Measuring Resident Physicians' Performance of Preventive Care

Comparing Chart Review with Patient Survey

Katri P. Palonen, MD, MSPH,^{1,2} Jeroan J. Allison, MD, MS,^{1,2,3} Gustavo R. Heudebert, MD,² Lisa L. Willett, MD,² Catarina I. Kiefe, MD, PhD,^{1,3,4} Terry C. Wall, MD, MPH,⁵ Thomas K. Houston, MD, MPH^{1,2,3}

¹The HSR&D Research Enhancement Award Program, Birmingham VA Medical Center, Birmingham, AL, USA; ²Division of General Internal Medicine, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA; ³Division of Preventive Medicine, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA; ⁴VA National Quality Scholars' Program, Office of Academic Affiliations, Veterans Health Administration; ⁵Division of General Pediatrics, Department of Pediatrics, University of Alabama at Birmingham, Birmingham, AL, USA.

BACKGROUND: The Accreditation Council for Graduate Medical Education has suggested various methods for evaluation of practice-based learning and improvement competency, but data on implementation of these methods are limited.

OBJECTIVE: To compare medical record review and patient surveys on evaluating physician performance in preventive services in an outpatient resident clinic.

DESIGN: Within an ongoing quality improvement project, we collected baseline performance data on preventive services provided for patients at the University of Alabama at Birmingham (UAB) Internal Medicine Residents' ambulatory clinic.

PARTICIPANTS: Seventy internal medicine and medicine-pediatrics residents from the UAB Internal Medicine Residency program.

MEASUREMENTS: Resident- and clinic-level comparisons of aggregated patient survey and chart documentation rates of (1) screening for smoking status, (2) advising smokers to quit, (3) cholesterol screening, (4) mammography screening, and (5) pneumonia vaccination.

RESULTS: Six hundred and fifty-nine patient surveys and 761 charts were abstracted. At the clinic level, rates for screening of smoking status, recommending mammogram, and for cholesterol screening were similar (difference <5%) between the 2 methods. Higher rates for pneumonia vaccination (76% vs 67%) and advice to quit smoking (66% vs 52%) were seen on medical record review versus patient surveys. However, within-resident ($N=70$) comparison of 2 methods of estimating screening rates contained significant variability. The cost of medical record review was substantially higher (\$107 vs \$17/physician).

CONCLUSIONS: Medical record review and patient surveys provided similar rates for selected preventive health measures at the clinic level, with the exception of pneumonia vaccination and advising to quit smoking. A large variation among individual resident providers was noted.

KEY WORDS: education; medical; preventive health services; patient survey; medical record review; cost evaluation.

DOI: 10.1111/j.1525-1497.2006.00338.x

J GEN INTERN MED 2006; 21:226-230.

Internal medicine residency programs are now required to evaluate performance in 6 competencies, including in "practice-based learning and improvement." The Accreditation Council for Graduate Medical Education new competency requirements define "practice-based" as investigation and evaluation of residents' own patient care, appraisal and

assimilation of scientific evidence, and improvements in patient care.¹ Accreditation Council for Graduate Medical Education has suggested various methods for evaluation of practice-based learning and improvement, but data on implementation of these methods are limited.

One widely implemented method for practicing physicians uses feedback of performance audits and peer-based comparison to promote improvement. Medical record review for performance audit and feedback has been explored as a tool to evaluate residents in outpatient clinics.² However, medical record review can underestimate the services offered,³⁻⁹ and it is expensive and logistically challenging.^{10,11} A feasible, inexpensive alternative that provides accurate data would be valuable to many residency programs.

Physician and patient surveys have been proposed as an alternative to medical record review for obtaining performance scores. However, studies have shown that physician surveys incorrectly estimate screening rates for preventive health services.¹²⁻¹⁴ Conversely, the Health Plan Employer Data and Information Set (HEDIS) has adopted patient surveys to evaluate physician performance for some quality measures, including questions on screening for tobacco and pneumonia vaccination.¹⁵ Current data suggest that patient surveys may be reliable in certain settings, such as in evaluation of health counseling and cancer screening.^{4,12,16} Previous comparison studies have not been conducted within the context of residency education.

Within the context of an ongoing quality improvement curriculum, we collected baseline resident performance data using both medical records and patient survey methods. In this paper, we compare the 2 methods as measures of performance at the resident and clinic levels. We hypothesized that patient surveys might provide a reliable representation of performance at a significantly lower cost, thus providing a sustainable tool for evaluation of resident performance in preventive health services and counseling.

