



Published in final edited form as:

Health Aff (Millwood). 2010 August ; 29(8): 1532–1538. doi:10.1377/hlthaff.2009.1091.

Cost Profiling in Health Care: Should the Focus be on Individual Physicians or Physician Groups?

A Mehrotra, JL Adams, JW Thomas, and EA McGlynn

RAND, Pittsburgh, Pennsylvania, USA

A Mehrotra: mehrotra@rand.org

Abstract

In an effort to rein in rising health care costs, health plans are using physician cost profiles as the basis for tiered networks that encourage patients to visit low-cost physicians. There are concerns that physician cost profiles are often unreliable and some have argued that physician groups should be profiled instead. Using data from Massachusetts, we address this debate empirically. While we find that physician group profiles are more reliable, the group profile is not a good predictor of individual physician performance within the group. Better methods for cost profiling providers are needed.

To rein in rising healthcare costs, health care purchasers are pursuing a number of consumer-directed policy applications that depend on individual physician cost profiles. Cost profiles, in conjunction with quality profiles, are publicly reported to encourage patients to choose high-value (i.e. low cost and high quality) physicians.[1] Other consumer-directed initiatives include selective networks and tiered networks.[2,3] In a selective network, a patient can be reimbursed for care only when they visit a high-value physician. In a tiered network, patients pay a lower co-payment when they visit a high-value physician. All of these consumer-directed policy interventions encourage patients to visit low-cost physicians.

Another approach to decreasing costs is to focus policies directly on the providers. More than half of HMOs in the United States, representing more than 80% of persons enrolled, use pay-for-performance in their provider contracts.[4] While pay-for-performance typically focuses on quality, newer programs are increasingly incorporating cost or efficiency measures.[4,5]

Both types of policy interventions, consumer-directed and provider-directed, typically use individual physician cost profiles. Individual physicians are seen as the primary driver of costs, because they order the diagnostic tests and treatments patients receive.[6] Also, patients are most concerned with selecting individual physicians rather than hospitals or physician networks.[7] Some physicians and profiling organizations have argued against this focus on the individual physician. They argue that the focus should instead be on physician groups, partly because the measurement of physician group costs will be based on larger sample sizes and therefore will be more reliable.[7,8]

Along with validity, reliability is a key test of the performance of a measurement system. [9-13] Reliability is a measurement of signal-to-noise, and in cost profiling applications it indicates the degree to which the performance of one provider can be distinguished from that of another. If reliability is low, then there is a greater risk that a provider will be misclassified as average-cost when he or she is actually low-cost. In previous work we have demonstrated that the majority of physician cost profiles do not meet minimum levels of reliability, and this results in considerable misclassification (e.g. average-cost physician

being misclassified as low-cost).[14] This finding is echoed in a recent paper by Nywiede and colleagues where they found the vast majority of individual physicians and small practices did not have sufficient sample size to detect differences in costs.[15] To our knowledge, no one has estimated the reliability of physician group cost profiles.

In this study we address the ongoing policy debate of whether one should profile individual physicians vs. physician groups. Using Massachusetts data we compare the sample size and reliability of cost profiles at the physician and physician group level. We quantify the fraction of physicians not in a group and the fraction of care provided by these solo physicians. Lastly, we address the degree to which a patient can effectively use physician group cost profiles when choosing an individual physician.

METHODS

In creating physician and group cost profiles, our goal was to replicate methods commonly used by individual health plans. Our data sources, physician sample, and method for creating individual physician cost profiles have been described in previous publications.[16] A concise overview is provided below.

Data Sources

We constructed an aggregated commercial claims data set that included all professional, facility, ancillary, and pharmacy claims from four health plans in Massachusetts for 2004-2005. We analyzed all claims for the 1.13 million enrollees between the ages of 18 and 65 who were continuously enrolled for the two years.

