ORIGINAL RESEARCH

Measuring Performance in Primary Care: What Patient Outcome Indicators Do Physicians Value?

Paul L. Dassow, MD, MSPH

Background: **Determining which patient outcome indicators may be appropriate to include in a primary care, practice performance tool is a difficult task. Unfortunately, no published studies currently document the opinions of these physicians regarding which indicators they most value.**

Objective: **To ascertain the level of agreement among primary care physicians regarding the most salient patient outcome indicators for measuring performance in primary care.**

Methods: **A random population survey of 115 adult primary care physicians. The survey consisted of a comprehensive list of health characteristics included in 8 validated instruments (eg, SF-36, Sickness Impact Profile), as well as 10 common clinical indicators and 6 health behaviors. Each item was ranked using a 5-point Likert scale regarding its value for inclusion in a performance measure.**

Results: **Analysis of 93 returned surveys (RR 81%) indicated strong agreement (≥75%) that 19 health characteristics were important or very important. These characteristics fit into 8 domains: physical functioning, psychological functioning, social functioning, pain, quality of life, physiologic symptoms, health behaviors, and clinical indicators. Notably absent were measures of social support and health perceptions.**

Conclusions: **Strong agreement exists among practicing primary care physicians regarding the most valued patient outcome indicators. Development of practice performance measures should be influenced by such data. (J Am Board Fam Med 2007;20:1–8.)**

Measuring the performance of medical care continues to be an important yet challenging part of clinical research, quality improvement, and more recently, as a possible mechanism for providing monetary incentives to practices. Various strategies for accomplishing this task have included examining process measures (eg, mammography rates), collecting objective patient data (eg, blood pressures), and obtaining subjective patient indicators (eg, patient satisfaction). Organizations such as the National Committee for Quality Assurance (NCQA) have issued well utilized sets of process indicators (ie, HEDIS) designed to assess the quality of care.[1] The Medical Outcomes Study, published in 1989, brought subjective patient reported indicators into the mainstream of outcomes assessment.[2] Numerous individual researchers have contributed to the effort by designing and testing new assessments focusing on patient function and satisfaction. All this activity has resulted in the creation of more than 30 global measures of health (or quality of life), in addition to more than 300 disease-specific measures.[3] This vast array of tools presents a significant challenge for those interested in choosing the best set of indicators to assess performance in primary care.

Recently, 5 prominent organizations joined together to form the Ambulatory Care Quality Alliance (AQA), a collaborative group whose stated goals are to agree on a strategy for measuring performance at the physician level, facilitate the collection of performance data, and report such data to consumers, physicians, and other stakeholders.[4] The AQA tasked a Performance Measurement Workgroup to arrive at a reasonable set of performance indicators that could be used in an ambulatory care setting. Representation on the Perfor-

This article was externally peer reviewed.
Submitted 17 May 2006; revised 24 October 2006; accepted 30 October 2006.
From the Department of Family and Community Medicine, University of Kentucky, Lexington, KY.
Funding: This study was supported in part by a Title VII Faculty Development Grant from the Health Resources and Services Administration.
Conflict of interest: none declared.
Corresponding author: Paul Dassow, MD, MSPH, Department of Family and Community Medicine, University of Kentucky, K-307, Kentucky Clinic, Lexington KY 40536 (E-mail: pdass1@email.uky.edu).

mance Measurement Workgroup included a broad range of stakeholders, including physician, hospital and health care professional groups, private sector employers and business coalitions, public purchasers and government agencies, health insurance plans, and accrediting organizations.[5] In early 2005, the Workgroup issued a starter set of 26 disease-specific performance indicators. These indicators are heavily weighted toward assessing process of care, with only 3 of 26 directly measuring a patient outcome.[6]

Inclusion of patient outcome indicators in any performance assessment tool remains somewhat controversial. A proposed resolution at the 2006 Annual American Academy of Family Physicians (AAFP) conference asks the Academy to reject the inclusion of any such outcome indicators tied to reimbursement, in lieu of process indicators.[7] Supporting arguments include problems with risk adjusting patient panels and holding physicians accountable for patient behaviors. Yet, measuring process rather than outcomes seems less preferable if indeed positive patient outcomes are the desired product. Is there a set of outcome indicators that most physicians would agree are valuable to measure? There is currently no published data documenting what assortment of outcome indicators primary care physicians would select to assess their performance if they were given the opportunity. The purpose of this study was to explore the opinions of practicing clinicians regarding the value of various global outcome indicators for performance measurement.