METHODS

Design

We collected baseline performance data on preventive health services provided by 70 internal medicine and medicine-pediatrics residents for patients seen at their outpatient continuity clinic. The University of Alabama at Birmingham

Address correspondence and requests for reprints to Dr. Palonen: 1530 3rd Avenue South, FOT 720, Birmingham, AL 35294-1495 (e-mail: palonen@uab.edu).

Institutional Review Board approved collection of anonymized patient data, linked and aggregated at the clinic and resident levels. Because of anonymous surveys, we were not able to link data at the patient level. The study was a medical education curricular-based research study, and waiver of written informed consent at the resident level was also approved.

Setting

All residents staff the Russell outpatient clinic one half-day per week. The clinic is located in downtown Birmingham and serves predominantly a lower income, lower education patient population, with a mean age of 43 years, and 71% were African American. Of the clinic patients, 36% had Medicaid.

Participants

We included all residents who were postgraduate year (PGY)-1 or -2 in the academic year 2002 to 2003, and 4 PGY-3 medicine-pediatrics residents.

Data Collection

Data were collected by 2 methods: patient survey and medical records review. Medical records contain dictated clinic notes, which are placed in paper format in patient's chart. Charts of all patients seen in clinic for primary care visit at least twice between June 2002 and September 2003 by participating residents were abstracted for 5 preventive health measures (Table 1). Measures were chosen based on the HEDIS guidelines¹⁵ and United States Preventive Services Task Force recommendations.¹⁷ Medical records were abstracted by blinded research assistants who were trained during a 2-week session using a random set of charts not included in the final study. Review of medical records was based on a standardized protocol using a database created by the customized electronic MEDQUEST tool (Fu and Associates, Arlington, VA)³ in the clinic. Data abstracted included patient demographics, and documentation of (1) screening for smoking status, (2) advising smokers to quit, (3) cholesterol screening, (4) mammography screening, and (5) pneumonia vaccination. A positive preventive measure was documented if either (a) physician documentation of the test was present in the clinic note, or (b) results of the test were available in the chart. Double data abstraction

was conducted for quality control and revealed an error rate for the primary abstractor of less than 2%.

All patients with primary care visits to participating residents from July 2003 to November 2003 were given exit survey cards. Because of the low literacy level of our patient population, the literacy level of the survey was evaluated using the Gunning Fog index¹⁸ and Flesch reading ease formula¹⁹ and revised to achieve Fog grade level 5.4 and Flesch reading ease score 83.2. The survey included (1) patient demographics, (2) physician name, (3) self-assessment of current smoking status, (4) patient's report of physician advice related to the preventive health measures, and (5) patient's self-report of previous preventive health measures. For example, patients were first asked "During your doctor's visit, did your doctor tell you to get a mammogram?" and later "Have you had a mammogram test for breast cancer in the past 2 years?" We used information on previously developed preventive health measures to determine whether a patient was due for any of the preventive health services. For example, if a patient reported not having had a mammogram within the last 2 years, she was included in patients who should be recommended for a mammogram during the visit.

Data Analyses

Because survey data were anonymous, we were not able to directly link medical record and survey to statistically assess agreement at the patient level. At the clinic level, we calculated the overall proportion and 95% confidence interval of eligible patients who received each preventive service, comparing the proportions as measured by charts and surveys.

Data were then examined by resident. Mean differences in resident performance as measured using surveys and charts were assessed. Scatterplots with correlation coefficients were created to further depict variations in agreement at the resident level. In addition, we collected data on the cost of the 2 methods. Statistical analysis was performed using STATA SE 8.

RESULTS

Resident Characteristics

Of the 70 residents in our study, 66% were male, 44% PGY-1, 50% PGY-2, and 6% PGY-3. The majority were in the categorical medicine track (63%) versus primary care track (20%) or

Table 1. Clinic-Level Performance Rates on 5 Preventive Health Services, as Assessed by Chart Abstraction and Patient Survey Samples

Variable	Survey		Chart		Percent Difference (%)
	Number Received Service/ Total N*	Percent (95% Confidence Interval) (%)	Number Received Service/ Total N*	Percent (95% Confidence Interval) (%)	
Number of patients asked about smoking status (out of all patients)	392/659	59.5 (55.6 to 63.3)	416/761	54.7 (51.1 to 58.2)	4.8
Number of patients counseled to quit tobacco products (out of all reported to be smokers)	124/237	52.3 (45.8 to 58.8)	167/255	65.5 (59.3 to 71.3)	-13.2
Pneumonia vaccine recommended (out of all patients >65y who had not received the vaccine)	76/113	67.3 (57.8 to 75.8)	114/150	76.0 (68.4 to 82.6)	-8.7
Mammogram recommended (out of all women ≥ 50 not screened for last 2 y)	167/182	91.8 (86.8 to 95.3)	217/245	88.6 (83.9 to 92.3)	3.2
Cholesterol screening recommended (out of men ≥ 35, women ≥ 45 who had not been screened in past 5 y)	371/451	82.3 (78.4 to 85.7)	635/737	86.2 (83.5 to 88.6)	-3.9

*Denominator determined by the number of ideal candidates surveyed or abstracted as explained after each variable.

