



Online article and related content
current as of September 11, 2009.

Improving Quality Improvement Using Achievable Benchmarks For Physician Feedback: A Randomized Controlled Trial

Catarina I. Kiefe; Jeroan J. Allison; O. Dale Williams; et al.

JAMA. 2001;285(22):2871-2879 (doi:10.1001/jama.285.22.2871)

<http://jama.ama-assn.org/cgi/content/full/285/22/2871>

Correction	Contact me if this article is corrected.
Citations	This article has been cited 117 times. Contact me when this article is cited.
Topic collections	Quality of Care; Quality of Care, Other; Randomized Controlled Trial Contact me when new articles are published in these topic areas.
Related Articles published in the same issue	June 13, 2001 JAMA. 2001;285(22):2921.
Related Letters	Reimbursement as Incentive to Improve Physicians' Quality of Care Albert J. Finestone et al. JAMA. 2001;286(13):1575.

Subscribe
<http://jama.com/subscribe>

Permissions
permissions@ama-assn.org
<http://pubs.ama-assn.org/misc/permissions.dtl>

Email Alerts
<http://jamaarchives.com/alerts>

Reprints/E-prints
reprints@ama-assn.org

Improving Quality Improvement Using Achievable Benchmarks For Physician Feedback

A Randomized Controlled Trial

Catarina I. Kiefe, PhD, MD

Jeroan J. Allison, MD, MS

O. Dale Williams, PhD

Sharina D. Person, PhD

Michael T. Weaver, PhD

Norman W. Weissman, PhD

GAPS BETWEEN MEDICAL CARE as actually practiced and the recommendations derived from evidence-based research are large and widespread.¹⁻³ Because more complete use of these recommendations should result in the prevention of considerable morbidity and mortality,^{4,5} research on methods to bridge these gaps is important. Quality improvement approaches such as medical record audit and feedback, opinion leaders, academic detailing, chart-based reminders, and computerized decision support have been evaluated.⁶⁻¹⁷

As explained recently by Samsa and Matchar,¹⁸ testing the general continuous quality improvement (CQI) approach to health care in randomized controlled trials (RCTs) is rare and, perhaps of necessity, inconclusive.¹⁹ Testing specific interventions deriving from a CQI approach in RCTs is more common, but still not abundant.¹⁸ These RCTs represent efforts to examine improvement activities with the same rigorous standards of evidence as those becoming increasingly accepted in the practice of evidence-based medicine.²⁰ Our study is an RCT that tests

Context Performance feedback and benchmarking, common tools for health care improvement, are rarely studied in randomized trials. Achievable Benchmarks of Care (ABCs) are standards of excellence attained by top performers in a peer group and are easily and reproducibly calculated from existing performance data.

Objective To evaluate the effectiveness of using achievable benchmarks to enhance typical physician performance feedback and improve care.

Design Group-randomized controlled trial conducted in December 1996, with follow-up through 1998.

Setting and Participants Seventy community physicians and 2978 fee-for-service Medicare patients with diabetes mellitus who were part of the Ambulatory Care Quality Improvement Project in Alabama.

Intervention Physicians were randomly assigned to receive a multimodal improvement intervention, including chart review and physician-specific feedback (comparison group; n=35) or an identical intervention plus achievable benchmark feedback (experimental group; n=35).

Main Outcome Measure Preintervention (1994-1995) to postintervention (1997-1998) changes in the proportion of patients receiving influenza vaccination; foot examination; and each of 3 blood tests measuring glucose control, cholesterol level, and triglyceride level, compared between the 2 groups.

Results The proportion of patients who received influenza vaccine improved from 40% to 58% in the experimental group ($P<.001$) vs from 40% to 46% in the comparison group ($P=.02$). Odds ratios (ORs) for patients of achievable benchmark physicians vs comparison physicians who received appropriate care after the intervention, adjusted for pre-intervention care and nesting of patients within physicians, were 1.57 (95% confidence interval [CI], 1.26-1.96) for influenza vaccination, 1.33 (95% CI, 1.05-1.69) for foot examination, and 1.33 (95% CI, 1.04-1.69) for long-term glucose control measurement. For serum cholesterol and triglycerides, the achievable benchmark effect was statistically significant only after additional adjustment for physician characteristics (OR, 1.40 [95% CI, 1.08-1.82] and OR, 1.40 [95% CI, 1.09-1.79], respectively).

Conclusion Use of achievable benchmarks significantly enhances the effectiveness of physician performance feedback in the setting of a multimodal quality improvement intervention.

JAMA. 2001;285:2871-2879

www.jama.com

Author Affiliations: Division of Preventive Medicine (Drs Kiefe, Williams, and Person), Center for Outcomes and Effectiveness Research and Education (Drs Kiefe, Allison, Williams, Person, Weaver, and Weissman), Division of General Internal Medicine (Dr Allison), School of Nursing (Dr Weaver), and Department of Health Services Administration (Dr Weissman), University of Alabama at Birmingham.

Disclosure: The Achievable Benchmark of Care (ABC) system is trademarked by the University of Alabama at Birmingham under an initiative of the Agency for Healthcare Research and Quality (AHRQ).

Corresponding Author and Reprints: Catarina I. Kiefe, PhD, MD, University of Alabama at Birmingham, 1717 11th Ave S, MT 700, Birmingham, AL 35205-4785 (e-mail: ckiefe@uab.edu).

the addition of a new tool, the Achievable Benchmarks of Care (ABC), to the "toolbox" for CQI.²¹⁻²³

Audit and feedback methods, in which clinicians receive reports of their performance and usually are compared to the mean performance of a peer group, have been used and studied extensively, but few of these studies have been RCTs.^{24,25} One underlying theory holds that viewing personal performance within the context of peer performance is a powerful motivator for change.²⁶⁻²⁸ However, researchers have reached few firm conclusions on the benefits of such an approach.²⁹ In general, only modest benefits have been described and long-term sustainability has not been demonstrated.

Seeking a method to increase the effectiveness of using performance feedback to clinicians, we developed the achievable benchmark method,³⁰⁻³² which is calculated from the performance of all members of a peer group and represents a realistic standard of excellence attained by the top performers in that group. The achievable benchmark method has desirable statistical characteristics and has been well received by physicians.³³⁻³⁵ Our next step, after developing the method and confirming its face validity, was to evaluate its effectiveness in improving care with the most rigorous approach available.