Physician Sample

Our study population included Massachusetts physicians who submitted at least one claim to one or more of the four participating health plans and were in a non-pediatric, non-geriatric specialty with direct patient contact (e.g., excluding radiologists, pathologists). We used a unique physician identifier previously created by Massachusetts Health Quality Partners (MHQP) to link data from the four health plans at the physician level.[17]

Defining Groups

We used MHQP's designations of physician groups. MHQP defines a physician group as a distinct set of physicians that together contract with health plans and share resources and leadership (e.g. medical director). MHQP staff assigned each physician to a group based on an algorithm that used variables in the health plan enrollment files such as Medicare Unique Physician Identification Number (UPIN). Physicians not allocated to a group by the algorithm were assigned by manual inspection. Physician group leaders reviewed and offered corrections to their group's roster of physicians. MHQP also assigns physicians to practice sites. We conducted separate analyses at the practice level. These are not included in the paper as cost profiling at the practice site appeared to provide little advantage over profiling individual physicians or physician groups.

A minority of physicians (n=1241, 9.8% of sample), mostly specialists, were members of more than one group. With the data available to us, we could not determine the group in which these physicians delivered specific services. We therefore randomly assigned these physicians (and all the episodes assigned to them) to a single group. In a sensitivity analysis we reanalyzed our results deleting these physicians from our analysis. The results of this sensitivity analysis were substantively the same.

Creating Provider Cost Profiles

We use the term “provider” generically to refer to a physician or group. Our methodology for creating cost profiles has been described in more detail in previous publications.[16] In brief, the method consisted of the following steps:

1. We created a standardized state wide price for each service. To eliminate the bias of extreme values, we set all charges below the 2.5th percentile and above the 97.5th percentile of the distribution for each service to the values at those cut points; [18,19]
2. We used Episode Treatment Groups® Version 6.0 (ETGs®) software to aggregate each patient’s claims into clinically related episodes of care (Symmetry Health Data Systems, Inc, Phoenix, AZ);
3. The cost of each patient episode was calculated by summing the standardized costs of each service received within the episode;
4. The total cost of an episode of care was attributed to the physician who accounted for the highest fraction (minimum 30%) of professional costs within the episode. If the physician worked in a group, the episodes assigned to a physician were also assigned to this group;
5. For each type of episode an expected cost level was calculated as the mean cost across all episodes attributed to physicians of the same specialty, for patients with the same level of co-morbidities (as measured using Symmetry’s Episode Risk Groups®);
6. Based on all episodes attributed to the provider (physician, group) we calculated the cost profile as the *Sum of Observed Costs* divided by *Sum of Expected Costs*. Profile scores greater than 1 are higher than the average of the peer group, profiles less than 1 are lower than average (i.e., more “efficient” use of resources).

Assign providers to cost categories

To put providers into average or low-cost categories, we used a t-test to determine whether a provider’s composite cost profile score was statistically better than the average score among their peers.[20] For physicians, the peer group was physicians within the same specialty. For groups, the peer group was all groups in the state.

Analyses

We described the number of physicians and mix of specialty in the groups. We measured the reliability of both individual physician and physician provider group profiles using methodology described in our previous work.[14] Reliability in this context describes how confidently we can distinguish the performance of one provider from another. Conceptually, it is a ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. Unfortunately, the use of very different methods to estimate the same underlying concept can be a source of some confusion. We use a common method in which reliability is characterized as a function of the components of a simple hierarchical model. [21]

RESULTS

In our data 12,615 physicians in 28 specialties were assigned 2.9 million episodes. Of the 12,615 physicians, 9,716 (77%) worked in one of the 185 groups in the state. Exhibit 1 shows the score distribution of both physicians and physician groups. The distribution of scores for physician groups is narrower (25th to 75th percentile 0.92-1.05) than the distribution for individual physicians (25th to 75th percentile 0.81-1.15).

Size and Composition of Groups

There is significant heterogeneity in the composition and size of groups (Exhibit 2). Among the 185 physician groups, 22 (12%) are composed of 2-3 primary care physicians while 47 (25%) are composed of 50 or more physicians in multiple specialties. The largest three groups in the state have 910, 636, and 326 physicians, respectively (19% of the physicians in a group).

Reliability of Physician and Group Cost Profiles

The median reliability of group cost profiles is 0.91 (IQR 0.78-0.96). In previous work we reported that the median reliability of individual physician cost profiles was 0.53 (IQR 0.21-0.79).[14] Groups with a larger number of physicians generally have higher reliability. For example, among groups of 2-3 physicians, the median reliability is 0.76 while among groups with more than 50 physicians it is 0.97 (Exhibit 3).