Table 2. Mean Difference Between Percentages of Preventive Service Recommendations for Eligible Patients as Assessed by Surveys Versus Charts at the Resident Level

	N*	Mean Difference (%) [†]	Inter-Quartile Range 25 to 75th Percentile (%)
Patients asked about smoking status (out of all patients)	70	1.3	– 17 to 21
Patients counseled to quit tobacco products (out of all reported to be smokers)	60	– 17.7	– 50 to 6
Pneumonia vaccine recommended (out of all patients > 65y who had not received the vaccine)	47	– 12.7	– 50 to 17
Mammogram recommended (out of all women ≥ 50 not screened for the last 2y)	60	1.6	0 to 17
Cholesterol screening recommended (out of men ≥ 35, women ≥ 45 who had not been screened in the past 5y)	68	– 2.7	– 19 to 14

*Number of residents includes only those with both medical record and survey data.

[†]Difference of mean performance at the resident level when comparing the proportion of patients in survey versus medical record review samples.

medicine-pediatrics (17%). Four participants (6%) were international medical graduates.

Patient Characteristics

A total of 678 out of 810 patients returned usable surveys (response rate = 83%), and medical records of 761 patients were reviewed. Because the survey was anonymous, we could not ascertain the overlap between the patients who were surveyed and those whose charts were reviewed. Surveyed patients had a mean age 6 years younger than those sampled by charts (56 vs 50 years, $P < .001$). They were also more likely to be female (63% vs 55%, $P = .002$).

Clinic-Level Agreement

Overall, the screening rates for preventive health services evaluated by medical records varied between 54.7% and 88.6% (Table 1). Patient reported rates varied from 52.3% to 91.8%. The rates ascertained by the 2 methods were similar (difference < 5%) for screening of smoking status, recommending a

mammogram, and for cholesterol screening. Pneumonia vaccination (difference = – 8.7%) and advice to quit smoking (– 13.2%) had somewhat greater disagreement between methods.

Resident Provider-Level Agreement

We abstracted a mean (SD) of 10.9 (3.9) and a range of 2 to 22 patient charts per resident. At the resident level, differences in screening rates as estimated by chart review and patient survey varied widely (Fig. 1, Table 2). Interquartile ranges noted variations up to 67%. Stratification by gender, PGY level, and residency track did not affect the results significantly.

Cost

Medical records review costs including programming the customized MedQuest tool, training research assistants, and time for chart abstraction. The total cost for medical records review data was \$7,510, approximately \$107 per resident physician. The largest cost was research assistant time.

FIGURE 1. Proportion of patients screened as ascertained by patient survey versus medical record review for preventive health services at a teaching ambulatory clinic. Each data point represents 1 resident, with the resident's screening rate by patient survey on the horizontal axis and the same resident's screening rate by medical record review on the vertical axis.

Patient survey costs included printing postcards, provision of free pens to patients, incentives to office staff for distributing surveys, and cost of data entry. The total cost for patient survey data was \$1,193, or approximately \$17 per resident physician. The largest cost for patient surveys was data entry. The difference between medical records review and patient survey costs was \$6,317 total and \$90 per resident.

DISCUSSION

We found that patient surveys and medical records review provided a fairly equivalent depiction of the overall clinic-level performance of selected preventive health measures. However, the performance assessment, as measured by survey versus medical records, varied widely for many individual residents. Thus, patient surveys were not adequate to measure the performance of individual residents or directly compare residents to their peers, if we consider the chart abstraction the standard. However, neither method may provide a stable estimate at the individual physician level, partly because of the low patient panel size.²⁰ However, the surveys were clearly less costly. When considering which method might be sustainable in the context of a residency program, program directors must weigh several issues, including the cost and logistics, accuracy of the performance audit method, and the purpose for which the audit will be used.

In addition to the considerable cost involved in implementing the chart audit, we encountered several challenging logistic issues. Identifying performance measures, programming the computerized data entry tool (MedQuest), and training the research assistants required more time than expected. Within-provider precision was limited by the number of patients seen twice or more during the calendar year (mean charts per resident = 11). If all charts were abstracted, time and cost would have increased considerably and may not have been feasible in the required time period.