We enlisted the population of Alabama physicians participating in the Ambulatory Care Quality Improvement Project (ACQIP) to test the effectiveness of achievable benchmarks in provider feedback. ACQIP was designed by the Health Care Financing Administration (HCFA) to improve quality of care for ambulatory Medicare patients with diabetes mellitus and conducted by peer review organizations (PROs) in Alabama, Iowa, and Maryland. To improve practice patterns, clinicians received multimodal interventions, including feedback of baseline performance data on quality measures. In this context, we performed an RCT to test the hypothesis that achievable benchmark-enhanced

feedback would result in more improvements in care than the "usual" feedback that is part of the multimodal interventions used nationwide in HCFA quality improvement projects.²⁶

METHODS

This group-randomized trial was conducted within the Alabama ACQIP, a HCFA-sponsored demonstration project designed to improve outpatient care of fee-for-service Medicare beneficiaries with diabetes mellitus.³⁶ We first describe the ACQIP design and then illustrate how we superimposed the achievable benchmark experiment on the ACQIP design with 70 ACQIP physicians.

Ambulatory Care Quality Improvement Project

Physicians in ACQIP were given performance feedback based on several quality measures. After feedback, the Alabama Quality Assurance Foundation (AQAF) and the Alabama PRO partnered with physicians to develop and implement quality improvement projects targeting the ACQIP performance measures. The baseline data collection period reflected performance of participating physicians from January 1, 1994, through June 30, 1995.^{37,38} After a structured and scheduled sequence of improvement efforts during 1996, follow-up data on the performance of the same physicians from January 1, 1997, through June 30, 1998, were collected.³⁶ HCFA has now implemented a national program of diabetes care improvement as part of its sixth scope of work and PROs are continuing to use methods similar to that of ACQIP.³⁹

Physician Selection. HCFA planned to recruit 100 physicians per state in 1995 from Alabama, Iowa, and Maryland. All physicians practicing family medicine, internal medicine, or endocrinology were identified from Medicare claims data. To be eligible, each was required to have a minimum of 25 eligible diabetic patients enrolled in a fee-for-service Medicare plan. From 561 eligible Alabama physicians, HCFA

generated a random sample that was used by AQAF to recruit for the ACQIP study. Physicians were invited to participate and when 100 had accepted, enrollment was closed.

Of the 97 initial Alabama ACQIP participating physicians enrolled in 1995, 70 completed the study in 1998. The 27 physicians lost to follow-up were practicing in a different environment, had retired, or were deceased.

Patient Selection. Eligible patients were identified from Medicare outpatient (Part B) files based upon a billing diagnosis of diabetes mellitus (*International Classification of Diseases, Ninth Revision: codes 105.00-250.9x*). Eligible patients were 65 years or older, had no end-stage renal disease, had a residence other than skilled nursing facility, and were alive at baseline. AQAF then assigned each patient to a primary care physician (family practice, general practice, internal medicine, osteopathy) or endocrinologist based on the number of office visits and the number of billable Medicare services provided.

For both baseline and follow-up assessment, we randomly selected and reviewed an average of 20 patient medical records for each physician. To ensure independence of baseline and follow-up observations, we planned to exclude patients from follow-up who had had their records reviewed at baseline.³⁶

Quality Measures. Through ACQIP and related projects, HCFA led the development of multiple quality measures for ambulatory diabetic patients with contributions from the American Diabetes Association (ADA), the National Committee for Quality Assurance, the American Academy of Family Physicians, the American College of Physicians, and the Veterans Health Administration.^{40,41} The indicators were designed to assess processes of care for quality improvement and were not intended to serve as standards of care. All indicators were dichotomous variables thought to be amenable to simple quality improvement measures. In general, the quality indicators allow for a

longer time frame to administer the clinical intervention (vaccine, aspect of physical examination, or laboratory test) than suggested by the ADA guidelines.⁴² This leniency means that performance for these indicators should be better than for the ADA guidelines because decreasing the time frame for the clinical intervention would probably decrease average indicator performance.

We ascertained from the medical record whether the following appropriate care was performed for each eligible patient at least once during the 18-month period: (1) measurement of long-term glucose control as reflected by at least 1 test of glycosylated hemoglobin (hemoglobin A_{1c}) or fructosamine, (2) measurement of serum cholesterol, (3) measurement of serum triglycerides, (4) measurement of serum creatinine, (5) performance of in-office foot examination, and (6) administration of influenza vaccine. Performance on an indicator was quantified by dividing the number of eligible patients who received the item by the total number of eligible patients.

During the study period, the guideline-recommended method for periodically assessing renal function in the absence of previously diagnosed proteinuria changed to the measurement of microalbuminuria.⁴³ Because we did not have quantitative assessment of microalbuminuria at baseline, it was not incorporated into the follow-up medical record abstraction. On the other hand, because periodic serum creatinine measurement is no longer recommended, we do not emphasize assessment of diabetic nephropathy screening but focus on the other 5 indicators.

Chart Review. All data for the quality measures were obtained from chart review according to methods previously described.⁴⁴ Charts were photocopied and abstracted centrally using MedQuest, publicly available software developed for HCFA (<http://www.hcfa.gov>). The ACQIP investigators developed a standardized chart review protocol and refined the protocol through pilot testing. As part of the protocol, abstractors underwent intense training with

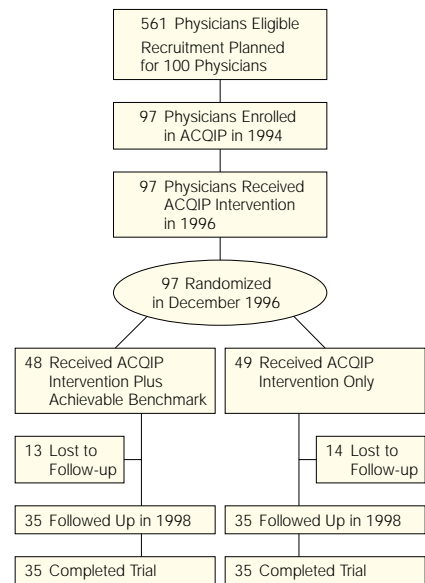
competency certification. The MedQuest chart review module contained standard lists for variable synonyms, medications, diagnoses, and procedures. Throughout the chart abstraction period, 5% of charts were randomly sampled for dual abstraction and physicians evaluated chart abstractions for validity. Validity and reliability of all key variables were at least 95%.