Of the 2.9 million assigned episodes, 2.0 million (70%) were assigned to individual physicians having cost profiles with reliability of 0.70 or above, a commonly used minimum reliability standard.[9,10] Physician groups satisfying this reliability standard were assigned 2.5 million (86%) episodes.

Relationship Between Classification of a Group and Individual Physicians That Work in that Group

We classified each group as either low-cost or average-cost. Based on their profile scores we also classified individual physicians within these groups as low-cost or average-cost. Among individual physicians in low-cost groups only 16% were classified as low-cost based on their own performance (Exhibit 4). Among individual physicians in average-cost groups, 8% were classified as low-cost based on their own performance. In Exhibit 5, we show the score distribution of the physicians who work in average-cost and low-cost groups respectively. There is significant heterogeneity in the cost profiles among both sets of physicians and considerable overlap in their distribution.

DISCUSSION

In an effort to decrease health care costs, there is growing interest in using physician cost profiles for policy applications such as tiered and selective networks, public reporting, and pay-for-performance incentives. There is an ongoing policy debate about whether cost measures should be applied at the individual physician or physician group level. Our results highlight some of the pros and cons of the two approaches. Profiling physician groups may be appealing because the cost profiles are more reliable—which means that we can confidently distinguish one group from another. On the other hand, if profiling were done only at the group level, a notable fraction of solo physicians would be excluded from profiling efforts. Further, there is considerable heterogeneity in what constitutes a group, and knowing the performance of a group does not predict the performance of individual physicians within the group.

Our results have different implications depending on the target of the different policies that use cost profiles. For provider-directed policies such as pay-for-performance incentives or provider feedback, our results support profiling groups instead of individual physicians. Compared to the cost profiles of individual physicians, group cost profiles have higher reliability. This is primarily driven by the larger number of patients assigned to a group and consistent with prior work that found practices with more than 50 physicians could effectively be profiled.[15] The higher reliability among physician groups means that we are more confident in our ability to distinguish one group from another. In the context of pay-for-performance incentives, this greater reliability means that a physician group is more likely to correctly receive (or not receive) an incentive payment.

However, the ability to accurately classify groups into cost performance tiers does not provide an adequate signal for consumer-directed policies where patients select an individual physician within a group. In a tiered plan physician groups would be placed into tiers based on their cost and quality profiles and patients would be given an incentive to choose care from a low-cost physician group. Within low-cost groups there is substantial heterogeneity in the relative costs of individual physicians. It is therefore unlikely that the individual physician within the low-cost group that cares for the patient will also be low-cost.

Our results also highlight the difficulty of defining a group. For these analyses we used the existing definitions of a physician group as determined by MHQP. MHQP's definition is logical and group leaders verified the roster of physicians within their groups. Nonetheless, there are concerns with face validity when comparing the relative costs of a 3 physician primary care group to a 910 physician multi-specialty group. This heterogeneity is the reality of how physicians are organized in Massachusetts and elsewhere in the United States. When debating the relative advantage of profiling physician groups or individual physicians, this heterogeneity must be kept in mind. Another logistical barrier to group profiling is that in Massachusetts, a quarter of the physicians are in solo-practice. When profiling at the group level, it is unclear how a health plan should treat these physicians. Within a tiered health plan, one option would be to assume that these physicians are low-cost, but this might be perceived as unfair by physicians in average cost groups because it would be easier for solo practice physicians to be included in high-performance networks. It would also decrease the potential cost savings from the health plan's perspective. Another option would be to assume that the solo physicians are average cost, but this might also be criticized as unfair by solo physicians who would be excluded from high-performance networks under such a policy.

There are numerous limitations to our analyses. Of note, our data come from Massachusetts. In other states, a larger fraction of physicians may work independently and profiling at the group or practice level may not be feasible. Also, it is less likely that a validated categorization of physicians exists outside of Massachusetts. Our analyses do not address of which level of the health care system actually drives the cost variation, individual physicians or the groups in which they practice.[22] This is something we hope to address in future work.

