Accounting for Variation in Technical Quality and Patient Satisfaction

The Contribution of Patient, Provider, Team, and Medical Center

Kelly L. Stolzmann, MS, Mark Meterko, PhD,* Michael Shwartz, PhD,*† Gary J. Young, JD, PhD,*‡ Erol A. Pekoz, PhD,† Justin K. Benzer, PhD,* Katerine Osatuke, PhD,§ Bert White, MBA, DMin,* and David C. Mohr, PhD**

Background: The delivery of healthcare depends on individual providers, coordination within teams, and the structure of the work setting. We analyzed the amount of variation in technical quality and patient satisfaction accounted for at the patient, provider, team, and medical center level.

Methods: Data abstracted from Veterans Health Administration patient medical records for 2007 were used to calculate measures of technical quality based on adherence to best practice guidelines in 5 domains. Outpatient satisfaction was obtained from a 2007 standardized national mail survey. Hierarchical linear models that accounted for the clustering of patients within providers, providers within teams, and teams within medical centers were used to partition the variation in technical quality and satisfaction across patients and components of the system (ie, providers, teams, and medical centers).

Results: Providers accounted for the largest percent of system-level variance for all technical quality domains, ranging from 46.5% to 71.9%. For the single-item measure of patient satisfaction, medical centers, teams, and providers accounted for about the same percent of system-level variance (31%–34%). For the doctor/patient interaction scale providers explained 59.9% of system-level variance, more than double that of teams and medical centers. For all the measures, the residual variance (composed of patient-level and random error) explained the largest proportion of the total variance.

Conclusions: Providers explained the greatest amount of system-level variation in technical quality and patient satisfaction. However, in both of these domains, differences between patients were the predominant source of nonrandom variance.

From the *VA Boston Healthcare System, Center for Organization, Leadership, and Management Research (COLMR); †Boston University School of Management; ‡Boston University School of Public Health; and §VA National Center for Organization Development (NCOD).

This research was supported by VHA HSR&D grant number: IIR 05–221.

Reprints: Kelly Stolzmann, MS, VA Boston Healthcare System, Mail Stop 152M, 150 South Huntington Avenue, Boston, MA 02130. E-mail: kelly.stolzmann@va.gov.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.lww-medicalcare.com).

ISSN: 0025-7079/10/4808-0676

Key Words: performance measurement, quality of care, patient satisfaction, quality improvement

(*Med Care* 2010;48: 676–682)

Clinical performance measures are increasingly being used to profile individual providers for the purpose of assessing quality of care.[1,2] One recent but rapidly expanding application of such profiling has been in the design of pay-for-performance (P4P) programs. In many P4P programs, financial incentives are used to reward individual providers, typically physicians, for achieving quality-related performance targets. However, evidence has been accumulating that calls into question the assumption that individual providers are the primary drivers of process and outcome variation[3–6] and studies on the impact of P4P programs have yielded mixed results.[7–13]

We identified limitations in the extant research on practice variation. First, research has focused on differences between geographic regions, hospitals, and individual providers with little or no attention to teams.[6,14–18] Several major conceptual developments including the application of total quality management/continuous quality improvement (QI) to healthcare and the evolution of the concept of clinical microsystems have led to an appreciation that the delivery of healthcare depends substantially on teamwork among different types of healthcare providers.[5,14] Given these developments and despite the difficulty of unambiguously defining a team in many situations, it is important to assess practice variation at the team level.

Second, most previous studies of practice variation have focused on provider variation with respect to technical quality and process outcomes.[4,5,16,18] However, patient ratings and reports of their healthcare experiences are commonly used as indicators of the quality of care and are an important dimension of overall healthcare delivery.[19,20]

Third, among existing studies on practice variation, there are substantial differences in both methods and findings.[4,5,21] In particular, many studies have been relatively limited in the size and nature of the patient groups studied or

have concentrated on patients in only one geographic location.[22] Furthermore, different analytic techniques have been used across studies, as well as different definitions of the denominator in computations of the percent of variance accounted for. Finally, these studies have been conducted in a wide variety of settings.