ACQIP Intervention. All ACQIP physicians participated in an intensive quality improvement program in which they were informed of their individual performance on the ACQIP indicators as well as of the mean performance of their peers (other participating Alabama physicians). Each physician received this information in mailings approximately 3 to 6 weeks apart during 1996, according to a schedule developed by HCFA and AQAF.³⁶ With assistance from AQAF, physician offices developed quality improvement plans (QIPs), currently on file at AQAF. The extensive and multimodal QIPs included formalized group meetings, root cause analysis, and changes of care at the office level, such as posting of patient educational material, use of chart interventions in the practice environment, reminders, clinical "flow sheets," and standing orders for appropriate administration of influenza vaccination. The QIPs were developed and documented according to a standardized and reproducible template.

Achievable Benchmark Experiment

We superimposed a group-randomized trial on the basic ACQIP design (FIGURE 1). In December 1996 we randomized the 97 Alabama ACQIP physicians to either the comparison or experimental achievable benchmark group. Of these 97 physicians, 27 were lost to follow-up. All ACQIP physicians not lost to follow-up agreed to participate in the achievable benchmark experiment, which consisted of adding, to the standard ACQIP intervention described above, an achievable benchmark for each indicator in the final report that was mailed to the achievable benchmark physicians, but not in

Figure 1. Trial Flow Diagram



Recruitment, randomization, and follow-up scheme for 97 Ambulatory Care Quality Improvement Project (ACQIP) physicians, 70 of whom participated in the achievable benchmark experiment.

the final report mailed to comparison physicians.³³ AQAF personnel who assisted the physician offices with developing the quality improvement projects were not informed as to which physicians received achievable benchmark feedback. Preintervention and postintervention changes in the experimental vs the comparison groups provided the main test of achievable benchmark effectiveness.

The achievable benchmark is calculated for a specific indicator of care, such as the percentage of eligible patients receiving influenza vaccination. In essence, the achievable benchmark represents the average performance for the top 10% of the physicians being assessed. In practice, adjustments are made to account for differences in the numbers of patients per physician and also to allow the inclusion of physicians with small numbers of eligible patients without unduly distorting the overall performance assessment.⁴⁵ Thus, an adjusted performance fraction (APF) is calculated for each physician by dividing the number of patients receiving the

Table 1. Characteristics of Eligible and Participating Alabama Physicians for the Ambulatory Care Quality Improvement Project (ACQIP), 1995*

	Eligible Physicians (n = 561)	Enrolled ACQIP Physicians (n = 97)	Physicians Completing Study (n = 70)
Year of first Alabama license			
Before 1979	52.0	39.2	44.3
1980-1989	42.8	43.3	47.1
1990 or after	3.4	7.2	4.3
Practice location			
Rural	32.8	30.9	35.7
Suburban	10.7	7.2	7.1
Urban	54.5	51.6	52.9
Graduate of US medical school			
Yes	87.5	76.3	81.4
No	10.7	13.4	14.3
Specialty			
Endocrinology	1.3	2.1	2.9
Family practice	50.8	35.1	37.1
Internal medicine	47.9	50.5	55.7

*Values are expressed as percentages, which may not add to 100 because of missing values. Comparison between eligible physicians vs physicians completing study and enrolled ACQIP physicians vs physicians completing study was not significant for any variable.

vaccination plus 1 by the number eligible for vaccination plus 2. The clinicians are then ranked, from highest to lowest, according to this APF until at least 10% of the patients for all the physicians have been included. The achievable benchmark calculation is then based on all the eligible patients for these top-ranked physicians and is the number of patients receiving the vaccination divided by the number eligible.

Details of the achievable benchmark method and its theoretical underpinnings are published elsewhere.³⁰⁻³³ A computer program for achievable benchmark computation, accompanied by a user manual, is posted on the Internet (<http://www.main.uab.edu/show.asp?durki=11311>) and will be provided upon request.

Statistical Analyses

The achievable benchmark experiment was a group-randomized trial, in that patients were nested within physicians, with physicians the unit of randomization and also the unit of some, but not all, of the analyses. To take full advantage of the available information, we also conducted some analyses with the patient as unit of analysis using techniques appropriate for the analysis of group-randomized trials.^{46,47}

We examined baseline demographics of the physicians and their patients. We also compared study physicians with nonparticipating Alabama physicians. Separate analyses were performed for each indicator. With the physician as unit of analysis, we used paired *t* tests to compare the mean baseline and follow-up performance of achievable benchmark intervention physicians (n=35) and then repeated this analysis for comparison physicians. To evaluate the statistical significance and magnitude of the achievable benchmark effect, ie, between-treatment differences in postintervention performance, we used generalized linear models. These models considered nesting of the 2978 patients within physicians and contained baseline performance as a covariate to adjust for any pre-intervention performance differences.^{48,49} We used a logit link to account for the binary nature of the response variable.

We also developed patient-level generalized linear models to estimate the odds of receiving a recommended intervention according to study arm after adjusting for physician characteristics. We did not adjust for patient characteristics because each quality measure specified a group of patients who were ideal candidates for the intervention. The applicability of each

quality measure in this study does not depend upon the patient characteristics. For example, diabetic patients should receive influenza vaccination regardless of whether they have hypertension, obesity, or coronary artery disease. Palmer⁵⁰ cogently argues that process measures that carefully identify ideal candidates for a procedure often do not require risk adjustment. Therefore, overadjustment for patient characteristics would obscure important findings. SAS Version 8.0 statistical software was used for the statistical analyses (SAS Institute Inc).

RESULTS

Physician and Patient Characteristics

In general, Alabama physicians completing the ACQIP study did not differ significantly from all physicians eligible for participation in ACQIP or from all physicians initially enrolled in ACQIP (TABLE 1). The physicians randomized to the experimental and comparison arms of the achievable benchmark experiment were not significantly different regarding years in practice, practice location, country of medical school attended, and specialty (TABLE 2). Patients of comparison and achievable benchmark physicians were similar in age, race, and pertinent comorbidity both at baseline and at follow-up (TABLE 3). Contrary to initial HCFA plans, 313 of the 1360 patients studied at follow-up had been included in the baseline group as well. To address this issue, we performed all analyses with and without these patients. Our results did not change substantially, although there was some loss of statistical significance with the reduced sample size.