Methods

Sample Selection

The study was designed as a cross-sectional survey of primary care physicians (family medicine and internal medicine). Inclusion criteria included Board certification, licensure in the state of Kentucky, involvement in direct patient care for at least 5 half-days per week, and completion of an accredited residency at least 1 year before this study. The survey instrument was hand delivered by the investigator to each physician in the sample to foster understanding, encourage return, and assure uniformity of participant instruction. After a short explanation of the study's purpose, the physicians were asked to consider the following question: "What health characteristics would you want included in a health measurement tool designed to assess the performance of your practice across your entire adult patient population?" (See Appendix A.)

The target population for this study included all family physicians and internists practicing in a 7-county region of central Kentucky. One of these counties (Fayette, population 270,000) is urban and contains a medical school and 2 universities; the 6 other adjoining counties are more rural (average population 32,000). Eligible physicians were identified through the latest published database of the Kentucky Board of Medical Licensure. Sample-size calculations indicated that a random sample of 92 physicians was needed to ensure an accuracy of $\pm 7\%$ in reported data at the 95% CI.

Instrument Creation

To facilitate the physicians' consideration of the many aspects, or characteristics, of health that could appear in a measurement instrument, a systematic analysis was conducted of the content of 8 generic, validated, outcomes tools: the Medical Outcomes Study's Long Form 149 (LF-149) and Short Form 36 (SF-36), the Dartmouth COOP Charts, the Functional Status Questionnaire, the Quality of Well-Being Scale, the Duke-UNC Health Profile, the Sickness Impact Profile, and the Nottingham Health Profile.[8–15] It should be noted that although each of these has been used as a generic health assessment, the Quality of Well-Being Scale was originally designed as a utility assessment tool. These tools were chosen for analysis given their prominence in recent health outcomes literature and after discussions with published health services researchers. The content of each of these tools was classified using a taxonomy similar to that used by the Medical Outcomes Study Group (Table 1).[10] In this system, as in the World Health Organization definition, health is considered to exist in 3 dimensions: physical, emotional, and social.[16] These 3 dimensions can be divided into several domains; for example, domains in the physical dimension would include such aspects as symptoms, functioning, and clinical indicators. Each domain can be further divided into sets of related health characteristics, which in turn can be assessed with single or multiple survey items.

After deleting redundancies in these health assessments and ensuring that all content was adequately represented, 32 separate health characteristics were identified. Added to this set of 32 health

Table 1. Sample of Content Analysis for Generic Health Assessments

	DCC*	FSQ	SF-36	SIP	QWB
Physical					
Functioning					
Sexual	X	X			
ADL	X	X	X	X	X
Mobility	X	X			
IADL	X	X	X	X	X
High level	X	X	X		
Pain					
Level	X	X	X	X	
Limitations	X	X			
Cognition					
Memory	X	X			
Language	X	X			
Symptoms					
Senses, CNS	X				
Elimination	X	X			
Lung, skin	X				
Eating/swallow	X	X			
Indicators†					
BP, BMI					
Emotional					
Functioning					
Distress	X	X	X	X	X
Well-being	X	X			
Self-esteem					
Sexual satisfaction	X				
Health perceptions					
Present		X	X	X	X
Future	X				
Recent change	X	X	X		
Behaviors‡					
Tob/Alc					
Vitality	X	X	X		
Quality of life	X				
Social					
Functioning					
Work/role	X	X	X	X	X
Pleasure	X	X	X		
Interactions					
Frequency	X				
Diversity	X				
Support	X				

* DCC, Dartmouth COOP charts; FSQ, functional status questionnaire; SF-36, MOS short form 36; SIP, sickness impact profile; QWB, quality of well-being s.cale
† From National Health and Nutrition Examination Survey (NHANES).
‡ From National Health Interview Survey (NHIS).

characteristics were 10 common clinical indicators taken from the National Health and Nutrition Examination Survey (NHANES III). The 6 health-related behaviors included in the National Health Interview Survey (NHIS) were also added given the prominent role preventive health counseling now holds in the primary care environment. This combination of patient reported health characteristics,

Table 2. Characteristics of Responders vs. Non-responders

	Responders (n = 93)	Non-responders (n = 22)	*P* Value
Gender (% male)	79	82	.73
Specialty (% FP)	45	36	.45
County (% Fayette)	46	55	.48

FP, family practice.

laboratory data, and health-related behaviors assured that participating physicians had the full range of possible patient outcomes to consider.

The survey instrument for this study asked physicians to rate each of these 48 health characteristics on a 5-point Likert scale ranging from unimportant (1) to very important (5). The survey also contained demographic items and an open-ended item that elicited any additional health characteristics that the physicians believed to be important. The instrument was piloted among 6 primary care physicians, with minor changes made in wording to facilitate understanding.