This study is designed to address these shortcomings. We sought to determine the amount of variation in both technical quality and patient satisfaction that is accounted for at the patient, provider, primary care team, and medical center level in a large national sample of patients who received care through the Veterans Health Administration (VA). This study is intended to help managers and clinicians identify the levels at which there is the most potential to reduce variations through targeted interventions.

METHODS

Overall Setting and Design

This study involved the secondary analysis on both technical quality and patient satisfaction data in the VA. The VA is the largest integrated healthcare system in the United States and operates 765 ambulatory care facilities and 153 acute-care inpatient medical centers, providing care to 5.58 million unique patients annually.[23]

The identification of primary care teams was a key element of the study design. We drew on the definition by Cohen and Bailey[24] of a team as an interdependent collection of individuals with shared responsibility for certain outcomes, and who are seen by themselves and others as an intact social entity, embedded within one or more larger social systems. Accordingly, we defined a primary care team as the group of providers responsible for the care of a specified set of patients. In the VA, primary care teams may be hospital based or work in a free-standing community-based outpatient clinic. At larger facilities, hospital-based primary care providers may be divided into multiple teams (eg, "Red Team" and "Blue Team").

We created 2 patient level analytic datasets by merging the technical quality of care data and the patient satisfaction survey data each with a file containing a provider identification number (ID) for the relevant episode of care (eg, the visit that triggered the patient satisfaction survey). The VA Primary Care Management Module database includes provider ID numbers (which are unique to each provider) and primary care team assignments. Thus, we were able to link outcomes with providers, teams, and medical centers.

Variables and Data Sources

Technical Quality

To measure the technical quality of care, we selected 12 key performance indicators for which VA senior managers are held accountable including: (1) measures of intermediate patient outcomes presumed to reflect adherence to disease management guidelines for 3 chronic conditions—diabetes, acute myocardial infarction (AMI), and hypertension; and (2) measures of adherence to preventive medicine recommendations in 2 domains—cancer screening and immunizations (Table 1). We created up to 5 disease-specific adherence measures for each patient by dividing the number of preventive medicine recommendations fulfilled and/or intermediate outcome levels achieved by the total number for which the patient was eligible.

TABLE 1. Technical Quality of Care Measures

Disease Category	Type	Description
Diabetes	Outcome	LDL-C <120
Diabetes	Outcome	BP ≤140/90
Diabetes	Outcome	BP ≥160/100 or not done*
Hypertension	Outcome	BP ≥160/100 or not recorded*
Hypertension	Outcome	BP ≤140/90
Cancer	Process	Breast screen (HEDIS) all ages
Cancer	Process	Cervical screen (HEDIS)
Cancer	Process	Colorectal screen (HEDIS)
Immunizations	Process	Influenza
Immunizations	Process	Pneumococcal
Acute myocardial infarction	Outcome	LDL-C <100 most recent and full lipid panel in past 2 yr
Acute myocardial infarction	Outcome	LDL-C ≥120*

*Measures were reverse coded in the calculation of adherence rates.
LDL-C indicates low-density lipoprotein cholesterol; BP, blood pressure; HEDIS, Healthcare Effectiveness Data and Information Set.

We obtained patient-level data on adherence from the VA fiscal year 2007 External Peer Review Program administered by the VA Office of Quality and Performance. The charts of selected patients with target diagnoses are reviewed by trained staff using detailed protocols to check whether specific evidence-based standards of care have been applied. These measures are then widely distributed and used as performance measures across VA.[25] Of the 42,836 records in the 2007 External Peer Review Program dataset, there were 3713 patients associated with 1327 providers whose IDs could not be matched to a primary care team, and an additional 1819 patients were not eligible for any of the 12 key performance indicators. Thus, our final technical quality analytic dataset contained measures from 37,304 patient records cared for by 4715 primary care providers within 1274 teams at 128 medical centers. The average number of patients per provider varied by disease category with the lowest number for AMI (2.6 patients per provider) and the highest for cancer screening (6.0 patients per provider). Although there were too few patients per provider to profile providers, there were a sufficient number to examine the relative magnitude of variances at different levels of the hierarchy.