Our research does not address other important aspects of cost profiling methodology. Other researchers have raised concerns about the validity of the claims data used to build the cost profiles[5] as well as the validity of the episode groupers used for cost profiles.[23] MedPAC found that in Miami overall patient costs are higher than average, but per episode costs are lower than average.[23] This is because overall patient costs are a product of the number of episodes and costs per episode. In Miami patients triggered a higher number of

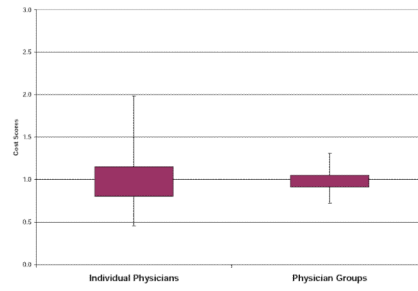
episodes possibly due to differences in physician billing patterns. More research is needed to address these validity concerns.

There is an ongoing policy debate on whether individual physicians or physician groups should be the focus of cost profiling. In previous work, we and others have raised concerns about the reliability of individual physician cost profiles. Though physician group profiles have the advantage of being more reliable, our findings highlight that they cannot be effectively used for consumer-directed policy applications as physician group profiles provide little useful information for patients. Together our findings from this work and previous research highlight the need for better methods for profiling providers (either physician or physician group) on their relative costs.

Endnotes

1. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med* 2008 Jan 15;148(2):111–23. [PubMed: 18195336]
2. Kazel R. Tiered physician network pits organized medicine vs. United. *Am Med News*. 2005 March 7;
3. Brennan TA, Spettell CM, Fernandes J, Downey RL, Carrara LM. Do managed care plans' tiered networks lead to inequities in care for minority patients? *Health Aff (Millwood)* 2008 Jul-Aug; 27(4):1160–6. [PubMed: 18607051]
4. Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM. Pay for performance in commercial HMOs. *N Engl J Med* 2006 Nov 2;355(18):1895–902. [PubMed: 17079763]
5. Robinson JC, Williams T, Yanagihara D. Measurement of and reward for efficiency In California's pay-for-performance program. *Health Aff (Millwood)* 2009 Sep-Oct;28(5):1438–47. [PubMed: 19738261]
6. Berwick DM. Measuring physicians' quality and performance: adrift on Lake Wobegon. *Jama* 2009 Dec 9;302(22):2485–6. [PubMed: 19996408]
7. Milstein A, Lee TH. Comparing physicians on efficiency. *N Engl J Med* 2007 Dec 27;357(26): 2649–52. [PubMed: 18160682]
8. Delmarva Foundation for Medical Care. Enhancing Physician Quality Performance Measurement and Reporting Through Data Aggregation: The Better Quality Information (BQI) to Improve Care for Medicare Beneficiaries Project. 2008. [cited May 13, 2009]; Available from: <http://www.cms.hhs.gov/bqi/>
9. Safran DG, Karp M, Coltin K, Chang H, Li A, Ogren J, et al. Measuring patients' experiences with individual primary care physicians. Results of a statewide demonstration project. *J Gen Intern Med* 2006 Jan;21(1):13–21. [PubMed: 16423118]
10. Hays, RD.; Revicki, D. Reliability and validity (including responsiveness). In: Fayers, P.; Hays, R., editors. *Assessing Quality of Life In Clinical Trials*. New York: Oxford University Press Inc; 2005.
11. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *Jama* 2001 Jul 25;286(4):415–20. [PubMed: 11466119]
12. Scholle SH, Roski J, Adams JL, Dunn DL, Kerr EA, Dugan DP, et al. Benchmarking physician performance: reliability of individual and composite measures. *Am J Manag Care* 2008 Dec; 14(12):833–8. [PubMed: 19067500]
13. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *Jama* 1999;281(22):2098–105. see comments. [PubMed: 10367820]
14. Adams JL, Mehrotra A, Thomas JW, McGlynn EA. Physician cost profiling--reliability and risk of misclassification. *N Engl J Med* 2010 Mar 18;362(11):1014–21. [PubMed: 20237347]
15. Nyweide DJ, Weeks WB, Gottlieb DJ, Casalino LP, Fisher ES. Relationship of primary care physicians' patient caseload with measurement of quality and cost performance. *Jama* 2009 Dec 9;302(22):2444–50. [PubMed: 19996399]

