

SPECIAL ARTICLE

The Quality of Health Care Delivered to Adults in the United States

Elizabeth A. McGlynn, Ph.D., Steven M. Asch, M.D., M.P.H., John Adams, Ph.D., Joan Keesey, B.A., Jennifer Hicks, M.P.H., Ph.D., Alison DeCristofaro, M.P.H., and Eve A. Kerr, M.D., M.P.H.

ABSTRACT

BACKGROUND

We have little systematic information about the extent to which standard processes involved in health care — a key element of quality — are delivered in the United States.

METHODS

We telephoned a random sample of adults living in 12 metropolitan areas in the United States and asked them about selected health care experiences. We also received written consent to copy their medical records for the most recent two-year period and used this information to evaluate performance on 439 indicators of quality of care for 30 acute and chronic conditions as well as preventive care. We then constructed aggregate scores.

RESULTS

Participants received 54.9 percent (95 percent confidence interval, 54.3 to 55.5) of recommended care. We found little difference among the proportion of recommended preventive care provided (54.9 percent), the proportion of recommended acute care provided (53.5 percent), and the proportion of recommended care provided for chronic conditions (56.1 percent). Among different medical functions, adherence to the processes involved in care ranged from 52.2 percent for screening to 58.5 percent for follow-up care. Quality varied substantially according to the particular medical condition, ranging from 78.7 percent of recommended care (95 percent confidence interval, 73.3 to 84.2) for senile cataract to 10.5 percent of recommended care (95 percent confidence interval, 6.8 to 14.6) for alcohol dependence.

CONCLUSIONS

The deficits we have identified in adherence to recommended processes for basic care pose serious threats to the health of the American public. Strategies to reduce these deficits in care are warranted.

From RAND, Santa Monica, Calif. (E.A.M., S.M.A., J.A., J.K., J.H., A.D.); the Veterans Affairs (VA) Greater Los Angeles Health Care System, Los Angeles (S.M.A.); the Department of Medicine, University of California Los Angeles, Los Angeles (S.M.A.); the VA Center for Practice Management and Outcomes Research, VA Ann Arbor Health Care System, Ann Arbor, Mich. (E.A.K.); and the Department of Medicine, University of Michigan, Ann Arbor (E.A.K.). Address reprint requests to Dr. McGlynn at RAND, 1700 Main St., P.O. Box 2138, Santa Monica, CA 90407, or at beth_mcglynn@rand.org.

N Engl J Med 2003;348:2635-45.

Copyright © 2003 Massachusetts Medical Society.

THE DEGREE TO WHICH HEALTH CARE in the United States is consistent with basic quality standards is largely unknown.^{1,2} Although previous studies have documented serious quality deficits, they provide a limited perspective on the issue.³⁻⁵ Most have assessed a single condition,^{6,7} a small number of indicators of quality,^{8,9} persons with a single type of insurance coverage,¹⁰ or persons receiving care in a small geographic area.^{11,12} The few national studies have been limited to specific segments of the population, such as Medicare beneficiaries¹³⁻¹⁵ or enrollees in managed-care plans¹⁶; have focused on a limited set of topics, such as preventive care,¹⁷ diabetes,¹⁸ or human immunodeficiency virus¹⁹; or have assessed health outcomes without a link to specific processes involved in care.²⁰ As a result, we have no comprehensive view of the level of quality of care given to the average person in the United States. This information gap contributes to a persistent belief that quality is not a serious national problem.¹

In this article, we report results from the Community Quality Index (CQI) study, a collateral study of the Community Tracking Study (CTS).²¹ The CTS, conducted by the Center for Studying Health System Change (CSHSC), monitors changes in health care markets in the United States. The CTS obtains self-reported information from a random sample of the U.S. population on their insurance coverage, patterns of utilization of health care services, and health status. The CSHSC has reported on trends in health care costs,²² factors affecting the choice of employer-sponsored or public insurance,²³ and changes in the structure of managed-care plans.²⁴ However, the CTS lacks detailed information about the implications of these variations in health care markets for the quality of health care. By collaborating with the CSHSC, we were able to assess the extent to which the recommended processes of medical care — one critical dimension of quality — are delivered to a representative sample of the U.S. population for a broad spectrum of conditions.