Patient Satisfaction

We measured patient satisfaction using patient-level data from the VA fiscal year 2007 Survey of Health Experiences of Patients (SHEP), also administered by Office of Quality and Performance. The satisfaction module of the ambulatory care SHEP is based on the Picker Institute surveys of patient-centered care, and standard scoring algorithms involve 38 multiple-response items that yield multi-item scales assessing patient perceptions on 8 dimensions (Table 2).[26] The outpatient SHEP also includes a

TABLE 2. Outpatient Satisfaction Indices and Component Scales

	Number of Items	Sample Item
Delivery system index		
Access	7	Were you able to get this clinic appointment as soon as you wanted?
Continuity	1	Is there one provider or team in charge of your VA care?
Visit coordination	5	Did someone tell you how you would find out about the results of your tests?
Overall coordination	6	Were the providers who cared for you familiar with your most recent medical history?
Doctor/patient interaction index		
Preferences	5	Did the provider listen to what you had to say?
Emotional support	4	Did you have confidence and trust in the provider you saw?
Education/information	7	When you asked questions, did you get answers you could understand?
Courtesy	2	Overall, how would you rate the courtesy of your provider?

single-item assessment of overall quality ("Overall, how would you rate the quality of care you received during the past 2 months?") that uses a 5-point "poor" to "excellent" Likert response scale, with a higher score indicating a more favorable assessment.

For this study, we used 3 measures of patient satisfaction. Based in part on the distinction between technical and interpersonal elements of quality proposed by Donabedian,[27] we computed 2 composite indices by taking the simple average of each patient's scores on the relevant SHEP scales (Table 2). As an index of satisfaction with the delivery system, we averaged the patient's scores on the access, continuity of care, visit coordination, and overall coordination (across visits over time) scales. As an index of satisfaction with the doctor/patient interaction, we averaged the patient's scores on the emotional support, patient education, attention to patient preferences, and courtesy scales. We also used the single-item overall quality score as a separate measure of patient satisfaction. The single-item measure has the advantage of allowing patients to implicitly weight various aspects of their experience according to their own evaluation of its importance when coming to an overall judgment of the quality of care they received.[28,29] Of the 90,795 SHEP respondents with an outpatient clinic stop code, 6802 had provider IDs that could not be matched to a primary care team. The final dataset consisted of 83,993 patients of 4786 providers within 1318 teams at 129 medical centers. Patient satisfaction scores were available for an average of 17.1 patients per provider.

There is literature suggesting that age and health status are the demographic factors with the greatest and most consistent influence on patient satisfaction scores.[30,31] Thus, we used veterans age in years, gender, and functional status as measured by the physical composite scale and mental composite scale scores from the SF-12 as individual-level risk-adjustors.[32] Adjusted models for the technical quality measures included patient age and gender.

Statistical Analysis

We used hierarchical linear models to account for the clustering of patients within providers, providers nested within teams, and teams nested within medical centers. Empty models with random intercepts and no explanatory variables were used to partition the variability for each technical quality and patient satisfaction measure by level (ie, provider, team, and medical center). The estimates from the variance-covariance matrix were used to calculate the intraclass correlation for each level in the hierarchy, which indicates the amount of variance in the dependant variable that is attributable to each level in the model.[33] We calculated the fraction of explained variance (excluding the patient level and random variation) in technical quality and patient satisfaction attributable to providers, teams, and medical centers. To test the robustness of these findings, the proportions of attributable variance were recomputed after eliminating all providers with <3 patients in the database. The multilevel analyses were performed using the PROC MIXED procedure in SAS Version 9.1 (SAS Institute Inc., Cary, NC).

