[4] Number of quality indicators statistically significantly different from the Boston MSA mean at the 1 percent level of significance.

SOURCE: Pope, G.C. and Kautter, J., RTI International, Waltham, MA.

percent (left ventricular function testing), with a composite score of 67 percent. The composite score means that, across the seven QIs, eligible beneficiaries received the specified service at only 67 percent of the clinically indicated rate, leaving considerable room for improvement.

On average, the composite QI performance of the 30 large, individually profiled organizations equals the overall MSA performance of 67 percent. But QI performance varies substantially across the organizations. The range in performance for the composite index is 22 percentage points, from 57 to 79 percent, and is larger than that for five of the seven individual indicators. Lipid management and profile show the largest variation across organizations, and HbA1c testing the least. Composite performance was statistically different than the MSA average for 21 of the 30 large organizations, with 10 organizations having lower than average performance.

Although average performance of the four physician networks is also about equal to the MSA average, the differences in QI rates among the networks are much smaller than among the 30 large organizations. Network composite indicator rates range only from 63 to 68 percent. Nevertheless, composite indicator rates at two of the four networks differ statistically from the MSA average.

QI performance differs only modestly by organization size, as measured by number of assigned Medicare beneficiaries. The composite rate, 64 percent, is lower than average for patients assigned to

the smallest practices with less than 100 assigned beneficiaries. It is slightly above average—ranging from 68 to 71 percent, on average—for patients assigned to organizations in three size categories, from 250 to 2,499 patients. Then it is about average, 66 to 68 percent, for patients assigned to the largest organizations in three size categories with 2,500 or more patients. The larger size, and presumably greater resources, of the largest POs, and of physician networks, does not translate into better QI performance. But patients seen primarily by the smallest practices receive recommended services slightly less often than average. Organizations of all sizes can improve their QI performance.

We did not adjust PO QI rates for patient characteristics, instead applying the same expectations to all populations. Adjusting for patient characteristics could affect relative performance rankings, for example,
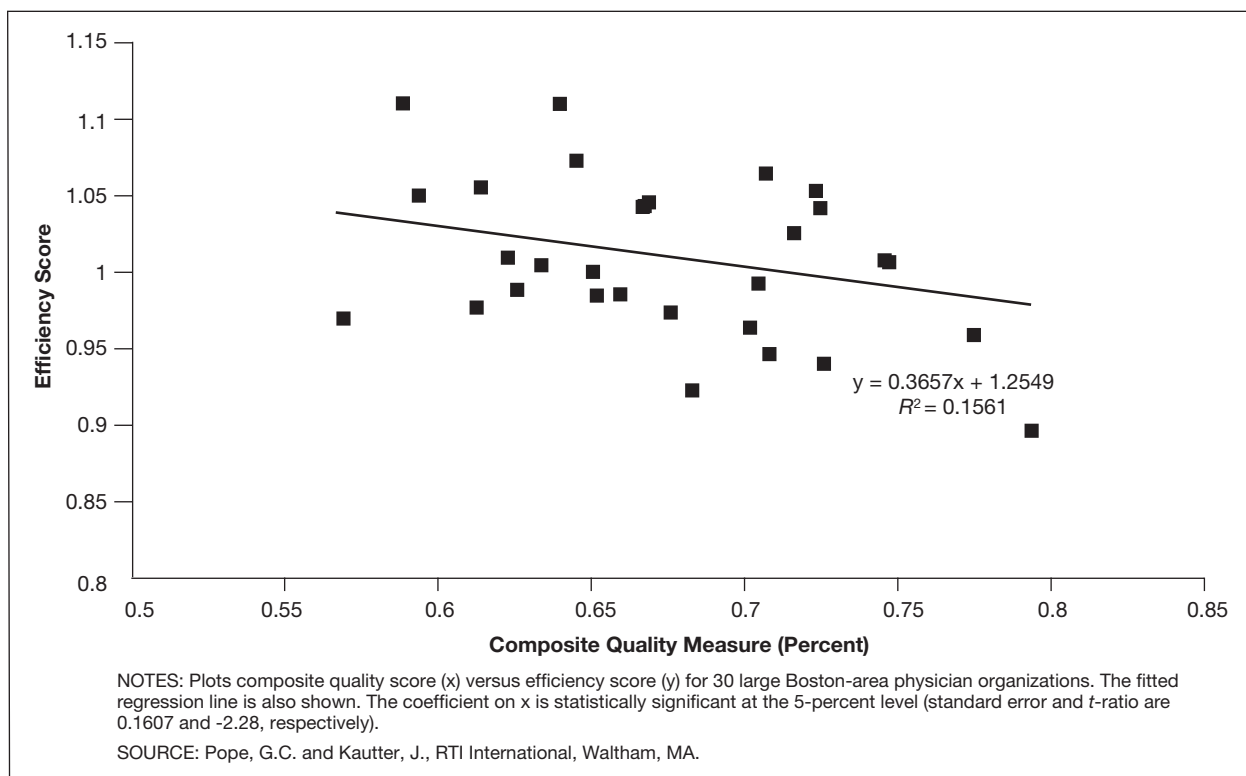
improving the ranking of organizations treating low socioeconomic status populations and narrowing the overall dispersion in performance. For example, the large organization with the lowest composite rate (57 percent, 10 percentage points lower than the MSA average) primarily serves a low-income, inner-city population.