Results

Personal contact was attempted with 140 physicians selected by random sampling of the target population. Of these, 23 could not be reached because they had either retired or relocated. Two physicians were ineligible for the study because they had accepted medical administration positions and were no longer in clinical practice. Of the remaining 115 physicians, 93 returned surveys for a response rate of 81%.

Table 2 compares the responding and non-responding physicians in terms of demographics. χ^2 statistics showed no significant differences between the 2 groups in gender, specialty, or county of practice. The gender mix was also not significantly different from the overall state statistics of 80.4% male.

A summary of the responses is presented in Table 3. Each individual item was ranked by the percentage of physicians indicating that the characteristic was either very important or important for inclusion in a practice performance assessment. The percentage of physicians ranking the item as somewhat important, as well as those noting it to be not very important or unimportant is also reported. Nineteen health characteristics (the first grouping in Table 3) were rated as either very important or important by at least 75% of the physicians. These characteristics, which fall within 8 domains of health, consist of 2 levels of physical functioning, 2 measures of physical pain, 4 types of physical symptoms (other than pain), 3 clinical indicators, a measure of emotional distress, 5 health behaviors, a general quality-of-life item, and an item regarding role functioning. Only 11 respondents answered the open-ended item, each suggesting the inclusion of other unique health-related behaviors or clinical indicators (eg, seat belt use, bone density).

Discussion

Practice performance data, in terms of patient outcomes, could be used for a variety of reasons including quality improvement, tracking systems changes, identifying best practices, and as a means of pay for performance. Including practicing physicians in the indicator determination process for all these uses is important, especially for uses tied to physician reimbursement. The initial failure of the pay for performance initiative created by United-Health Group highlights some of the difficulties in tying reimbursement to indicators determined by third party payers.[17] Prominent among the reasons cited for this failure was the lack of input received by practicing physicians. Indeed, numerous analysts warn that any set of indicators that has not been created with significant clinician input will invariably fail.[18,19]

The data presented in this study represents the first to document the opinions of primary care physicians regarding the relative value of various patient outcome indicators for performance assessment. The physicians responding to the survey question displayed a surprising degree of agreement regarding which indicators they felt were most important. Those items felt to be important or very important by 75% or more of the physicians in this study were considered to show a high level of agreement and were placed in the first grouping in Table 3. Those items achieving a simple majority of physicians who deemed the items important or

Table 3. Percent of Physicians Rating Each Health Characteristic Regarding Its Importance for Inclusion in a Performance Measurement Tool

Health Characteristic	Unimportant or Not Very Important 1 + 2	Somewhat Important 3	Important or Very Important 4 + 5
Blood Pressure	2.2	6.6	91.2
Alcohol/drug use	2.2	9.0	88.8
Physical functioning—ADL	5.5	6.7	87.7
Tobacco use	2.2	11.2	86.5
Psychological distress	3.4	12.4	84.2
Breathing difficulties	3.4	12.6	83.9
Role functioning	5.6	11.2	83.1
Bowel or bladder difficulties	3.4	13.6	82.9
Pain—level	3.4	14.6	82.0
Pain—limitations	3.4	14.6	82.0
Dietary habits	5.7	12.5	81.8
Body mass index	6.7	14.4	78.9
Eating/swallowing difficulties	3.4	18.2	78.4
Seizures/syncope	9.1	12.5	78.4
Usual physical activity	3.4	18.2	78.4
Physical functioning—IADL	11.1	11.1	77.8
Quality of life (general)	8.0	14.8	77.3
HDL/LDL cholesterol	6.7	16.7	76.7
Vaccination status	3.4	20.5	76.1
Cognitive functioning	6.7	19.1	74.1
Dysfunction of the senses	6.9	21.8	71.3
Psychological well-being	6.7	22.5	70.8
Mobility	3.4	25.8	70.8
Health perception—change	14.6	19.1	66.3
Sleep characteristics	4.5	31.5	64.0
Fasting blood sugar	9.0	27.0	64.0
Health perception—present	9.0	28.1	62.9
Vitality	14.6	22.5	62.9
Total cholesterol	14.4	23.3	62.2
Communication difficulties	14.9	23.0	62.0
Sexual activity	9.1	30.7	60.2
Social functioning	15.7	27.0	57.3
Sexual functioning	10.1	33.7	56.2
Health perceptions—future	18.0	28.1	53.9
Spiritual health	18.0	29.2	52.8
Social support	18.2	31.8	50.0
Sexual satisfaction	16.9	34.8	48.3
Hemoglobin A1C	25.8	25.8	48.3
Hemoglobin/hematocrit	18.9	37.8	43.3
Skin symptoms	25.0	36.4	38.7
Recreational limitations	13.5	48.3	38.2
Creatinine	24.4	37.8	37.8
Thyroid-stimulating hormone	17.8	45.6	36.7
Self-esteem	29.5	35.2	35.3
Electrolytes	32.2	35.6	32.2
Home management problems	27.3	43.2	43.2
Physical functioning high	21.3	49.4	29.2
Quality of social interactions	28.1	44.9	26.9