Although PROC MIXED assumes a normal distribution for the dependent variable, Bayesian models allow for more flexibility in specifying the underlying distribution of the data. To check the robustness of the findings obtained using hierarchical linear models in SAS, we also ran Bayesian hierarchical models using WinBUGS (see Appendix, Supplemental Digital Content 1, online only available at: http://links.lww.com/MLR/A84).[34] In addition, models from PROC MIXED provide the residual variation for each model, which includes both patient level and random variation. The WinBUGs models estimate patient-level variation directly. We used estimates of patient-level variation from WinBUGS to divide the PROC MIXED residual variation into a patient level and a random component (see Appendix Supplemental Digital Content 1, online only available at: http://links.lww.com/MLR/A84). We also used WinBUGs to estimate our confidence in results when undertaking pairwise comparisons of variances at different levels in the hierarchy.

RESULTS

The technical quality scores were generally high but varied by disease category with the lowest rate for AMI at 76% and the highest for hypertension at 87%. Patient satisfaction was also quite high in this study population. The mean overall satisfaction score was 4.2 on a 5-point Likert scale. Descriptively, scores on the doctor/patient interaction composite index were slightly higher than those

TABLE 3. Summary Statistics for Technical Quality and Patient Satisfaction Measures by Level of Care

	Patient			Provider		Team		Medical Center	
	N	Mean	SD	p25th	p75th	p25th	p75th	p25th	p75th
Technical quality measures*									
Diabetes	9698	0.86	0.23	0.78	1.00	0.79	0.93	0.83	0.89
Hypertension	21583	0.87	0.27	0.79	1.00	0.81	0.95	0.85	0.90
Cancer	26249	0.80	0.39	0.67	1.00	0.73	0.92	0.78	0.85
Immunizations	25489	0.84	0.34	0.75	1.00	0.78	0.94	0.80	0.86
Acute myocardial infarction	6089	0.76	0.38	0.50	1.00	0.62	1.00	0.71	0.80
Satisfaction measures									
Single-item overall quality†	77117	4.19	0.87	4.00	4.44	4.00	4.32	4.08	4.25
Delivery system index‡	82872	82.64	20.03	76.17	89.28	77.83	86.83	79.29	84.97
Doctor/patient interaction index‡	83150	85.92	19.58	79.84	92.64	81.28	89.53	82.87	87.75

*Disease-specific ratios (0.0–1.0).
†5-point Likert scale; higher scores indicate a more favorable assessment.
‡0–100 point satisfaction scale; higher scores indicate a more favorable assessment.

TABLE 4. Technical Quality and Patient Satisfaction: Percent of Explainable Variance Across Levels of the Delivery System

	Percentage of Total Variance Explained at the System Level*	Percentage (%) of System-Level Variance Explained By		
		Provider	Team	Medical Center
Technical quality measures				
Diabetes	3.6	61.1	21.4	17.5
Hypertension	4.7	71.9	16.6	11.4
Cancer	4.2	46.5	24.3	29.2
Immunizations	3.4	55.3	33.0	11.7
Acute myocardial infarction	4.9	50.0	41.9	8.1
Satisfaction measures				
Single-item overall quality	4.6	34.3	34.4	31.3
Delivery system index	7.2	40.6	27.7	31.7
Doctor/patient interaction index	8.6	59.9	19.4	20.7

*The column entitled "Percentage of Total Variance Explained at the System Level" reflects the proportion of total variance, including patient level and random error, explained by the provider, team, and medical center levels.

on the delivery system composite index (86% and 83%, respectively) (Table 3).

Of the total variation in the technical quality data, at most 4.9% was explained by the combination of provider, team, and medical center levels, which together comprise the system level. Table 4 reports the percentage of total variation and of system-level variation accounted for at each level. In all disease categories, the provider level explained the greatest amount of system-level variation, ranging from 46.5% (Cancer) to 71.9% (Hypertension). The team and medical center levels each explained a moderate percentage (11%–33%), with the exception of AMI, where 41.9% of variation was explained at the team level and 8.1% at the medical center level.