Effectiveness of Achievable Benchmark Method

The achievable benchmarks for each indicator were: (1) influenza vaccination, 82%; (2) foot examination, 86%; (3) long-term glucose control measurement, 97%; (4) cholesterol measurement, 99%; and (5) triglycerides mea-

surement, 98%. Both groups of physicians had a mean preintervention influenza vaccination rate of 40%; physicians receiving achievable benchmarks improved to a postintervention rate of 58%, while comparison physicians improved to 46% (FIGURE 2). In addition, both experimental and comparison groups improved significantly on foot examination (46% to 61% vs 32% to 45%) and long-term glucose control measurement (31% to 70% vs 30% to 65%). For cholesterol measurement, the experimental group improved significantly (66% to 72%) while the comparison group did not (66% to 69%). The changes for triglyceride measurement (61% to 65% vs 57% to 60%) were not significant.

Patients of achievable benchmark physicians had significantly higher adjusted odds of receiving appropriate care at follow-up compared with patients of comparison physicians for influenza vaccination, foot examination, and measurement of long-term glucose control (TABLE 4).

Adjustment for Physician Characteristics

After adjustment for the physician characteristics and baseline performance using generalized linear models, patients of urban physicians tended to receive more appropriate care (TABLE 5). In addition, international medical graduates were more likely to perform foot examinations. Physicians who graduated after 1970 were more likely to order influenza vaccination but less likely to order lipid testing for their patients. Family practitioners were more likely than internists to order influenza vaccination. Finally, even after adjustment for multiple physician characteristics, patients of physicians assigned to the experimental study arm had significantly higher odds of receiving appropriate care at follow-up on all 5 measures.

COMMENT

In this RCT, we demonstrated that achievable benchmark feedback im-

proved clinician performance beyond the effect produced by an underlying improvement intervention, which in itself was associated with significant overall improvement for most quality measures. For influenza vaccination, foot

Table 2. Characteristics of Physicians by Experimental Group, at Baseline and Follow-up*

	Initial ACQIP Physicians (1995)		Physicians Completing ACQIP Study (1998)	
	Experimental (n = 48)	Comparison (n = 49)	Experimental (n = 35)	Comparison (n = 35)
Year of graduation from medical school				
Before 1970	9 (19)	11 (22)	7 (20)	9 (26)
1970-1979	15 (31)	18 (37)	12 (34)	13 (37)
1980 or after	18 (37.5)	16 (33)	15 (43)	11 (31)
Unknown	6 (12.5)	4 (8)	1 (3)	2 (6)
Practice location				
Rural	15 (31)	15 (31)	12 (34)	13 (37)
Suburban	3 (6)	4 (8)	2 (6)	3 (9)
Urban	24 (50)	26 (53)	20 (57)	17 (49)
Unknown	6 (13)	4 (8)	1 (3)	2 (6)
Graduate of US medical school				
Yes	35 (73)	39 (80)	28 (80)	29 (83)
No	7 (15)	6 (12)	6 (17)	4 (11)
Unknown	6 (13)	4 (8)	1 (3)	2 (6)
Specialty				
Endocrinology	0 (0)	2 (4)	0 (0)	2 (6)
Family practice	15 (31)	19 (39)	13 (37)	13 (37)
Internal medicine	26 (54)	23 (47)	21 (60)	18 (51)
Other	1 (2)	1 (2)	0 (0)	0 (0)
Unknown	6 (13)	4 (8)	1 (3)	2 (6)

*Values are expressed as number of participants and percentages (%). Percentages may not add to 100 due to rounding. No comparisons are significant at P<.05. ACQIP indicates Ambulatory Care Quality Improvement Project.

Table 3. Patient Characteristics for Baseline and Remeasurement Cohort, by Experiment Group, Alabama ACQIP*

	Patients of Experimental Physicians	Patients of Comparison Physicians	P Value
Baseline Cohort, 1994-1995			
	(n = 965)	(n = 966)	
Mean age, y	75.9	76.1	.66
Race			
White	37.8	34.0	.33
Black	18.2	21.3	.33
Hypertension	83.9	83.6	.88
Obesity	25.7	23.1	.18
Coronary artery disease	39.0	39.5	.81
Peripheral vascular disease	19.8	21.6	.36
Remeasurement Cohort, 1997-1998			
	(n = 678)	(n = 682)	
Mean age, y	75.2	74.8	.20
Race			
White	45.1	44.7	.88
Black	16.5	19.5	.15
Other	0.0	0.0	.32
Undetermined	38.2	35.3	.27
Hypertension	84.9	87.4	.18
Obesity	31.2	31.2	.99
Coronary artery disease	41.7	38.6	.24
Peripheral vascular disease	22.0	20.4	.46

*Values are expressed as percentages. ACQIP indicates Ambulatory Care Quality Improvement Project.

examination, and long-term glucose control measurement, physician receipt of achievable benchmark feedback was associated with 33% to 57% higher odds of patients receiving appropriate care at follow-up compared with patients of comparison physicians.

Interpreting the magnitude of the observed achievable benchmark effect demands consideration of the relative harm, benefit, and cost of each clinical intervention and of the incremental cost of adding an achievable benchmark to existing feedback programs.⁵¹ Because achievable benchmarks are easy to calculate from existing data, they are a simple tool for enhancing audit and feedback approaches. Also,

there is no foreseeable harm from adding achievable benchmarks to clinician profiles. Although we did not quantify the cost of using achievable benchmarks, it is modest given that the tool requires no data collection beyond that necessary for the usual audit and feedback process.