IADL, instrumental activities of daily living; HDL/LDL, high-density lipoprotein/low-density lipoprotein.

very important, but less than 75%, were placed in group 2. These cut points for determining the level of physician agreement were determined before data collection and were based on the premise that a simple majority, although noteworthy, would not carry the weight nor be as compelling as items that achieved agreement among three-quarters or more of the physicians surveyed. The initial pilot data indicated that these 3 categories would each contain at least some of the rated health characteristics.

Examination of those characteristics deemed important or very important by at least 75% of the respondents is somewhat surprising. Of the 19 characteristics in this grouping, 5 involved health-related behaviors. These behaviors are typically not included in global patient outcome assessments because many regard them as representing a risk to health rather than health itself. But it is clear that physicians recognize the salience of these behaviors to a person's overall health status and to assessing their own work. Operationalizing these behaviors for an assessment tool has been done in the NHIS and others. Adapting these items for a practice performance tool is certainly plausible.

Physicians are often criticized for being more concerned with lab and other objective data than with a patient's own interpretation of their health status. It is reassuring to see that only 3 of the items in the 1st grouping involve such data. Two of these, blood pressure and body mass index, are routinely measured at outpatient appointments. The third, high-density lipoprotein/low-density lipoprotein levels, is now a part of routine health screening. Obtaining these data from practices should be relatively simple. Inclusion of such data in a practice performance tool should not limit its ease of use; in fact, many patients possess an accurate knowledge of these parameters and could report these, along with subjective data, on a mailed questionnaire. In addition, as electronic medical records become more prevalent, pooling this data with a patient's self-assessment will become even less problematic.

Conspicuously absent from the 1st grouping of characteristics were social indicators of health, such as social support and social interaction. Most modern quality of life measures, including the oft-used SF-36, adhere to the World Health Organization's definition of health, and thus include a variety of such social health items. Post-survey physician feedback to the author indicated that most physicians felt social health to be outside the purview of routine primary care. Also absent were measures of patients' general health perceptions (eg, present and future), and this despite a body of research showing a high degree of correlation with future health-related events.[20] Exclusion of these latter characteristics may reflect physicians' opinions about their inability to substantially alter these patient perceptions.

The study was limited by a number of factors. First, it was conducted in a relatively small geographical area. Concerns over homogeneity of the sample led to questioning the physicians about location of medical and residency training. Responses indicated that 40% of the sample trained outside of Kentucky, thus ensuring at least some degree of diversity. The second limitation involved judgments made during the process of item classification. Many of the health characteristics involving symptoms were placed in the "pain" category rather than creating a separate organ-related symptom. An example of this was indigestion. Although a separate category of "stomach symptoms" could have been included in the survey tool, the judgment was made that these symptoms nearly always involve some kind of discomfort and thus rightly belong in the "pain" item. Other researchers could interpret these symptoms differently, which would alter a comprehensive listing of potential patient outcomes. Finally, the sample size was such that items clustering around the cut points of 50% and 75% could easily fall into or out of the groupings subject to their measurement error. To decrease the 7% uncertainty in the individual measurements would entail gathering responses from a larger group.

Other important considerations regarding data interpretation remain. First, the physicians involved in this study were not instructed to consider a specific use for the performance data. Rather, they were asked to judge the value of indicators for assessing their practice's clinical performance in general. If asked to consider a specific use, such as pay for performance, respondents may have altered their responses, although the resulting sets of indicators would probably be related. Thus, the data identified in this study could serve as a starting point for a collaborative discussion when considering any specific use. Second, process indicators were not included in the survey tool because this study was particularly interested in opinions about outcome indicators. A few physicians included var-

ious screening tests in the open-ended section of the survey, indicating recognition of their importance in assessing primary care practice. This study in no way indicates what mix of process and outcome indicators is most appropriate. Further research would be needed to understand practicing physicians' preferences in this regard.