We found similar amounts of explained system-level variance at the provider, team, and medical center levels for the single-item measure of overall patient satisfaction (31%–34%) (Table 4). The percent of variance explained at the provider, team, and medical center levels was also fairly uniform for the delivery system satisfaction index (28%–41%). Regarding satisfaction with the doctor/patient interaction, however, the provider level explained 59.9% of the system-level variation, more than double the amount of variance explained at the team and medical center levels (Table 4). Eliminating all providers with <3 patients did not affect the amount of variance explained at the provider, team, or medical center levels for the technical quality or patient satisfaction measures (data not shown).

Models adjusting for patient age and gender did not affect the amount of variation accounted for among the technical quality measures (results not shown). The inclusion of age, gender, and SF-12 scores explained ~8% of the total variation in satisfaction with the doctor/patient interaction (data not shown). Patient-level covariates also explained about 8% of the variation in the delivery system satisfaction index. However, adjusting the patient satisfaction scores had minimal impact on the variation at the system level.

For the technical quality and patient satisfaction measures, the residual estimate (composed of patient level and random error) explained a large proportion of the total variance ($>90\%$ for all measures; Table 5). We used the results from WinBUGS to estimate the percent of variance explained at the patient level separate from random error (see Appendix, Supplemental Digital Content 1, online only available at: http://links.lww.com/MLR/A84). For patient satisfaction, the amount of variance explained at the patient level varied widely by measure, with the lowest for the doctor/patient interaction index (4.1%) and the highest for the single-item

TABLE 5. Technical Quality and Patient Satisfaction: Percent Variance Explained at the Patient Level

	Estimated Patient Level (%)	Estimated Random Error (%)	Total Residual (%)*
Technical quality measures			
Diabetes	23.6	72.8	96.4
Hypertension	30.0	65.3	95.3
Cancer	28.4	67.4	95.8
Immunizations	25.9	70.8	96.6
Acute myocardial infarction	64.3	30.9	95.1
Satisfaction measures			
Single-item overall quality	78.6	16.8	95.4
Delivery system index	52.2	40.7	92.8
Doctor/patient interaction index	4.1	87.3	91.4

*The column entitled "Total Residual" is the percent of total variance explained by the residual, which is composed of patient-level and random variation. Estimated patient-level variation is extrapolated from the WinBUGS results, as explained in the Appendix.

measure of patient satisfaction (78.6%). For technical quality, the patient level explained between 23.6% (Diabetes) and 64.3% (AMI) of the total variance (Table 5). As shown in Appendix Table 4 (Supplemental Digital Content 1, online only available at: http://links.lww.com/MLR/A84), for the technical quality measures, we are certain that patient-level variation is greater than system-level variation; more likely than not, it is at least 50% greater.

DISCUSSION

We found a small amount of total variation explained at the provider, team, and medical center levels for both technical quality and patient satisfaction. The provider level explained the largest proportion of system-level variance among all technical quality measures. Regarding patient satisfaction, the percent variance accounted for was relatively small at the provider, team, and medical center level for both the single-item measure and the delivery system index, whereas for the doctor/patient interaction satisfaction index, the provider level explained more than double the amount of variance than the team and medical center.

Multilevel studies examining variation in technical quality measures at the provider, team, and facility levels have yielded mixed results.[21] Reports of total variance attributable to the physician level ranged from 1%–2%[6] to 10%[4]; at the team level, estimates of 4% or less have been typical.[3,4] In a study of adherence to diabetes care guidelines in one VA regional healthcare system, Krein et al[4] found that the percent of total variation accounted for at the physician level varied from a moderate amount (8%–10%) on some process measures (eg, obtaining laboratory tests) to a modest amount (2%–7%) on measures of resource utilization (eg, cost of hypoglycemic medications) to virtually none (<2%) on measures of intermediate outcomes (eg, percent of patients with last HbA1c value ≥9.5%). At the team level, the percent of variation accounted for across the same set of measures did not exceed 3%; however, at the next higher level of aggregation, the percent of variation accounted for at the facility level again varied from virtually none to what Krein et al described as moderate amounts (8%–10%) and even higher on some intermediate outcomes and resource use variables (12%–18%). In another study that focused exclusively on hemodialysis resource use, facility-level variation was substantially larger than physician-level variation, both before and after case-mix adjustment.[5] O'Connor et al[35] found that physicians and clinics accounted for a small amount of variance in hemoglobin values among patients with diabetes; over 95% of the variance was attributable to the patient level. The results of this study are generally consistent with these previous studies with regard to the percent of variance accounted for at the individual physician level.