Because the quality measures we studied are backed by evidence and are applicable to a high proportion of diabetic patients, using the achievable benchmark in population-wide initiatives could benefit a substantial number of patients.⁵² For example, the prevalence of type 2 diabetes mellitus is estimated to be 15.6 million in the United States.⁵³ Between 1994 and 1997, some 20 000 to 40 000 deaths per

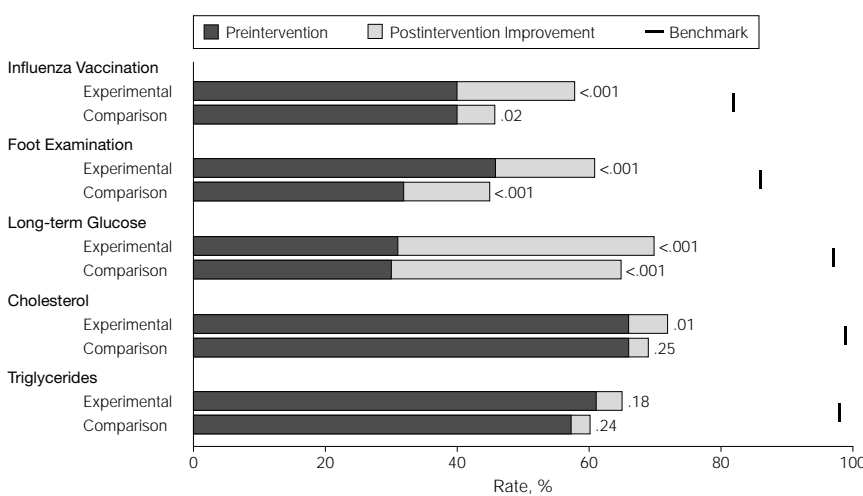
year were attributable to influenza and subsequent pneumonia.⁵⁴ Of all patients dying with influenza and pneumonia, approximately 18% are estimated to be diabetic,⁵⁵ resulting in some 3600 to 7200 deaths per year in diabetic patients. A meta-analysis revealed the efficacy of influenza vaccine in preventing death to be about 68%.⁵⁶ Given that 12% more patients of physicians who received achievable benchmark care compared with patients of physicians who did not receive achievable benchmark care were vaccinated for influenza and assuming that none of those dying from influenza were vaccinated, between 294 and 587 deaths, and substantially more episodes of influenza and pneumonia, could be prevented per year.

A previously published survey of achievable benchmark physicians (81% response rate) conducted after feedback provides some insight into possible mechanisms of achievable benchmark influence on practice patterns.^{33,34} Of all respondents, 74% considered specific new approaches to improve care, 63% identified new approaches specifically for their office, and 55% actually implemented new approaches. The most frequently reported process change was the incorporation of practice flow sheets and the most frequently reported change in clinical practice was an increase in office foot examinations.

Our study was conducted in the office setting, where the individual physician has direct influence. We speculate that comparison of individual performance with an achievable benchmark might have provided additional motivation for change. This speculation is consistent with social cognition models of change that emphasize provider perceptions and attitudes.^{57,58} However, we recognize the complexity and incomplete picture presented by the literature on changing provider behavior.⁵⁹

The concept of benchmarking is widely included in the improvement literature. However, traditional definitions of benchmarks and benchmark providers have been subjective and opin-

Figure 2. Preintervention and Postintervention Performance of 5 Quality of Care Measures



Mean unadjusted performance of clinicians before Health Care Financing Administration (HCFA) Ambulatory Care Quality Improvement Project (ACQIP) intervention and after ACQIP intervention (entire bars, showing preintervention performance plus postintervention change), by achievable benchmark experimental status. Each pair of bars represents 1 indicator. P values at the end of each bar represent preintervention and postintervention comparison in each group.

Table 4. Adjusted Odds Ratios for Receipt of Appropriate Postintervention Care for Patients of Achievable Benchmark Physicians vs Patients of Comparison Physicians, Alabama ACQIP*

Quality Measure	Odds Ratio (95% Confidence Interval)	P Value
Influenza vaccine	1.57 (1.26-1.96)	<.001
Foot examination	1.33 (1.05-1.69)	.02
Long-term glucose measurement	1.33 (1.04-1.69)	.02
Serum cholesterol measurement	1.20 (0.95-1.51)	.13
Serum triglycerides measurement	1.15 (0.92-1.44)	.22

*Odds ratios, 95% confidence intervals, and P values are based on generalized linear models with the patient as the unit of analysis. Appropriate care receipt at follow-up is the dependent variable and physician experimental arm is the main independent variable. Models adjust baseline performance and for nesting of patients within physicians. Each row represents a separate model. ACQIP indicates Ambulatory Care Quality Improvement Project.

ion driven, rather than data driven.⁶⁰⁻⁶³ Since our earlier publications of the achievable benchmark method,³¹⁻³³ we have received many requests for information regarding its use. For example, the achievable benchmark method is now included in the toolkit published by the US Public Health Service for states wishing to reach Healthy People 2010 goals⁶⁴ and has also been adopted by PROs conducting HCFA-sponsored quality improvement projects.³⁵

Our finding that physicians in rural settings were less likely to improve with feedback is intriguing. There is substantial literature on rural health care delivery, yet little is available on improvement efforts based on quality of care measurement in rural areas.⁶⁵⁻⁶⁹ Also, our data suggest that recent medical school graduates, foreign medical graduates, and family practitioners may be more responsive to improvement efforts. Number of years in practice and country of medical school have received some peripheral attention in the literature on the relationship between physician characteristic and quality, but no clear pattern of association has emerged.⁷⁰⁻⁷⁵ The relationship between physician specialty and quality of care has recently come under intense scrutiny.⁷⁶⁻⁷⁸ Because the identification of physician characteristics predicting success of quality improvement was not the focus of our analysis, we intend to study this issue separately.

Our study has several limitations. First, we included volunteers from a subset of Alabama physicians. How-

ever, physicians were chosen by a stratified randomization process, and we did not find significant differences between eligible physicians and physicians who completed this study. Because we were investigating how to improve quality improvement methods, working with volunteer physicians is consistent with our research objectives. In fact, many of the current improvement efforts, including those promoted by HCFA and other organizations, follow general principles of CQI and rely on voluntary participation of providers.

