

Availability of Data for Measuring Physician Quality Performance

Sarah Hudson Scholle, MPH, DrPH; Joachim Roski, PhD, MPH;
Daniel L. Dunn, PhD; John L. Adams, PhD; Donna Pillitterre Dugan, MS;
L. Gregory Pawlson, MD, MPH; and Eve A. Kerr, MD, MPH

Measurement of physician quality performance is increasingly used by health plans as the basis for quality improvement, network design, and financial incentives.¹ Still, efforts to measure physician performance face a number of challenges, in particular the need for sufficient sample size to support reliable measurement and the lack of consensus on methods for attributing patient measures to clinicians.^{2,3}

Researchers have noted that measurement and comparison of physician quality can be hampered by sample size.⁴ A minimum threshold of 30 patients is a common guideline for supporting comparisons for an individual measure,⁵ and evidence suggests that at least 35 to 45 observations are needed to make valid comparisons.^{6,7} One challenge in obtaining sufficient sample size relates to the measure itself. Many quality measures describe a select group of patients and, by definition, will yield a small number of patients for any physician. Other measures apply to larger proportions of patients, but the ability to capture information on a physician's entire panel of patients is limited (as when performance measurement relies on data from a single health plan).

A related issue in quality measurement is attribution. Which physicians should be responsible for a quality measure? Given the current focus on team-based chronic disease care and the reality that most patients receive care from multiple clinicians,⁸ some authors argue that the most appropriate level of accountability is not the individual physician but rather a formal or informal group of physicians.⁹ Healthcare organizations often attribute patient quality measures based on utilization or a specific set of services, despite the challenges in identifying which physician should be held responsible for the fulfillment (or lack of fulfillment) of a quality measure.

Efforts are needed to understand how these issues may affect the meaningfulness and soundness of physician profiling efforts. In this study, we used a data set that is typical of the information used by health plans to characterize physician performance.

Using 27 well-accepted measures that can be obtained from administrative data, we evaluated (1) how many quality events were available per physician and (2) how different attribution rules affect the number of quality events.

Objective: To evaluate measurement of physician quality performance, which is increasingly used by health plans as the basis of quality improvement, network design, and financial incentives, despite concerns about data and methodological challenges.

Study Design: Evaluation of health plan administrative claims and enrollment data.

Methods: Using administrative data from 9 health plans, we analyzed results for 27 well-accepted quality measures and evaluated how many quality events (patients eligible for a measure) were available per primary care physician and how different approaches for attributing patients to physicians affect the number of quality events per physician.

Results: Fifty-seven percent of primary care physicians had at least 1 patient who was eligible for at least 1 of the selected quality measures. Most physicians had few quality events for any single measure. As an example, for a measure evaluating appropriate treatment for children with upper respiratory tract infections, physicians on average had 14 quality events when care was attributed to physicians if they saw the patient at least once in the measurement year. The mean number of quality events dropped to 9 when attribution required that the physician provide care in at least 50% of a patient's visits. Few physicians had more than 30 quality events for any given measure.

Conclusions: Available administrative data for a single health plan may provide insufficient information for benchmarking performance for individual physicians. Efforts are needed to develop consensus on assigning measure accountability and to expand information available for each physician, including accessing electronic clinical data, exploring composite measures of performance, and aggregating data across public and private health plans.

(*Am J Manag Care.* 2009;15(1):67-72)

In this issue

Take-away Points / p71

www.ajmc.com

Full text and PDF

Web exclusive

eAppendix Tables 1 to 3

For author information and disclosures, see end of text.

METHODS

Data Sources

Administrative claims and enrollment data from the Ingenix Impact Pro database¹⁰ for individuals enrolled in 9 health plans for 2003 and 2004 were available for this study. The Impact Pro database is built from deidentified health insurance claims and enrollment information contributed by different managed care organizations. Each of 9 plans selected for the study had at least 250,000 members and accounted for 15% to 50% of managed care enrollees in their markets. During the study period, 170,168 primary care physicians (PCPs) provided care to members of these plans. More details on the study methods are available elsewhere.¹¹

Selection of Measures and Attribution to Physicians

We focused on 27 measures describing acute, chronic, and preventive care activities performed by PCPs. Only measures that could be obtained through administrative claims data were included. **eAppendix Table 1** (available at www.ajmc.com) lists all quality measures used in this study, as well as the period used to attribute patients and quality events to physicians.