Relationship Between Quality and Efficiency

If the highest quality POs are also the most efficient, Medicare, and other payers, could improve both the quality of care and its efficiency by directing patients to these providers. Also, providers that score high on both quality and efficiency could provide models for other providers to adopt.

Figure 1 shows the relationship between quality and efficiency for the 30 large POs in the Boston MSA. Quality is measured

Figure 1

Quality Versus Efficiency Scores of Large Physician Organizations in the Boston Metropolitan Statistical Area: 2002

$$y = 0.3657x + 1.2549$$
$$R^2 = 0.1561$$

NOTES: Plots composite quality score (x) versus efficiency score (y) for 30 large Boston-area physician organizations. The fitted regression line is also shown. The coefficient on x is statistically significant at the 5-percent level (standard error and *t*-ratio are 0.1607 and -2.28, respectively).

SOURCE: Pope, G.C. and Kautter, J., RTI International, Waltham, MA.

by each organization's composite quality score and efficiency by its fully adjusted efficiency index. In the figure, high values of the composite quality measure indicate relatively high quality, whereas high values of the efficiency index indicate relatively low efficiency (i.e., higher than predicted costs). There is a statistically significant positive relationship between quality and efficiency (Figure 1), with POs exhibiting higher quality also exhibiting greater efficiency (the *t*-statistic of the regression slope coefficient is -2.28, which is significant at the 5-percent level). The fitted line indicates that a 10-percentage point increase in the composite quality score is associated with a 3.7-percentage point decrease in the efficiency index, that is, a 3.7-percentage point reduction in cost of care. The relationship between quality and efficiency is statistically weak (the R^2 statistic is 16 percent), meaning there is considerable variation in the relationship between quality and efficiency among the profiled organizations.

CONCLUSIONS

We show that it is feasible to identify POs and their patients in Medicare claims, and profile the cost efficiency and process quality of care they provide. We find that patient case mix—and to a much lesser extent, geographic location and hospital payment add-ons—account for most (87 percent) of the variation in the per capita Medicare expenditures of patients assigned to large POs. After these adjustments, the efficiency of 4 of 30 large organizations differed statistically from the Boston MSA average. Residual expenditure variation—which could be the result of efficiency differences, or of unmeasured factors—was within 11 percent above or below the area average.

We conclude that some possible efficiency differences among large POs are identified, but proper adjustments greatly reduce the large initial cost differences. The potential savings from redirecting patients from inefficient to efficient organizations may be worthwhile to capture. But they do not appear to be particularly large given that only a few organizations were identified as having efficiency different from average. Transferring all patients from the few least to the few most efficient organizations (even if that were feasible) would affect only a small proportion of the total patient population.