Second, the improvement attributable to achievable benchmark feedback may or may not persist over time. This is a limitation of what is known about audit and feedback, as well as about improvement efforts in general, and constitutes an area of active research.^{4,24,25,29}

Third, to inexpensively calculate an achievable benchmark, one needs a readily available data source. Although we used existing chart-review data in this project, we have demonstrated in previous publications that achievable benchmarks may be calculated from other data sources such as the National Health Interview Survey³⁰ and Medicare administrative data.³⁷

Finally, questions regarding the achievable benchmark method itself remain, including whether it can be shown to be equally effective in other settings, such as in inpatient or managed care outpatient improvement efforts. Also, we do

not know that adding the achievable benchmark method to a less-intense quality improvement program would produce similar incremental benefit. However, we note again that intensive multimodal quality improvement projects are frequently used. HCFA has now implemented several national Medicare programs based on the ACQIP model.³⁹ Similar multimodal quality improvement programs are also espoused by the Joint Commission on Accreditation of Healthcare Organizations,²³ and managed care organizations frequently conduct multimodal quality improvement interventions of equal intensity.

We believe that, as the measurement of quality becomes ever more important in health care, it behooves researchers in this area to subject their approaches to rigorous scrutiny where feasible. Our study represents a methodological advance in that it demonstrates the effectiveness of a simple, new quality improvement tool. In addition, we rigorously expand the available literature on evaluating quality improvement tools.

In conclusion, with the achievable benchmark method we have added a new tool to the "toolbox" for translating evidence into practice. In the RCT reported here, we have demonstrated this tool's effectiveness in enhancing audit and feedback based on medical record review in the ambulatory setting. With the current imperative to improve health care delivery, any effective addition to improvement efforts deserves close attention. When calculated from existing data

Table 5. Odds Ratios for Receipt of Appropriate Care Postintervention With Adjustment for Physician Characteristics, Alabama ACQIP*

Quality Measure	Physician Characteristics					
	Experimental Study Group (vs Comparison)	Rural Practice (vs Urban)	International (vs US Medical Graduate)	Graduation Year (vs Before 1970)		Internal Medicine (vs Family Practice)
				1970-1979	≥1980	
Influenza vaccine	1.54 (1.21-1.96)	0.50 (0.35-0.69)	1.17 (0.77-1.75)	2.05 (1.26-3.35)	2.27 (1.22-4.25)	0.60 (0.42-0.85)
Foot examination	1.32 (1.00-1.72)	0.83 (0.59-1.16)	1.72 (1.13-2.63)	1.06 (0.65-1.73)	1.10 (0.58-2.06)	1.09 (0.76-1.56)
Long-term glucose	1.32 (1.01-1.74)	1.38 (0.96-2.01)	0.65 (0.42-1.01)	0.91 (0.53-1.53)	0.71 (0.36-1.38)	0.91 (0.62-1.35)
Serum cholesterol	1.40 (1.08-1.82)	0.67 (0.47-0.95)	1.31 (0.84-2.08)	0.58 (0.35-0.96)	1.60 (0.86-2.99)	1.39 (0.94-2.04)
Serum triglycerides	1.40 (1.09-1.79)	0.49 (0.35-0.69)	1.21 (0.80-1.85)	0.63 (0.39-0.99)	1.74 (0.99-3.08)	0.86 (0.59-1.24)

*Odds ratios and 95% confidence intervals are based on generalized linear models with the patient as the unit of analysis and appropriate care receipt at follow-up as the dependent variable, adjusting for baseline performance, for nesting of patients within physicians, and for physician characteristics. Each row represents a separate model. Referent category named in parentheses.

and used as an enhancement to an existing feedback program, the achievable benchmark method has the advantages of being neither resource intensive nor inherently hazardous. With high face validity, the peer-based, data-driven achievable benchmark method has many advantages over subjectively defined benchmarks and represents an advance in the methodology of quality measurement and improvement.

Author Contributions: *Study concept and design:* Keife, Allison, Williams, Weissman. *Acquisition of data:* Keife, Allison, Weissman. *Analysis and interpretation of data:* Keife, Allison, Person, Weaver, Weissman. *Drafting of the manuscript:* Keife, Allison, Williams, Weissman. *Critical revision of the manuscript for important intellectual content:* Keife, Allison, Williams, Person, Weaver, Weissman. *Statistical expertise:* Keife, Williams, Person, Weaver, Weissman. *Obtained funding:* Keife, Allison, Weissman. *Administrative, technical, or material support:* Allison, Williams. *Study supervision:* Keife, Allison.

Funding/Support: This work was supported by grants HS09446, HS11124, and HS/GM10389 from the Agency for Healthcare Research and Quality and conducted in cooperation with the Alabama Quality Assurance Foundation and the Health Care Financing Administration. Dr Keife is a Senior Scholar in the Veterans Affairs National Quality Scholars Fellowship Program.

Disclaimer: The analyses upon which this publication is based were performed under contract 500-99-AL02 entitled "Utilization and Quality Control Peer Review Organization for the State (Commonwealth) of Alabama," sponsored by the Health Care Financing Administration, Department of Health and Human Services. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. The authors assume full responsibility for the accuracy and completeness of the ideas presented. This article is a direct result of the Health Care Quality Improvement Program initiated by the Health Care Financing Administration, which has encouraged identification of quality improvement projects derived from analysis of patterns of care, and therefore required no special funding on the part of this contractor. Ideas and contributions to the authors concerning experience in engaging with issues presented are welcomed.

REFERENCES

1. Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *BMJ*. 1998;317:465-468.
2. Col NF, McLaughlin TJ, Soumerai SB, et al. The impact of clinical trials on the use of medications for acute myocardial infarction: results of a community-based study. *Arch Intern Med*. 1996;156:54-60.
3. Greer AL. The state of the art versus the state of the science: the diffusion of new medical technologies into practice. *Int J Technol Assess Health Care*. 1988;4:5-26.