However, our results differed from other recent multilevel studies with regard to the amount of variation accounted for at levels of aggregation above the physician, particularly at the facility level: our estimates were generally lower. One possible reason for this may be that differences between facilities, and teams within facilities, on these closely monitored technical quality measures have decreased over time in response to previous QI interventions at the team level and incentives at the medical center level.[36] Another factor may be differences in the definition of team, which could in turn affect the variance left over to be accounted for at higher levels.

Although previous studies of variation in medical care have generally focused on clinical processes and outcomes, some have examined aspects of patient satisfaction. Again, results have varied depending on the number of levels examined and the definition of variance accounted for (eg, whether total variance or only explainable variance was used as the denominator). However, in general physicians have been found to account for a noteworthy amount of variation in patient satisfaction when compared with other levels.[37–39] In this study, we also found that the provider level accounted for the greatest amount of variation in patient satisfaction. This finding was most pronounced with regard to the doctor/patient interaction index, which included patient evaluations of provider emotional support and courtesy. However, it was also true that providers accounted for the largest proportion of variance in the delivery system index, which one might expect to be more strongly associated with the team and/or medical center levels. This may reflect the subdivision of the ambulatory care domains into the 2 summary indices. One component of the delivery system index—coordination of care—included questions that refer to the providers who have cared for the patient during the past 2 months. Another component—the continuity measure—also explicitly references the "…provider…in charge of your care." Patients responses to such questions may well be driven largely by their experiences with their providers.

In contrast to the inconsistent results from previous studies regarding the percent of variance in technical quality and patient satisfaction accounted for at different levels of aggregation, it is worth noting that the greatest proportion of observed variation has been consistently observed at the patient level.[3–5,21,22,35,37–39] Similarly, in this study, the ma-

jority of the variation in both technical quality and patient satisfaction was due to the unmeasured patient factors or chance. These results suggest that eliminating systematic variation at the provider, team, and medical center levels would have a relatively modest effect on variation in technical quality and patient satisfaction.

Providers were the primary source of system-level variation in both quality and satisfaction, which offers support for assessing quality at the provider level. However, it is important to note that it is not clear what level of the system should be targeted to most effectively improve scores and reduce variation. For example, Selby et al[39] have shown that facility-level QI efforts resulted in significant improvements in outcomes, even though the facility level accounted for a smaller proportion of variance than individual providers. Improvement efforts aimed at levels higher than the provider may also elicit changes in provider behavior.[5,21] More longitudinal studies are needed to better understand the relative impact of QI efforts targeted at different levels of the system. Health plans and provider groups may also want to investigate the efficacy of directing incentives to patients to modify their own behavior. We also interpret these results as supportive of the increased patient participation in medical care decision-making and customization of clinical encounters that are hallmarks of patient-centered care.[40]

This study has a number of strengths. We used a large, national sample of patients and examined both technical quality and patient satisfaction measures. Technical quality measures were based on multiple indicators for both process and outcome measures spanning multiple disease categories; patient satisfaction measures included evaluations of both features of the delivery system and of the doctor/patient interaction. We conducted the analyses using 2 statistical packages (SAS and WinBUGS) representing different analytic models (maximum likelihood and simulation, respectively) to confirm the results by comparison (see Appendix, Supplemental Digital Content 1, online only available at: http://links.lww.com/MLR/A84).