There were also logistic challenges with the patient surveys. Patient surveys had to be assessed for the appropriate reading level. This limits the performance measures that can be feasibly assessed with surveys. In addition, we spent additional time with our front office staff encouraging them to distribute the patient surveys and, as discussed above, provided a \$100 incentive to each staff member for their effort.

What about the accuracy and reliability of either method? Agreement between methods of performance evaluation of physicians in practice has varied. A study by Montano found an association between patient survey and medical records for cancer screening by primary care physicians in practice,¹² although some other studies found the association to be dependent on the health measure in question.²¹ In many studies, patient report is substantially higher than the recorded preventive services in the medical record.^{5,9,22-25} Kell et al.²⁶ found that the rate of mammogram screening was lower on medical records than in claims data, pointing out a limitation in the medical record data found in other studies as well.^{24,25,27,28} Thus, neither charts nor surveys are necessarily a gold standard. Some measures, such as cholesterol screening, may not be directly discussed with patient, and are therefore less likely to be reported in the survey.⁴ Alternatively, some measures, including tobacco use screening, may be addressed during the patient visit, but not adequately documented by physician.^{4,5,9,25,28}

A limitation of our study is that data are limited to 1 residency program and 1 clinic where each resident sees patients only one half-day per week. Patients served in this clinic are, in general, of low socioeconomic status and often with multiple medical problems, making preventive health services somewhat more difficult to address. Patient surveys were adjusted to a lower reading level, which may have affected the accuracy of the questions when compared with the medical records. Also, patient surveys were anonymous, and we were not able to assess agreement at the patient level. At the provider and clinic levels we have assumed that the 2 methods represent repeated measures of aggregate performance. However, differences may have resulted from the different patient populations. In addition, we chose only a limited number of preventive health measures based on relevance to residency education, presumed variability among residents, and possibility for change. Our costs were largely determined by the type of chart abstraction tool, and having medical records available only in paper format. Variability in costs would have likely existed if these methods had been used in a different clinical setting.

To our knowledge, this is the first study comparing medical records with patient surveys in a residency population. Previous studies have concentrated on medical records as a tool for quality improvement projects, and have in general had fewer charts abstracted.^{2,11,29} Our study is unique in measuring a broad range of preventive health measures, with a large number of patient charts and patient surveys. We had well-defined quality indicators and standardized computer-assisted abstraction methodology, which was based on previously published methods.¹⁰

Based on our data, individual residents' performance may not be accurately reflected by either patient surveys or medical records review. Residents may perceive one source to be more valid over another when receiving feedback and this will not only impact the quality of the feedback given to improve preventive health services, but also the potential improvements in performance in response to the feedback. Research assessing residents' acceptance of alternative methods of performance evaluation is needed.

In conclusion, residency programs need to consider cost, feasibility, accuracy and, most importantly, the purpose for which the performance evaluation will be used when choosing a method. Findings from this study suggest that patient surveys or chart abstraction could be used to evaluate the overall performance of residents in a training program. However, there was considerable variability at the resident level between the 2 methods, and neither may be a completely accurate representation of actual performance. Thus, if findings of preventive health measures are to be used only for educational purposes, some variability may be acceptable, and patient surveys are an obvious choice because of their low cost. Based on our results, we cannot recommend either method to be used in a system of rewards or punishment where the need for a precise and accurate measurement of individual performance exists.

Sources of Funding: This study was supported by a grant from the UAB Health Services Foundation General Endowment Fund. Dr. Palonen was supported by Office of Academic Affiliation, Veterans Health Administration, as a VA National Quality Scholars Fellow.