We identified physicians by the unique identifiers used by each health plan. Primary care physicians, including family physicians, general internists, and general pediatricians, were identified based on their specialty designated in health plan credentialing records.

In selecting an attribution approach, we considered the interactions between clinicians and patients in the course of delivering care, the kinds of services involved, the evidence of a physician's involvement in the patient's care, and the data sources available. For this study, we applied a measure-specific attribution logic based on administrative data. Measures were attributed to PCPs based on the outpatient visits they provided to patients during a prescribed time frame specific to each measure. Visits were defined using Healthcare Effectiveness Data and Information Set codes for preventive and ambulatory health services.⁵ To test a less stringent approach to attribution, a patient measure was attributed to a physician if the patient had 1 or more visits during the prescribed time frame. In addition to this "1-visit" rule, 2 more stringent rules were assessed: a PCP was attributed responsibility for a patient's measure (1) if the patient completed at least 30% of his or her ambulatory visits with that physician (30% rule) and (2) if the patient completed at least 50% of his or her ambulatory visits with that physician (50% rule).

A quality event occurred each time a patient was eligible for a quality measure. Therefore, a single patient could contribute multiple quality events if he or she was eligible

for multiple measures (eg, preventive screening and another measure).

Statistical Analysis

We computed summary information describing the number and proportion of physicians attributed with quality events for eligible patients for each attribution approach. We also examined the proportion of physicians with more than 30 quality events for each individual measure and the proportion of quality events accounted for by those physicians with more than 30 quality events. More detailed results are provided in **eAppendix Table 2** and **eAppendix Table 3** (available at www.ajmc.com). All analyses were conducted by staff at Ingenix and the National Committee for Quality Assurance using SAS version 9.0 (SAS Institute, Inc, Cary, NC).¹²

RESULTS

Overall, 57% of 170,168 PCPs represented in the study claims data could be attributed responsibility for at least 1 quality event (ie, ≥ 1 of their patients was eligible for ≥ 1 of our selected quality measures). **Table 1** summarizes findings based on the 1-visit rule and describes the percentage of PCPs with more than 30 quality events for a measure. Except for preventive measures, few PCPs had more than 30 observations for any given measure. However, these high-volume providers account for a larger share of quality events overall, particularly for preventive care measures. For example, only 17% of physicians had more than 30 quality events for colorectal cancer screening, but these physicians accounted for 78% of the quality events for this indicator. Only 1% of physicians had more than 30 quality events for annual glycosylated hemoglobin testing among patients with diabetes mellitus, but they accounted for 16% of the quality events for this measure.

Table 2 summarizes how moving from a less stringent rule to a more stringent rule for attribution affects the number of patients available for characterizing physician performance. For example, using the 1-visit rule for the measure assessing appropriate care for upper respiratory tract infections in children, physicians on average had 14 eligible patients for that measure in the measurement year. The mean number of quality events dropped to 11 when care was attributed using the 30% rule and to 9 when care was attributed using the 50% rule. Relative to the 1-visit rule, the 50% rule reduced by about half the number of quality events per physician for a measure. Adopting a more stringent rule for a measure also reduced the number of PCPs with at least 1 quality event for that measure (data not shown).

Availability of Data for Measuring Physician Quality Performance

Table 1. Primary Care Physicians (PCPs) Attributed More Than 30 Quality Events^a

Quality Measure	Total PCPs	PCPs With >30 Attributed Quality Events, %	Total Quality Events	Quality Events Accounted for by PCPs With >30 Attributed Quality Events, %
Preventive care				
Breast cancer screening	52,056	9	536,127	54
Colorectal cancer screening	68,063	17	1,233,428	78
Cervical cancer screening	76,856	18	1,606,255	80
Chlamydia screening in women	10,649	1	64,415	15
Glaucoma screening in older adults	28,104	6	250,800	51
Chronic care				
Use of appropriate medications for people with asthma	18,381	0	58,772	2
Antidepressant medication management				
Acute phase	13,419	0	38,947	3
Continuation phase	13,419	0	38,947	3
Follow-up after hospitalization for mental illness				
30 d	1775	0	2766	0
7 d	1775	0	2766	0
Follow-up care for children prescribed attention-deficit/hyperactivity disorder medication				
Initiation phase	4853	0	10,228	0
Continuation and maintenance phase	2871	0	4282	0
β-Blocker treatment after a heart attack				
At discharge	779	0	917	0
Persistence	3724	0	4592	0
Comprehensive diabetes care				
Glycosylated hemoglobin testing	20,601	1	114,179	16
Low-density lipoprotein cholesterol testing	11,814	4	100,925	25
Medical attention for nephropathy	30,567	2	183,207	16
Osteoporosis management in women who had a fracture	1627	0	2358	0
Annual monitoring for patients taking persistent medications				
Angiotensin-converting enzyme inhibitor	28,604	6	250,767	32
Anticonvulsant	9228	0	15,640	0
Digoxin	7184	0	12,103	1
Diuretic	26,720	4	196,937	22
Statin	21,344	7	192,423	37
Acute care				
Appropriate treatment for children				
With upper respiratory tract infections	20,615	13	295,041	68
With pharyngitis	13,400	8	130,775	54
Inappropriate antibiotic treatment for adults with acute bronchitis	18,289	1	81,787	12
Use of imaging studies for low back pain	23,498	1	103,464	8