4. Chassin MR, Galvin RW. The urgent need to improve health care quality. *JAMA*. 1998;280:1000-1005.
5. Schuster MA, McGlynn EA, Brook RH. How good is the quality of health care in the United States? *Milbank Q*. 1998;76:517-563.
6. Greco PJ, Eisenberg JM. Changing physicians' practices. *N Engl J Med*. 1993;329:1271-1273.
7. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance: a systematic review of the effect of continuing medical education strategies. *JAMA*. 1995;274:700-705.
8. Davis DA, Thomson MA, Oxman AD, Haynes RB. Evidence for the effectiveness of CME: a review of 50 randomized controlled trials. *JAMA*. 1992;268:1111-1117.
9. Thomson O'Brien MA, Oxman AD, Haynes RB, Davis DA, Freemantle N, Harvey EL. Local opinion leaders: effects on professional practice and health care outcomes [Cochrane Database of Systematic Reviews]. Oxford, England: Cochrane Library, Update Software; 2000;issue 2.
10. Thomson O'Brien MA, Oxman AD, Davis DA, Haynes RB, Freemantle N, Harvey EL. Educational outreach visits: effects on professional practice and health care outcomes [Cochrane Database of Systematic Reviews]. Oxford, England: Cochrane Library, Update Software; 2000;issue 2.
11. Tierney WM, Overhage JM, Takesue BY, et al. Computerizing guidelines to improve care and patient outcomes: the example of heart failure. *J Am Med Inform Assoc*. 1995;2:316-322.
12. Marshall PS, Cooper S. Influencing physician prescribing through academic detailing with benchmarks. *Am J Health Syst Pharm*. 1998;55:2234-2236.
13. Soumerai SB, Avorn J. Principles of educational outreach ("academic detailing") to improve clinical decision making. *JAMA*. 1990;263:549-556.
14. Buntinx F, Winkens R, Grol R, Knotterus JA. Influencing diagnostic and preventive performance in ambulatory care by feedback and reminders. *Fam Pract*. 1993;10:219-228.
15. Lomas J, Enkin M, Anderson GM, Hannah WJ, Vayda E, Singer J. Turning practice guidelines into actual practice: a randomized controlled trial of opinion leaders and audit/feedback in community hospitals. Center for Health Economics and Policy Analysis. Vol 19. Hamilton, Ontario: McMaster University; 1990.
16. Calkins E, Katz LA, Karuza J, Wagner A. The small group consensus process for changing physician practices: influenza vaccination. *HMO Pract*. 1995;9:107-110.
17. Soumerai SB, McLaughlin TJ, Gurwitz JH, et al. Effect of local medical opinion leaders on quality of care for acute myocardial infarction: a randomized controlled trial. *JAMA*. 1998;279:1358-1363.
18. Samsa G, Matchar D. Can continuous quality improvement be assessed using randomized trials? *HSR: Health Serv Res*. 2000;35:687-700.
19. Shortell SM, Bennett CL, Byck GR. Assessing the impact of continuous quality improvement on clinical practice: what it will take to accelerate progress. *Milbank Q*. 1998;76:593-624.
20. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine: How to Practice and Teach EBM*. 2nd ed. New York, NY: Churchill Livingstone; 2000.
21. Berwick DM, Godfrey AB, Rossner J. *Curing Health Care: Best Strategies for Quality Improvement*. San Francisco, Calif: Jossey-Bass; 1990.
22. Griffith JR, Sahney VK, Mohr RA. *Reengineering Health Care*. Ann Arbor, Mich: Health Administration Press; 1995.
23. Nelson EC, Batalden PB, Ryer JC. *Clinical Improvement Action Guide*. Oakbrook Terrace, Ill: Joint Commission on Accreditation of Healthcare Organizations; 1998.
24. Thomson O'Brien MA, Oxman AD, Davis DA,

- Haynes RB, Freemantle N, Harvey EL. Audit and feedback: effects on professional practice and health care outcomes [Cochrane Database of Systematic Reviews]. Oxford, England: Cochrane Library, Update Software; 2000;issue 2.
25. Thomson O'Brien MA, Oxman AD, Davis DA, Haynes RB, Freemantle N, Harvey EL. Audit and feedback versus alternative strategies: effects on professional practice and health care outcomes [Cochrane Database of Systematic Reviews]. Oxford, England: Cochrane Library Update Software; 2000;issue 2.
26. Jencks SF, Wilensky GR. The health care quality improvement initiative: a new approach to quality assurance in Medicare. *JAMA*. 1992;268:900-903.
27. Jencks SF. Changing health care practices in Medicare's Health Care Quality Improvement Program. *Jt Comm J Qual Improv*. 1995;21:343-347.
28. Hayes RP, Ballard DJ. Review: feedback about practice patterns for measurable improvements in quality of care—a challenge for PROs under the Health Care Quality Improvement Program. *Clin Perform Qual Health Care*. 1995;3:15-22.
29. Oxman AD, Thomson MA, Davis DA, Haynes RB. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *CMAJ*. 1995;153:1423-1431.
30. Allison JJ, Keife CI, Weissman NW. Can data-driven benchmarks be used to set the goals of Healthy People 2010? *Am J Public Health*. 1999;89:61-65.
31. Keife CI, Woolley TW, Allison JJ, Box JB, Craig AS. Determining benchmarks: a data driven search for the best achievable performance. *Clin Perform Qual Health Care*. 1994;2:190-194.
32. Keife CI, Weissman NW, Allison JJ, Farmer RM, Weaver M, Williams OD. Identifying achievable benchmarks of care (ABCs): concepts and methodology. *Int J Qual Health Care*. 1998;10:443-447.
33. Weissman NW, Allison JJ, Keife CI, et al. Achievable benchmarks of care: the ABCs of benchmarking. *J Eval Clin Pract*. 1999;5:203-215.
34. Pemberton JK, Keife C, Weissman NW. Physician perspectives on clinical performance feedback with achievable benchmarks of care. *Abstract Book/Association for Health Services Research*. 1998;15:202-203.
35. Texas Medical Foundation. CCP benchmarking provides achievable goals for hospitals. *Natl Perspect*. 1998;1:2-3.
36. *Narrative Project Document*. Alabama Quality Assurance Foundation Inc; 1996.
37. Kell SH, Allison JJ, Brown KC, Weissman NW, Farmer R, Keife C. Measurement of mammography rates for quality improvement. *Qual Manage Health Care*. 1999;7:11-19.
38. Kell SH, Drass J, Bausell RB, Thomas KA, Osborn MA, Gohdes D. Measures of disease control in Medicare beneficiaries with diabetes mellitus. *J Am Geriatr Soc*. 1999;47:417-422.
39. Office of Clinical Standards and Quality. *Peer Review Organization Manual*. Washington, DC: Health Care Financing Administration; 1999.
40. Mayfield J. Who cares about the quality of diabetes care? almost everyone! *Clin Diabet*. 1998;16:161-167.
41. Diabetes Quality Improvement Project. *Diabetes Quality Improvement Project MedQuest Data Abstraction & Analysis System User's Guide*. Austin: Texas Medical Foundation; 1999.
42. American Diabetes Association. Clinical practice recommendations 1999. *Diabetes Care*. 1999;22 (suppl 1):S1-S114.
43. American Diabetes Association. Clinical practice recommendations 2000. *Diabetes Care*. 2000;23 (suppl 1):S1-S24.
44. Allison JJ, Wall TC, Spettell CM, et al. The art and science of chart review. *Jt Comm J Qual Improv*. 2000; 26:115-136.
45. Agresti A. *Categorical Data Analysis*. New York, NY: John Wiley & Sons; 1990.