METHODS

RECRUITMENT OF PARTICIPANTS

In 12 metropolitan areas (Boston; Cleveland; Greenville, S.C.; Indianapolis; Lansing, Mich.; Little Rock, Ark.; Miami; Newark, N.J.; Orange County, Calif.; Phoenix, Ariz.; Seattle; and Syracuse, N.Y.), using random-digit-dial telephone surveys, the CTS deliberately recruited enough participants to assess how

structural characteristics in each market (e.g., the penetration of managed care) affect patterns of access to and utilization of health care services. Between October 1998 and August 2000, we recontacted by telephone households that had participated in the CTS interviews. Participants were asked to complete a telephone interview regarding their health history and to provide a listing of all individual or institutional health care providers whom they had seen during the previous two years. Participants who orally agreed to provide access to their medical records were sent written consent forms to sign and return to RAND. Photocopies of the medical records of participants providing written consent were sent to RAND for central abstracting.

RESPONSE RATES

Because of the complex, multistage nature of the study design, several calculations of the response rate are provided. Among the 20,028 adults in the initial sample, 2091 (10 percent) were deemed ineligible, primarily because they had left the area. Among the 17,937 eligible adults, 13,275 (74 percent) participated in the telephone interview regarding their health history, including 863 (7 percent) who had had no visits to a health care provider during the previous two years. Among the 12,412 participants who had had visits, 10,404 (84 percent) agreed orally to provide access to their medical records. We obtained written consent from 7528 (61 percent of those with visits to a provider). Participants reported having seen between 1 and 17 providers (mean, 2.6) during the study period. We obtained at least one record for 6712 (89 percent) of those who returned their consent forms. Overall, we received 84 percent of the records for which we had consent forms; we received all expected records for 4612 of the 6712 participants with consent forms and records (69 percent) and all but one record for 1547 of these participants (23 percent). Sensitivity analyses revealed few differences in results related to the completeness of records, so all participants for whom we obtained at least one record were included in the results we report (37 percent of the sample of eligible adults).

DEVELOPMENT OF INDICATORS OF QUALITY

The indicators of quality used in the study were derived from RAND's Quality Assessment Tools system.²⁵ RAND staff members selected acute and chronic conditions that represented the leading causes of illness, death, and utilization of health

care in each age group, as well as preventive care related to these causes. For each condition, staff physicians reviewed established national guidelines and the medical literature and proposed indicators of quality for all phases of care or medical functions (screening, diagnosis, treatment, and follow-up). We developed indicators to assess potential problems with the overuse and underuse of key processes. We primarily chose measures of processes as indicators, because they represent the activities that clinicians control most directly, because they do not generally require risk adjustment beyond the specification of eligibility, and because they are consistent with the structure of national guidelines.^{5,26}

Four nine-member, multispecialty expert panels were convened to assess the validity of the indicators proposed by the staff, using the RAND–UCLA modified Delphi method.²⁷ The members of the panels, nominated by the appropriate specialty societies, were diverse with respect to geography, practice setting, and sex. Indicators were rated on a 9-point scale (with 1 denoting not valid and 9 very valid). Only indicators with a median validity score of 7 or higher were included in the Quality Assessment Tools system. This method of selecting indicators is reliable²⁸ and has been shown to have content, construct, and predictive validity in other applications.^{29–32}

The criteria for the selection of conditions, reviews of the literature, the process followed by the panels, and the final indicators have been published elsewhere.^{33–36} (Further information on all the quality indicators used in this study is available at http://www.rand.org/health/mcglynn_appa.pdf or from the National Auxiliary Publications Service. *) Table 1 provides a brief description and classifications for a sample of the indicators we used. The classifications enabled us to examine quality from the perspective of what is being done (type of care), why it is being done (function), how it is being delivered (mode), and the nature of the quality problem (underuse or overuse). Results are based on 439 indicators for 30 conditions and preventive care.

HEALTH HISTORY INTERVIEW

We obtained selective information directly from respondents to augment information in their medical records. The health history took an average of 13 minutes to complete. The data obtained in this in-

*See NAPS document no. 05610 for 50 pages of supplementary material. To order, contact NAPS, c/o Microfiche Publications, 248 Hempstead Tpke., West Hempstead, NY 11552.

terview were used to refine the analysis of a respondent's eligibility for inclusion in the analysis or to augment the scoring for 22 of the 439 indicators. For example, we used reports of symptoms from participants with asthma to classify those with moderate-to-severe disease. We augmented scores for influenza or pneumococcal immunizations and screening for cancer on the basis of self-reports.