^aQuality events were attributed to a physician using the 1-visit rule. If the physician had a preventive care or ambulatory visit with the health plan member anytime during the eligibility period for the measure, the quality event was attributed to that physician. See eAppendix Table 1 for quality measure specifications and attribution periods.

■ TRENDS FROM THE FIELD ■

■ **Table 2.** Quality Events Among Primary Care Physicians (PCPs) by Attribution Rule

Quality Measure	Mean No. of Quality Events		
	≥1 Preventive Care or Ambulatory Visit	≥30% of the Patient's Preventive Care or Ambulatory Visits	≥50% of the Patient's Preventive Care or Ambulatory Visits
Preventive care			
Breast cancer screening	11.2	7.9	5.9
Colorectal cancer screening	18.1	14.0	10.7
Cervical cancer screening	20.9	13.6	9.6
Chlamydia screening in women	6.9	4.7	3.7
Glaucoma screening in older adults	8.9	7.2	5.9
Chronic care			
Use of appropriate medications for people with asthma	3.2	2.3	2.0
Antidepressant medication management (2 measures)	2.9	2.0	1.8
Follow-up after hospitalization for mental illness (2 measures)	5.9	6.0	6.6
Follow-up care for children prescribed attention-deficit/hyperactivity disorder medication			
Initiation phase	2.1	1.7	1.6
Continuation and maintenance phase	1.5	1.3	1.2
β-Blocker treatment after a heart attack			
At discharge	5.3	5.0	5.3
Persistence	1.2	1.1	1.1
Comprehensive diabetes care (3 measures)			
Osteoporosis management in women who had a fracture	1.5	1.3	1.2
Annual monitoring for patients taking persistent medications			
Angiotensin-converting enzyme inhibitor	8.8	6.8	5.6
Anticonvulsant	1.7	1.4	1.3
Digoxin	1.7	1.5	1.4
Diuretic	7.4	5.6	4.7
Statin	9.0	6.9	5.6
Acute care			
Appropriate treatment for children			
With upper respiratory tract infections	14.3	10.5	8.7
With pharyngitis	13.7	10.4	8.7
Inappropriate antibiotic treatment for adults with acute bronchitis			
Use of imaging studies for low back pain	4.4	3.7	3.2
Total patient quality events			
Total quality events	56.9	41.5	31.4

DISCUSSION

Even evaluating the large health plans included in the study and using a less stringent approach to measure attribution, few

physicians had more than 30 quality events available to characterize their performance on key quality measures such as colorectal cancer screening or diabetes care. Thirty observations represent a common threshold for adequate denomina-

tor size in performance measurement; the number of observations needed to gain reliable measurement at the physician level may be higher (or lower) depending on the between-physician variation in performance on a given measure.^{6,7,11}

Still, for many measures, physicians with a high volume of quality events account for a significant percentage of all quality events observed.

Using more stringent and specific rules to assign patients to physicians (by requiring that a larger proportion of a patient's care was managed by that physician) further decreased the number of quality events attributable to any given physician.

Our findings illustrate the challenges of benchmarking individual physician performance using available administrative data from individual health plans. Pham and colleagues⁸ recently noted that care for patients covered by Medicare is frequently shared among multiple providers and concluded that this dispersion of patients could limit the effectiveness of pay-for-performance initiatives because of the lack of accountability on the part of individual physicians. Our data go further to show that, even if accountability is assumed, there is limited information available to characterize physician performance on actual quality measures for single private sector health plans.