The data reported in this study do represent important information regarding practicing clinicians' judgments about patient outcome indicators that could be used to assess performance. Physicians continue to be wary of pay for performance programs and are especially sensitive to measurement tools designed by third party payers.[21] Gathering physician preference data and using it to lead discussions on the most meaningful performance metrics can only serve to enhance the success of such programs. The performance of primary care physicians seems especially likely to be measured by such programs in the coming years. It is appropriate that this group be given an opportunity to voice its understanding of what performance indicators are most salient.

Appendix A

Study Script

"Good afternoon. My name is Paul Dassow and I am currently working as a family physician at the University of Kentucky. I have become increasingly interested in the issue of measuring performance in primary care. Many organizations, including insurance companies, quality assurance agencies, and governmental groups are considering measuring patient outcomes as an indicator of practice performance. As a practicing clinician myself, I feel it is very important for clinicians to be involved in deciding what patient outcomes, if any, might be important to include in a measurement tool meant to assess practice performance."

"To facilitate consideration of this question, I have included 48 health characteristics that were taken from validated health assessment tools and placed them in a survey format. For each health characteristic, please indicate how valuable you feel inclusion of that item would be in a tool meant to assess practice performance across your adult patient population. You may decide that none are valuable or that all of them are valuable—please consider each item on its own merits. You may add additional items that you feel would be valuable at the end of the survey. In addition, consider that the items would be assessed on all your adult patients not just on those with certain diagnoses."

"Return of this survey will be considered as consent to participate in this research study. Thank you for considering being a part of this study."

References

1. HEDIS Benchmarks [homepage on the internet]. Washington (DC): National Committee for Quality Assurance [cited 10 May 2006]. Available from: http://www.ncqa.org.
2. Tarlov AR, Ware JE Jr., Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical Outcomes Study. An application of methods for monitoring the results of medical care. JAMA 1989;262:925–30.
3. McHorney CA. Generic health measurement: past accomplishments and a measurement paradigm for the 21st century. Ann Intern Med 1997;127:743–50.
4. AQA [homepage on the internet]. The Ambulatory Care Quality Alliance; ©2005 [cited 2006 May 10]. Available from http://www.ambulatoryqualityalliance.org.
5. AQA Workgroup Membership. The Ambulatory Care Quality Alliance; c2005 [cited 10 May 2006]. Available from: http://www.ambulatoryqualityalliance.org/files/AQAWorkgroupMembershipLists.doc.
6. AQA releases ambulatory quality measures. Med Health 2005;59:3–4.
7. Resolution No. 302 (Kentucky A) [cited 20 Sept 2006]. Available from: http://www.aafp.org.
8. Kaplan RM, Anderson JP. The General Health Policy Model: An integrated approach. In: Spilker B, editor. Quality of life and pharmacoeconomics in clinical trials. New York: Raven Press; 1996. p. 309.
9. Ware JE, Sherbourne CD. The MOS 36-item short form health survey. Conceptual framework and item selection. Med Care 1988;30:473–83.
10. Stewart AL, Ware JE. Measuring functioning and well-being: the medical outcomes study approach. Durham (NC): Duke University Press; 1992.
11. Nelson E, Wasson J, Kirk J, et al. Assessment of function in routine clinical practice: Description of the COOP chart method and preliminary findings. J Chron Dis 1987;40:55S–63S.
12. Parkerson GR, Gehlbach SH, Wagner EH, James SA, Clapp NE, Muhlbaier LH. The Duke-UNC Health Profile: An adult health status instrument for primary care. Med Care 1981;19:806–28.
13. Hunt SM, McEwen J. The development of a subjective health indicator. Soc Health Illness 1980;2:232–46.
14. Bergner M, Bobbitt RA, Kressel S, Pollard WE, Gilson BJ, Morris JR. The Sickness Impact Profile: conceptual formulation and methodology for the de-

velopment of a health status measure. Int J Health Serv 1976;6:393–415.
15. Jette AM, Davies AR, Cleary PD, et al. The Functional Status Questionnaire: reliability and validity when used in primary care. J Gen Intern Med 1986; 1:143–9.
16. World Health Organization. Constitution of the World Health Organization. New York: World Health Organization; 1947.
17. Weber DO. The dark side of P4P. Physician Exec 2005;31:20–5.
18. Williams TR. Practical design and implementation considerations in pay-for-performance programs. Am J Man Care 2006;12:77–80.
19. Mirsky RS. Physician buy-in is essential for pay for performance. Physician Exec 2005;31:16–9.
20. Idler EL, Kasl S. Health perceptions and survival: do global evaluations of health status really predict mortality? J Gerontol 1991;46:S55–S65.
21. Steiger B. Poll finds physicians very wary of pay-for-performance programs. Physician Exec 2005;31:6–11.