Among the 30 large POs, 21 provided composite quality that differed statistically from the Boston MSA average. The range of composite quality was from 57 to 79 percent, compared to the MSA average of 67 percent. We conclude that there is meaningful variation in process quality among the large POs, and some potential for improving average quality through patient reallocations from low- to high-performing organizations. But as with efficiency, the quality differences are not so great, and the potential for patient movement so large, that substantial improvement in average quality can be attained by reallocating patients among organizations. Also, adjustment for patient characteristics affecting adherence to physician recommendations could narrow the observed differences among organizations. If observed quality differences among organizations are partly due to patient characteristics, patient movement from lower- to higher-quality organizations may improve overall average quality less than expected.

We find that POs exhibiting higher process quality also tend to exhibit greater cost efficiency. We estimate that a 10-percentage point increase in the composite quality score is associated with a 3.7-percentage point reduction in actual versus predicted

cost of care, on average. This implies that insurers will tend to improve both cost efficiency and process quality by redirecting patients to organizations exhibiting either one. But the correlation between quality and efficiency is weak (R^2 of 16 percent). Insurers will need to explicitly identify organizations scoring highly on both efficiency and quality to ensure that redirecting patients will enhance both.

Our conclusion that redirecting patients among organizations has only a modest potential to improve average performance assumes that organizations' efficiency and quality is static. However, the threat of losing patients may spur all organizations to improve their performance, and could result in a significant impact on average efficiency and quality. Similarly, explicit incentives for efficiency and quality improvements supplied to all organizations by pay-for-performance programs have the potential to significantly improve average performance (Rosenthal et al., 2004). In short, with static physician behavior the gains from redirecting patients are likely to be limited, but significant improvements might be realized from establishing dynamic market and financial incentives for improvement in all organizations over time.

This study has a number of limitations. We focus only on large POs, which account for about one-third of Boston area Medicare beneficiaries. Our statistical power analysis shows that only large POs with 2,000 or more assigned patients can be reliably profiled for efficiency. But feedback on large groups of physicians is less specific than profiling individual physicians or small practices. Methods of aggregating smaller practices—the extended hospital medical staff has been suggested, for example (Fisher et al., 2006)—or combining multiple years of data would need to be developed for efficiency profiling to be reliably applied to the majority of the beneficiaries in the market area.

We analyzed only one market area, the Boston area, which has several unique characteristics, such as a high concentration of teaching hospitals. Although our methodology is generalizable, our empirical findings may not generalize to other areas. We are constrained to QIs that can be measured in claims data, and consequently were only able to examine seven process indicators, four of which focus on diabetes. Our measure of cost of care, payments, is not a perfect measure of resource costs, although it is probably highly correlated with resources used and it is what payers care about. Our case-mix measure, although state-of-the-art, may not incorporate all relevant patient risk characteristics.

Our efficiency index, although commonly used, is not directly actionable by physicians. We rely on population-based profiling, which is more comprehensive and succinct, but less clinically detailed than episode-based profiling. Several feasible extensions would improve the actionability of population-based efficiency profiling. These could include comparing hospitalization rates of assigned patients to market-area norms, and profiling subpopulations of assigned beneficiaries such as congestive heart failure patients. Another limitation is that we assigned patients to POs based on their utilization, but some physicians may not feel responsible for quality and efficiency without explicit patient enrollment or choice of a primary care physician (McCall, Pope, and Adamache, 1998).

Finally, and importantly, health outcomes are not considered in our efficiency or quality measures. Efficiency is measured by the inputs used to treat patients with a certain diagnostic profile compared to the average resources used to treat them. Quality is measured by process

indicators. If organizations that use above-average amounts of resources achieve better patient outcomes, they may be the most efficient organizations. Similarly, organizations that achieve better outcomes, even if they score poorly on process indicators, may be the highest-quality organizations.

REFERENCES

Cave, D.G.: Profiling Physician Practice Patterns Using Diagnostic Episode Clusters. *Medical Care* 33(5):463-486, May 1995.