Several limitations to this study should also be noted. The study was conducted within the VA, and thus, it is possible that some of the results may not generalize to other healthcare delivery contexts with a less uniform medical informatics infrastructure or less evolved and integrated performance measurement system. However, the similarity of our results to those in much of the literature suggests this is not a major problem. Second, levels of performance on both the technical quality measures and patient satisfaction were relatively high and uniform, which restricted the amount of variability in the data. However, although many of our measures had a very high average, similar results in the percent of variance explained were found across all measures, even those with lower means (eg, AMI; mean = 0.76). Finally, as noted, each provider had only a small number of eligible patients for the measures. Hence, we could not reliably profile providers. However, our purpose in this article was not to profile but to analyze variances. We have shown that in pairwise comparisons, we can determine with a high level of confidence which level of the hierarchy explains the most variance in technical quality and patient satisfaction scores.

In summary, this study used multilevel modeling to estimate the proportion of variation explained at the patient, provider, team, and medical center level among technical quality indicators and patient satisfaction measures in a national sample of VA patients. We found differences in the levels of variation explained across disease-specific technical quality measures and patient satisfaction indices, which may suggest areas of focus for improving care. The results of this study are applicable to policymakers and managers interested in investing resources in QI and patient satisfaction.

REFERENCES

1. Kassirer JP. The use and abuse of practice profiles. *N Engl J Med*. 1994;330:634–636.
2. Wennberg DE. Variation in the delivery of health care: the stakes are high. *Ann Intern Med*. 1998;128:866–868.
3. Hofer TP, Hayward RA, Greenfield S, et al. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *JAMA*. 1999;281:2098–2105.
4. Krein SL, Hofer TP, Kerr EA, et al. Whom should we profile? Examining diabetes care practice variation among primary care providers, provider groups, and health care facilities. *Health Serv Res*. 2002;37:1159–1180.
5. Turenne MN, Hirth RA, Pan Q, et al. Using knowledge of multiple levels of variation in care to target performance incentives to providers. *Med Care*. 2008;46:120–126.
6. Hayward RA, Manning WG Jr, McMahon LF Jr, et al. Do attending or resident physician practice styles account for variations in hospital resource use? *Med Care*. 1994;32:788–794.
7. Rosenthal MB, Frank RG. What is the empirical basis for paying for quality in health care? *Med Care Res Rev*. 2006;63:135–157.
8. Pearson SD, Schneider EC, Kleinman KP, et al. The impact of pay-for-performance on health care quality in Massachusetts, 2001–2003. *Health Aff (Millwood)*. 2008;27:1167–1176.
9. Mehrotra A, Damberg CL, Sorbero ME, et al. Pay for performance in the hospital setting: what is the state of the evidence? *Am J Med Qual*. 2009;24:19–28.
10. Herrin J, Nicewander D, Ballard DJ. The effect of health care system administrator pay-for-performance on quality of care. *Jt Comm J Qual Patient Saf*. 2008;34:646–654.
11. Damberg CL, Raube K, Teleki SS, et al. Taking stock of pay-for-performance: a candid assessment from the front lines. *Health Aff (Millwood)*. 2009;28:517–525.
12. Conrad DA, Perry L. Quality-based financial incentives in health care: can we improve quality by paying for it? *Annu Rev Public Health*. 2009;30:357–371.
13. Petersen LA, Woodard LD, Urech T, et al. Does pay-for-performance improve the quality of health care? *Ann Intern Med*. 2006;145:265–272.
14. Casparie AF. The ambiguous relationship between practice variation and appropriateness of care: an agenda for further research. *Health Policy*. 1996;35:247–265.
15. Ashton CM, Petersen NJ, Souchek J, et al. Geographic variations in utilization rates in Veterans Affairs hospitals and clinics. *N Engl J Med*. 1999;340:32–39.
16. Fisher ES, Wennberg JE, Stukel TA, et al. Hospital readmission rates for cohorts of Medicare beneficiaries in Boston and New Haven. *N Engl J Med*. 1994;331:989–995.
17. Bensing JM, van den Brink-Muinen A, de Bakker DH. Gender differences in practice style: a Dutch study of general practitioners. *Med Care*. 1993;31:219–229.
18. Greenfield S, Nelson EC, Zubkoff M, et al. Variations in resource utilization among medical specialties and systems of care. Results from the medical outcomes study. *JAMA*. 1992;267:1624–1630.
19. Eisen SV, Shaul JA, Clarridge B, et al. Development of a consumer survey for behavioral health services. *Psychiatr Serv*. 1999;50:793–798.
20. Shaul JA, Eisen SV, Stringfellow VL, et al. Use of consumer ratings for