ABSTRACTING OF CHARTS

We developed computer-assisted abstraction software on a Visual Basic platform (version 6.0, Microsoft). The software allowed the manual abstraction of charts to be tailored to the specific record being reviewed and provided interactive checks of the quality of the data (for consistency and range), calculations (e.g., the determination of the presence of high blood pressure), and classifications (e.g., the determination of drug class) during abstraction.

Data for the study were abstracted by 20 trained registered nurses who had successfully abstracted a complex standard chart after a two-week training program. Charts were abstracted separately for each health care provider of each participant (i.e., at the dyad level). The average time required to abstract a chart for a participant–provider dyad was 50 minutes.

To assess interrater reliability, we re-abstracted charts from a randomly selected 4 percent sample of participants. Average reliability, with the use of the kappa statistic, ranged from substantial to almost perfect³⁷ at three levels: the presence or absence of a given condition ($\kappa=0.83$), the participant's eligibility for the process represented by a given indicator ($\kappa=0.76$), and scoring of a given indicator ($\kappa=0.80$).

STATISTICAL ANALYSIS

We specified the combination of variables necessary to determine whether each participant was or was not eligible for the process specified by each indicator and whether each participant did or did not receive each process or some proportion of it. Each indicator was scored at one of three levels — that of the individual participant, that of the participant–provider dyad, or that of the episode — depending on the nature of the process being evaluated. The level at which an indicator was scored affected the number of times a participant was eligible for the specified process; the resulting number served as the denominator in the calculation of the aggregate score. For participant-level indicators, we gave

Table 1. Selected Quality-of-Care Indicators and Classifications Used in the Community Quality Index Study.*

| Condition† | Description of Selected Indicator | Classification for Aggregate Scores | | | |
|---|---|-------------------------------------|-----------|-----------------------------------|----------------------|
| | | Type of Care | Function | Mode | Problem with Quality |
| Alcohol dependence (5 indicators) | | | | | |
| Indicator 2 | Assessment of alcohol dependence among regular or binge drinkers | For chronic condition | Diagnosis | History | Underuse |
| Indicator 4 | Treatment referral for persons given a diagnosis of alcohol dependence | For chronic condition | Treatment | Encounter or other intervention | Underuse |
| Asthma (25 indicators) | | | | | |
| Indicator 4 | Long-acting agents for patients with frequent use of short-acting beta-agonists | For chronic condition | Treatment | Medication | Underuse |
| Indicator 6 | Inhaled corticosteroids for patients receiving long-term systemic corticosteroid therapy | For chronic condition | Treatment | Medication | Underuse |
| Breast cancer (9 indicators) | | | | | |
| Indicator 1 | Appropriate follow-up of palpable mass | For chronic condition | Diagnosis | Laboratory testing or radiography | Underuse |
| Indicator 5 | Choice of surgical treatments for stage I or II cancer | For chronic condition | Treatment | Surgery | Underuse |
| Cerebrovascular disease (10 indicators) | | | | | |
| Indicator 4 | Antiplatelet therapy for noncardiac stroke or transient ischemic attack | For chronic condition | Treatment | Medication | Underuse |
| Indicator 5 | Carotid imaging for patients with symptomatic cardiovascular disease or transient ischemic attack | For chronic condition | Diagnosis | Laboratory testing or radiography | Underuse |
| Colorectal cancer (12 indicators) | | | | | |
| Indicator 1 | Screening for high-risk patients starting at 40 yr of age | Preventive | Screening | Laboratory testing or radiography | Underuse |
| Indicator 7 | Appropriate surgical treatment | For chronic condition | Treatment | Surgery | Underuse |
| Congestive heart failure (36 indicators) | | | | | |
| Indicator 1 | Ejection fraction assessed before medical therapy | For chronic condition | Diagnosis | Laboratory testing or radiography | Underuse |
| Indicator 32 | ACE inhibitors for patients with congestive heart failure and an ejection fraction <40% | For chronic condition | Treatment | Medication | Underuse |
| Coronary artery disease (37 indicators) | | | | | |
| Indicator 3 | Counseling on smoking cessation | For chronic condition | Treatment | Counseling or education | Underuse |
| Indicator 11 | Avoidance of nifedipine for patients with an acute myocardial infarction | For chronic condition | Treatment | Medication | Overuse |
| Diabetes (13 indicators) | | | | | |
| Indicator 9 | Diet and exercise counseling | For chronic condition | Treatment | Counseling or education | Underuse |
| Indicator 12 | ACE inhibitors for patients with proteinuria | For chronic condition | Treatment | Medication | Underuse |

Table 1. (Continued.)