16. Adams, JL.; Mehrotra, A.; Thomas, JW.; McGlynn, EA. Detailed Methodology and Sensitivity Analyses. Santa Monica CA: RAND; 2010. Physician Cost Profiling - Reliability and Risk of Misclassification.
17. Friedberg MW, Coltin KL, Pearson SD, Kleinman KP, Zheng J, Singer JA, et al. Does affiliation of physician groups with one another produce higher quality primary care? *J Gen Intern Med* 2007 Oct;22(10):1385–92. [PubMed: 17594130]
18. Tukey JW. The Future of Data Analysis. *The Annals of Mathematical Statistics* 1962;33(1):1–67.
19. Thomas JW, Ward K. Economic profiling of physician specialists: use of outlier treatment and episode attribution rules. *Inquiry* 2006 Fall;43(3):271–82. [PubMed: 17176969]
20. Adams JL, McGlynn EA, Thomas JW, Mehrotra A. Incorporating statistical uncertainty in the use of physician cost profiles. *BMC health services research* 2010;10:57. [PubMed: 20205736]
21. Raudenbush, S.; Bryk, A. *Applications and Data Analysis Methods*. 2. Newbury Park, CA: Sage; 2002. Hierarchical Linear Models.
22. Krein SL, Hofer TP, Kerr Eve A, Hayward RA. Whom Should We Profile? Examining Diabetes Care Practice Variation among Primary Care Providers, Provider Groups, and Health Care Facilities. *HSR: Health Services Research* 2002;37(5):1159–80.
23. Medicare Payment Advisory Commission. Report to the Congress: Increasing the Value of Medicare. Washington, DC: Medicare Payment Advisory Commission; 2006. Using episode groupers to assess physician resource use; p. 1-27.

**Exhibit 1.****Distribution of Provider Cost Profiles***

Providers with cost scores above 1 have higher costs than expected while providers with cost profiles less than 1 have lower costs than expected. Box represent 25th and 75th percentile, tails represent 5th and 95th percentiles

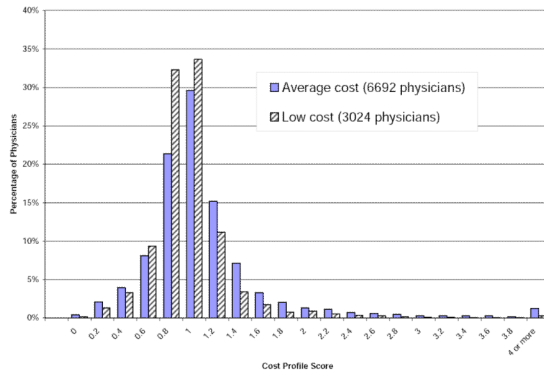


Exhibit 5.
Cost Profile Distribution of Individual Physicians Among Average Cost and Low Cost Physician Groups

Exhibit 2

Size and Composition of Massachusetts Groups

Groups	Number of MD					Total
	2-3 MD	4-9 MD	10-49 MD	50+ MD		
	n (% of all groups)					
PCP MDs Only	22 (12%)	16 (9%)	7 (4%)	0 (0%)	45	
PCP MDs >= 75%	0 (0%)	7 (4%)	16 (9%)	0 (0%)	23	
Specialty Only	9 (5%)	2 (1%)	1 (1%)	0 (0%)	12	
Multispecialty	6 (3%)	9 (5%)	43 (23%)	47 (25%)	105	
Total	37	34	67	47	185	

Exhibit 3

Sample size and reliability of individual physician and physician group profiles

	Number	Mean number of episodes assigned	Median Reliability (25 th and 75 th percentile)	
Individual MD	12,615	255	0.53	(0.21, 0.79)
All Groups	185	13,790	0.91	(0.78, 0.96)
Groups with 2-3 MD	37	1,102	0.76	(0.54, 0.87)
Groups with 4-9 MD	34	3,091	0.85	(0.71, 0.92)
Groups with 10-49 MD	67	8,658	0.93	(0.82, 0.96)
Groups with 50+ MD	47	38,835	0.97	(0.93, 0.98)

Exhibit 4

Relationship between Category of Group's Cost and Category of Physician's Costs in the Group

		MD Cost Category		Total
		Low Cost	Average Cost	
		n (% of MD within group cost category)		
Group Cost Category	Low Cost	487 (16%)	2537 (84%)	3024
	Average Cost	503 (8%)	6189 (92%)	6692