Limitations

Limitations of this study included the number of measures studied, the reliance on administrative data only, and the lack of direct information about the physician's relationship with the patient. These findings are based on only 27 quality measures from administrative data. However, all are well tested and nationally endorsed, and most are included in health plans' and employers' physician performance measurement programs for PCPs. Administrative data were used because most physician-level measurement efforts around the country rely on these data. However, administrative data limit the type of clinical actions that can be profiled.¹³ These limitations demonstrate the issues that health plans often face in developing meaningful provider profiles. Finally, the study describes findings for PCPs only. Within the context of this study, we found similar challenges in achieving sample sizes that provided more than 30 quality events for specialist physicians.

Implications

As our results demonstrate, several practical steps are needed to ensure that physician profiles based on administra-

Take-away Points

Available administrative data for a typical health plan may provide insufficient information for benchmarking performance among individual physicians.

- For any single quality measure, most physicians had few quality events attributed to them even using a less stringent rule for attributing patient measures to physicians.
- To promote confidence in physician performance data, health plans should share with physicians information on the number of quality events measured and the rules for attributing responsibility for care.
- Efforts are needed to encourage the aggregation of databases across public and private health plans and to maximize the data available for characterizing physician performance.

tive data have sufficient information for reliable estimates. First, pooling administrative data within communities across all health plans, government purchasers, and other entities is critical to construct a more complete database representing most or all care rendered by that community's physicians. Regional and national quality initiatives promoted by the Centers for Medicaid & Medicare Services¹⁴ and by the Robert Wood Johnson Foundation¹⁵ are examples of such data pooling.

Second, composite measures should be considered. Care should be taken in selecting and weighting the individual measures for inclusion in a composite. Furthermore, the use of composites creates additional challenges in interpreting quality results and specific actions for improving care. However, composites constructed around a particular condition or patient care activity may provide insights into quality performance and increase the number of quality events available for comparing providers.

Third, efforts are needed to encourage physician practices, health plans, and other entities to make readily available more clinically detailed data for quality measurement construction. Initially, the more widespread availability of electronic data that is already present in some settings (eg, laboratory results) will allow for a larger number of potentially more meaningful quality measures to be constructed without the expense of medical record review, and efforts are critically needed to improve the capabilities of electronic medical records to report quality measures.¹⁶ Efforts to augment routinely available administrative data by including additional codes (eg, *Current Procedural Terminology II*) that capture critical information about outcomes or results of patient care may be promising but have not yet been tested for accuracy or reliability in widespread applications, to our knowledge.

CONCLUSIONS

Our results highlight the challenges that individual health plans face in using administrative data to measure physician performance. Creating large multipayer administrative data-

bases that capture a wide proportion of care in physician practices may aid in that effort. Consensus is needed on implementation methods, including the selection of endorsed measures, the method and level of attribution, and the minimum sample sizes needed to capture performance. Performance measurement holds significant promise for improving quality of care. Yet, performance measurement is not without potential for unintended consequences for physicians and patients.^{17,18} As the United States moves forward on efforts to offer financial rewards based on quality performance and to report performance data publicly,¹⁹ there is growing urgency in improving our ability to base these efforts on sound, meaningful, and actionable performance results.

Author Affiliations: National Committee for Quality Assurance (SHS, JR, DPD, LGP), Washington, DC; RAND Corporation (JLA), Santa Monica, CA; Ingenix (DLD), Waltham, MA; and Ann Arbor Veterans Affairs Medical Center and University of Michigan (EAK), Ann Arbor. Dr Roski is now with the Engelberg Center for Healthcare Reform at the Brookings Institution, Washington, DC.

Funding Source: This study was supported by a grant from the Commonwealth Fund and by grant 1 R13 HS016277 from the Agency for Healthcare Research and Quality. Dr Kerr was supported in part by grant DIB 98-001 from the Veterans Affairs Health Services Research and Development Quality Enhancement Research Initiative for Diabetes Mellitus and by Michigan Diabetes Research and Training Center Grant P60DK-20572 from the National Institute of Diabetes and Digestive and Kidney Diseases.

Previous Presentations: Portions of this work were presented at the 2006 Annual Research Meeting of AcademyHealth; June 26, 2006; Seattle, WA; and at an invitation-only conference convened by the National Committee for Quality Assurance entitled "Benchmarking Physician Performance: Current Practice and Research Needs"; January 11, 2006; Rockville, MD.