REFERENCES

1. **The Accreditation Council for Graduate Medical Education.** Outcome Project. Available at: www.acgme.org/Outcome/. Accessed July 12, 2004.
2. **Kern DE, Harris WL, Boekeloo BO, Barker LR, Hogeland P.** Use of an outpatient medical record audit to achieve educational objectives: changes in residents' performances over six years. *J Gen Intern Med.* 1990;5:218-24.
3. **Callahan EJ, Bertakis KD.** Development and validation of the Davis Observation Code. *Fam Med.* 1991;23:19-24.
4. **Stange KC, Zyzanski SJ, Smith TF, et al.** How valid are medical records and patient questionnaires for physician profiling and health services research? A comparison with direct observation of patient's visits. *Med Care.* 1998;36:851-67.
5. **Wilson A, McDonald P.** Comparison of patient questionnaire, medical record, and audio tape in assessment of health promotion in general practice consultations. *BMJ.* 1994;309:1483-5.
6. **Luck J, Peabody JW, Dresselhaus TR, Lee M, Glassman P.** How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record. *Am J Med.* 2000;108:642-9.
7. **Peabody JW, Luck J, Glassman P, et al.** Measuring the quality of physician practice by using clinical vignettes: a prospective validation study. *Ann Intern Med.* 2004;141:771-80.
8. **Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M.** Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA.* 2000;283:1715-22.
9. **Nicholson JM, Hennrikus DJ, Lando HA, McCarty MC, Vessey J.** Patient recall versus physician documentation in report of smoking cessation counselling performed in the inpatient setting. *Tob Control.* 2000;9:382-8.
10. **Allison JJ, Wall TC, Spettell CM, et al.** The art and science of chart review. *Jt Comm J Qual Improv.* 2000;26:115-36.
11. **Holmboe E, Scranton R, Sumption K, Hawkins R.** Effect of medical record audit and feedback on residents' compliance with preventive health care guidelines. *Acad Med.* 1998;73:901-3.
12. **Montano DE, Phillips WR.** Cancer screening by primary care physicians: a comparison of rates obtained from physician self-report, patient survey, and chart audit. *Am J Public Health.* 1995;85:795-800.
13. **Gilchrist VJ, Stange KC, Flocke SA, McCord G, Bourguet CC.** A comparison of the National Ambulatory Medical Care Survey (NAMCS) measurement approach with direct observation of outpatient visits. *Med Care.* 2004;42:276-80.
14. **McPhee SJ, Richard RJ, Solkowitz SN.** Performance of cancer screening in a university general internal medicine practice: comparison with the 1980 American Cancer Society Guidelines. *J Gen Intern Med.* 1986;1:275-81.
15. **National Committee for Quality Assurance.** The health plan employer data and information set (HEDIS). Available at: <http://www.ncqa.org/Programs/HEDIS/>. Accessed July 25, 2004.
16. **Zapka JG, Bigelow C, Hurley T, et al.** Mammography use among sociodemographically diverse women: the accuracy of self-report. *Am J Public Health.* 1996;86:1016-21.
17. **Agency for Healthcare Research and Quality.** U.S. Preventive Services Task Force. Available at: <http://www.ahrq.gov/clinic/uspstfix.htm>. Accessed May 13, 2003.
18. **Gunning R.** The Technique of Clear Writing. New York, NY: McGraw-Hill International Book Company; 1952.
19. **Flesch R.** A new readability yardstick. *J Appl Psychol.* 1948;32:221-33.
20. **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;281:2098-105.
21. **Harwell TS, Moore K, Madison M, et al.** Comparing self-reported measures of diabetes care with similar measures from a chart audit in a well-defined population. *Am J Med Qual.* 2001;16:3-8.
22. **Armstrong K, Long JA, Shea JA.** Measuring adherence to mammography screening recommendations among low-income women. *Prev Med.* 2004;38:754-60.
23. **Dresselhaus TR, Peabody JW, Lee M, Wang MM, Luck J.** Measuring compliance with preventive care guidelines: standardized patients, clinical vignettes, and the medical record. *J Gen Intern Med.* 2000;15:782-8.
24. **Fowles JB, Rosheim K, Fowler EJ, Craft C, Arrichiello L.** The validity of self-reported diabetes quality of care measures. *Int J Qual Health Care.* 1999;11:407-12.
25. **Mant J, Murphy M, Rose P, Vessey M.** The accuracy of general practitioner records of smoking and alcohol use: comparison with patient questionnaires. *J Public Health Med.* 2000;22:198-201.
26. **Kell SH, Allison JJ, Brown KC, Weissman NW, Farmer R, Kiefe C.** Measurement of mammography rates for quality improvement. *Qual Manag Health Care.* 1999;7:11-9.
27. **Skinner KM, Miller DR, Lincoln E, Lee A, Kazis LE.** Concordance between respondent self-reports and medical records for chronic conditions: experience from the Veterans Health Study. *J Ambul Care Manage.* 2005;28:102-10.
28. **Ward MM, Doebbeling BN, Vaughn TE, et al.** Effectiveness of a nationally implemented smoking cessation guideline on provider and patient practices. *Prev Med.* 2003;36:265-71.
29. **Cardozo LJ, Steinberg J, Lepczyk MB, Binns-Emerick L, Cardozo Y, Aranha AN.** Improving preventive health care in a medical resident practice. *Arch Int Med.* 1998;158:261-4.