Centers for Medicare & Medicaid Services: *CMS Physician Focus Quality Initiative: Chronic Disease and Prevention Measures.* 2005. Internet address: http://www.cms.hhs.gov/PhysicianFocusedQual-Inits/downloads/PFQIClinicalMeasures200512.pdf (Accessed 2007.)

Centers for Medicare & Medicaid Services: *CMS HQI Demonstration Project: Composite Quality Score Methodology Review.* 2004. Internet address: http://www.cms.hhs.gov/HospitalQualityInits/downloads/HospitalCompositeQualityScoreMethod-ologyOverview.pdf (Accessed 2007.)

Fisher, E.S., Staiger, D.O, Bynum, J.P.W., et al.: Creating Accountable Care Organizations: The Extended Hospital Medical Staff. *Health Affairs* 26(1):w44-w57, 2006.

Hofer, T.P., Hayward, R.A., Greenfield, S., et al.: The Unreliability of Individual Physician "Report Cards" for Assessing the Costs and Quality of Care of a Chronic Disease. *Journal of the American Medical Association* 281(22):2098-2105, June 9, 1999.

Landrum, M.B., Bronskill, S.E., and Normand, S.T.: Analytic Methods for Constructing Cross-Sectional Profiles of Health Care Providers. *Health Services & Outcomes Research Methodology* 1(1):23-47, 2000.

Leapfrog Group & Bridges to Excellence: *Measuring Provider Efficiency, Version 1.0.* December 31, 2004. Internet address: http://www.bridgestoexcellence.org/assets/Documents/Program_Evaluation_Documents/Provider_Efficiency_Whitepaper/Measuring_Provider_Efficiency_Version1_12-31-20041.pdf (Accessed 2007.)

McCall, N.T., Pope, G.C., and Adamache, K.W.: *Research and Analytic Support for Implementing Performance Measurement in Medicare Fee for Service: First Annual Report.* Health Economics Research, Inc. Waltham, MA. 1998.

Medicare Payment Advisory Commission: *Report to the Congress: Increasing the Value of Medicare.* Washington, DC. 2006.

Medicare Payment Advisory Commission: *Report to the Congress: Medicare Payment Policy.* Washington, DC. 2005.

Pope, G.C., Kautter, J., Ellis, R.P., et al.: Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model. *Health Care Financing Review* 25(4):119-141, Summer 2004.

Robinson, J.C.: Hospital Tiers in Health Insurance: Balancing Consumer Choice with Financial Incentives. Web Exclusive. *Health Affairs* W3-135—W3-146, March 19, 2003. Internet address: http://content.healthaffairs.org/cgi/reprint/hlthaff.w3.135v1 (Accessed 2007.)

Rosenthal, M.B., Fernandopulle, R., Song, H.R., et al.: Paying for Quality: Providers' Incentives for Quality Improvement. *Health Affairs* 23(2):127-141, March/April 2004.

Rosenthal, M.B. and Epstein, A.M.: Pay for Performance in Commercial HMOs. *New England Journal of Medicine* 355(18):1895-1902, November 2, 2006.

Sandy, L.G.: The Future of Physician Profiling. *Journal of Ambulatory Care Management* 22(3):1-16, 1999.

Thomas, J.W.: Should Episode-Based Economic Profiles Be Risk Adjusted to Account for Differences in Patients' Health Risks? *Health Services Research* 41(2):581-598, 2006.

Thomas, J.W., Grazier, K.L., and Ward, K.: Economic Profiling of Primary Care Physicians: Consistency Among Risk-Adjusted Measures. *Health Services Research* 39(4):Part I, 985-1003, August 2004.

Tucker, A.M., Weiner, J.P., Honigfeld, S., et al.: Profiling Primary Care Physician Resource Use: Examining the Application of Case Mix Adjustment. *Journal of Ambulatory Care Management* 19:60-80, 1996.

U.S. Government Accountability Office: *Consumer-Directed Health Plans.* Pub. No. GAO-06-514. U.S. Government Printing Office. Washington, DC. April 2006.