Author Disclosure: The authors (SHS, JLA, DPD, LGP) report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article. Dr Roski reports having served as a consultant for Children's Hospital and Clinic of Minneapolis, NCQA, and Impact Education, Inc, and has received grants from Robert Wood Johnson Foundation, Commonwealth Fund, and Agency for Healthcare Research and Quality. Dr Dunn is an employee of Ingenix, a company that sells quality measurement products and services. However, none of those products are mentioned in this study. Dr Kerr reports serving as an unpaid consultant to the National Committee for Quality Assurance.

Authorship Information: Concept and design (SHS, JR, DLD, JLA, DPD, LGP, EAK); acquisition of data (SHS, DLD, DPD); analysis and interpretation of data (SHS, JR, DLD, JLA, EAK); drafting of the manuscript (SHS, JR, DLD, JLA); critical revision of the manuscript for important intellectual content (SHS, JR, JLA, LGP, EAK); statistical analysis (DLD, JLA); obtaining funding (SHS, JR); administrative, technical, or logistic support (DPD, LGP); and supervision (JR, LGP).

Address correspondence to: Sarah Hudson Scholle, MPH, DrPH, National Committee for Quality Assurance, 1100 13th St NW, Ste 1000, Washington, DC 20005. E-mail: scholle@ncqa.org.

REFERENCES

- Galvin R, Milstein A. Large employers' new strategies in health care. *N Engl J Med*. 2002;347(12):939-942.
- Landon BE, Normand SL, Blumenthal D, Daley J. Physician clinical performance assessment: prospects and barriers. *JAMA*. 2003;290(9):1183-1189.
- Lee TH, Meyer GS, Brennan TA. A middle ground on public accountability. *N Engl J Med*. 2004;350(23):2409-2412.
- 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):2098-2105.
- National Committee for Quality Assurance. *HEDIS 2007 Technical Specifications for Physician Measurement*. Washington, DC: National Committee for Quality Assurance; 2007.
- Safran DG, Karp M, Coltin K, et al. Measuring patients' experiences with individual primary care physicians: results of a statewide demonstration project. *J Gen Intern Med*. 2007;21(1):13-21.
- Kaplan SH, Griffith JL, Price LL, Pawlson LG, Greenfield S. Measuring the physician effect on quality of care measures: improving the reliability of composite physician performance assessment scores. *Med Care*. In Press.
- Pham HH, Schrag D, O'Malley AS, Wu B, Bach PB. Care patterns in Medicare and their implications for pay for performance. *N Engl J Med*. 2007;356(11):1130-1139.
- Krein SL, Hofer TP, Kerr EA, Hayward RA. Whom should we profile? examining diabetes care practice variation among primary care providers, provider groups, and health care facilities. *Health Serv Res*. 2002;37(5):1159-1189.
- Ingenix.com Web site. Impact Pro. 2008. <http://www.ingenix.com/Products/Employers/HealthandProductivity/EvidenceBasedHealthEMP/IngenixImpactProPAYUA/>. Accessed October 31, 2008.
- Scholle SH, Roski J, Adams J, et al. Reliability of individual and composite measures for profiling physician performance. *Am J Manag Care*. 2008;14(12):829-838.
- SAS [computer program]. Version 9.0. Cary, NC: SAS Institute; 2004.
- Paulson LG, Scholle SH, Powers A. A comparison of administrative-only versus administrative plus chart review data. *Am J Manag Care*. 2007;13(10):553-558.
- AQA Web site. AQA Alliance. 2005. <http://www.aqaalliance.org>. Accessed October 29, 2008.
- RWJF Web site. The Robert Wood Johnson Foundation: health and health care improvement. <http://www.rwjf.org>. Accessed October 29, 2008.
- Kerr EA, Smith DM, Hogan MM, et al. Building a better quality measure: are some patients with 'poor quality' actually getting good care? *Med Care*. 2003;41(10):1173-1182.
- Casalino LP, Elster A, Eisenberg A, Lewis E, Montgomery J, Ramos D. Will pay-for-performance and quality reporting affect health care disparities? *Health Aff (Millwood)*. 2007;26(3):w405-w414.
- Werner RM, Asch DA, Polsky D. Racial profiling: the unintended consequences of coronary artery bypass graft report cards. *Circulation*. 2005;111(10):1257-1263.
- Epstein AM. Pay for performance at the tipping point. *N Engl J Med*. 2007;356(5):515-517. ■