quality improvement in behavioral health insurance plans. *Jt Comm J Qual Improv*. 2001;27:216–229.
21. Fung V, Schmittdiel JA, Fireman B, et al. Meaningful variation in performance: a systematic literature review. *Med Care*. 2010;48:140–148.
22. Rodriguez HP, Scoggins JF, von Glahn T, et al. Attributing sources of variation in patients' experiences of ambulatory care. *Med Care*. 2009; 47:835–841.
23. *VA Stats at a Glance*. Washington, D.C.: Government Printing Office, National Center for Veterans Analysis and Statistics; 2009.
24. Cohen SG, Bailey DE. What makes teams work: group effectiveness research from the shop floor to the executive suite. *J Manag*. 1997;23: 238–290.
25. Department of Veterans Affairs. Office of Quality and Performance. Performance Measurement Overview. http://vaww.archive.oqp.med.va.gov/oqp_services/performance_measurement/pm.asp. Accessed on September 18, 2009.
26. Cleary PD, Edgman-Levitan S, Roberts M, et al. Patients evaluate their hospital care: a national survey. *Health Aff (Millwood)*. 1991;10:254–267.
27. Donabedian A. The quality of care: how can it be assessed? *JAMA*. 1988;260:1743–1748.
28. Vavra TG. *Improving Your Measurement of Customer Satisfaction: A Guide to Creating, Conducting, Analyzing, and Reporting Customer Satisfaction Measurement Programs*. Milwaukee, WI: ASQ Quality Press; 1997.
29. Nagy MS. Using a single-item approach to measure facet job satisfaction. *J Occup Organ Psychol*. 2002;75:77–86.
30. Carlson MJ, Shaul JA, Eisen SV, et al. The influence of patient characteristics on ratings of managed behavioral health care. *J Behav Health Serv Res*. 2002;29:481–489.
31. Young GJ, Meterko M, Desai K. Patient satisfaction with hospital care: effects of demographic and institutional characteristics. *Med Care*. 2000;38:325–334.
32. Ware JE, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34:220–233.
33. Singer J. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *J Educ Behav Stat*. 1998;24: 323–355.
34. Lunn DJ, Thomas A, Best N, et al. WinBUGS – a Bayesian modeling framework: concepts, structure, and extensibility. *Stat Comput*. 2000; 10:325–337.
35. O'Connor PJ, Rush WA, Davidson G, et al. Variation in quality of diabetes care at the levels of patient, physician, and clinic. *Prev Chronic Dis*. 2008;5:1–9.
36. The Congress of the United States, Congressional Budget Office. *Quality Initiatives Undertaken by the Veterans Health Administration*. Pub No. 3234 August 2009.
37. Sixma HJ, Spreeuwenberg PM, van der Pasch MA. Patient satisfaction with the general practitioner: a two-level analysis. *Med Care*. 1998;36: 212–229.
38. Hekkert KD, Cihangir S, Kleefstra SM, et al. Patient satisfaction revisited: a multilevel approach. *Soc Sci Med*. 2009;69:68–75.
39. Selby JS, Schmittdiel JA, Lee J, et al. Meaningful variation in performance: what does variation in quality tell us about improving quality? *Med Care*. 2010;48:133–139.
40. Allshouse KD. Treating patients as individuals. In: Gerteis M, Edgman-Levitan S, Daley J, Delbanco T, eds. *Through the Patients' Eyes: Understanding and Promoting Patient-Centered Care*. San Francisco, CA: Josey-Bass; 1993:19–44.