46. Murray DM. *Design and Analysis of Group-Randomized Trials*. New York, NY: Oxford University Press; 1998.
47. Hennekens CH, Buring JE. Validity versus generalizability in clinical trial design and conduct. *J Card Fail*. 1998;4:239-241.
48. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73:13-22.
49. McCullagh P, Nelder J. *Generalized Linear Models*. 2nd ed. New York, NY: Chapman & Hall; 1989.
50. Palmer RH. Process-based measures of quality: the need for detailed clinical data in large health care databases. *Ann Intern Med*. 1997;127:733-738.
51. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine: panel on cost-effectiveness in health and medicine. *JAMA*. 1996;276:1172-1177.
52. Hennekens C, Buring J. Measures of disease frequency and association. In: Marrent S, ed. *Epidemiology in Medicine*. Boston, Mass: Little Brown & Co; 1987:54-98.
53. Harris MI, Flegal KM, Cowie CC, et al. Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in US adults: the Third National Health and Nutrition Examination Survey, 1988-1994. *Diabetes Care*. 1998;21:518-524.
54. Brammer TL, Izurieta HS, Fukuda K, et al. Surveillance for influenza—United States, 1994-95, 1995-96, and 1996-97 seasons. *MMWR Morbid Mortal Wkly Rep. CDC Surveill Summ*. 2000;49:13-28.
55. Valdez R, Narayan KM, Geiss LS, Engelgau MM. Impact of diabetes mellitus on mortality associated with pneumonia and influenza among non-Hispanic black and white US adults. *Am J Public Health*. 1999;89:1715-1721.
56. Gross PA, Hermogenes AW, Sacks HS, Lau J, Levandowski RA. The efficacy of influenza vaccine in elderly persons: a meta-analysis and review of the literature. *Ann Intern Med*. 1995;123:518-527.
57. Marteau T, Sowden A, Armstrong D. Implementing research findings into practice: beyond the information deficit model. In: Haines A, Donald A, eds. *Getting Research Findings Into Practice*. London, England: BMJ Books; 1998.
58. Conner M, Norman P. The role of cognition in health behaviors. In: Conner M, Norman P, eds. *Predicting Health Behavior*. Buckingham, England: Open University Press; 1996.
59. Getting evidence into practice. Effective Health Care. Available at: <http://www.cochranelibrary.com>. Accessibility verified May 17, 2001.
60. Camp RC, Tweet AG. Benchmarking applied to health care. *J Qual Improve*. 1994;20:229-238.
61. Campbell AB. Benchmarking: a performance intervention tool. *J Qual Improve*. 1994;20:225-228.
62. Barnes RV, Lawton L, Briggs D. Clinical benchmarking improves clinical paths: experience with coronary artery bypass grafting. *J Qual Improve*. 1994;20:267-276.
63. Mohr JJ, Mahoney CC, Nelson EC, Batalden PB, Plume SK. Improving health care, part 3: clinical benchmarking for best patient care. *J Qual Improve*. 1996;22:599-616.
64. US Public Health Foundation. *Healthy People 2010 Toolkit*. Item RM-005. Washington, DC: US Public Health Foundation; 1999.
65. McKenzie A, Grylls J. Diabetic retinal photographic screening: a model for introducing audit and improving general practitioner care of diabetic patients in a rural setting. *Aust J Rural Health*. 1999;7:237-239.
66. Improving quality in small hospitals. *Qual Letter Healthcare Leaders*. 1999;11:2-7.
67. Fortney J, Rost K, Zhang M, Warren J. The impact of geographic accessibility on the intensity and quality of depression treatment. *Med Care*. 1999;37:884-893.
68. Goins RT, Mitchell J. Health-related quality of life: does rurality matter? *J Rural Health*. 1999;15:147-156.
69. Szafran O, Bell NR. Use of walk-in clinics by rural and urban patients. *Can Fam Physician*. 2000;46:114-119.
70. Richards C, Klabunde C, O'Malley M. Physicians' recommendations for colon cancer screening in women: too much of a good thing? *Am J Prev Med*. 1998;15:246-249.
71. Preisser JS, Cohen SJ, Wofford JL, et al. Physician and patient predictors of health maintenance visits. *Arch Fam Med*. 1998;7:346-351.
72. Lurie N, Margolis K, McGovern PG, Mink P. Physician self-report of comfort and skill in providing preventive care to patients of the opposite sex. *Arch Fam Med*. 1998;7:134-137.
73. Bindman AB, Grumbach K, Vranizan K, Jaffe D, Osmond D. Selection and exclusion of primary care physicians by managed care organizations. *JAMA*. 1998;279:675-679.
74. Heath K, Hogg RS, Singer J, Schechter MT, O'Shaughnessy MV, Montaner JS. Physician concurrence with primary care guidelines for persons with HIV disease. *Int J STD AIDS*. 1997;8:609-613.
75. Glatstein IZ, Harlow BL, Hornstein MD. Practice patterns among reproductive endocrinologists: the infertility evaluation. *Fertil Steril*. 1997;67:443-451.
76. MacLean CH, Louie R, Leake B, et al. Quality of care for patients with rheumatoid arthritis. *JAMA*. 2000;284:984-992.
77. Hillner BE, Smith TJ, Desch CE. Hospital and physician volume or specialization and outcomes in cancer treatment: importance in quality of cancer care. *J Clin Oncol*. 2000;18:2327-2340.
78. Chen J, Radford MJ, Wang Y, Krumholz HM. Care and outcomes of elderly patients with acute myocardial infarction by physician specialty: the effects of comorbidity and functional limitations. *Am J Med*. 2000;108:460-469.

Curiosity is one of the permanent and certain characteristics of a vigorous mind.
—Samuel Johnson (1709-1784)