| Condition | Description of Selected Indicator | Classification for Aggregate Scores | | | |
|---|---|-------------------------------------|-----------|-----------------------------------|----------------------|
| | | Type of Care | Function | Mode | Problem with Quality |
| Headache (21 indicators) | | | | | |
| Indicator 11 | CT or MRI for patients with new-onset headache and an abnormal neurologic examination | Acute | Diagnosis | Laboratory testing or radiography | Underuse |
| Indicator 15 | Use of appropriate first-line agents for patients with acute migraine | Acute | Treatment | Medication | Overuse |
| Hip fracture (9 indicators) | | | | | |
| Indicator 6 | Prophylactic antibiotics given on day of hip-repair surgery | Acute | Treatment | Medication | Underuse |
| Indicator 7 | Prophylactic antithrombotic drugs given on admission for patients with hip fracture | Acute | Treatment | Medication | Underuse |
| Hyperlipidemia (7 indicators) | | | | | |
| Indicator 4 | Treatment of high LDL cholesterol levels in patients with coronary artery disease | For chronic condition | Treatment | Medication | Underuse |
| Hypertension (27 indicators) | | | | | |
| Indicator 16 | Lifestyle modification for patients with mild hypertension | For chronic condition | Treatment | Counseling or education | Underuse |
| Indicator 18 | Pharmacotherapy for uncontrolled mild hypertension | For chronic condition | Treatment | Medication | Underuse |
| Indicator 27 | Change in treatment when blood pressure is persistently uncontrolled | For chronic condition | Follow-up | Medication | Underuse |
| Acute low back pain (6 indicators) | | | | | |
| Indicator 1 | Rule out cancer, fracture, infection, cauda equina syndrome, and neurologic causes | Acute | Diagnosis | History | Underuse |
| Indicator 6 | Avoidance of prolonged bed rest | Acute | Treatment | Other | Overuse |
| Preventive care (38 indicators) | | | | | |
| Indicator 1 | Screening for problem drinking | Preventive | Screening | History | Underuse |
| Indicator 2 | Mammographic screening for breast cancer | Preventive | Screening | Laboratory testing or radiography | Underuse |
| Indicator 3 | Screening for colorectal cancer in persons at average risk | Preventive | Screening | Laboratory testing or radiography | Underuse |
| Indicator 8 | Influenza vaccine for persons ≥65 yr of age | Preventive | Treatment | Immunization | Underuse |
| Indicator 21 | HIV testing for those at risk | Preventive | Screening | Laboratory testing or radiography | Underuse |
| Indicator 25 | Screening for cervical cancer | Preventive | Screening | Laboratory testing or radiography | Underuse |
| Indicator 29 | Smoking status documented | Preventive | Screening | History | Underuse |
| Indicator 31 | Annual advice for smokers to quit smoking | Preventive | Treatment | Counseling or education | Underuse |

| Condition | Description of Selected Indicator | Classification for Aggregate Scores | | | |
|--|--|-------------------------------------|-----------|-----------------------------------|----------------------|
| | | Type of Care | Function | Mode | Problem with Quality |
| Sexually transmitted diseases (26 indicators) | | | | | |
| Indicator 9 | Chlamydia screening for high-risk women | Preventive | Screening | Laboratory testing or radiography | Underuse |
| Indicator 24 | HIV screening in patients with sexually transmitted diseases | Acute | Screening | Laboratory testing or radiography | Underuse |

* ACE denotes angiotensin-converting enzyme, CT computed tomography, MRI magnetic resonance imaging, LDL low-density lipoprotein, and HIV human immunodeficiency virus.

† The number of indicators given in parentheses after each condition is the total number of indicators of quality of care for that condition; the indicators listed below each condition are examples.

the participant a score of “pass” if at least one of his or her health care providers had delivered the indicated care (e.g., influenza vaccination). For indicators scored at the level of the participant–provider dyad (e.g., smoking status noted in the chart), we scored every dyad separately, so the number of times the participant was counted in the denominator depended on the number of providers who saw the participant and could have performed the specified process. For indicators scored at the episode level (e.g., follow-up after hospitalization for an exacerbation of asthma), we scored every event rendering the participant eligible for the specified process and involving any of the participant’s providers, so the number of eligibility events depended on the number of episodes that occurred.

In order to produce aggregate scores, we divided all instances in which recommended care was delivered by the number of times participants were eligible for indicators in the category. For example, Table 1 presents information about seven of the indicators for acute care; the number of times participants were eligible for these indicators would constitute the denominator for the acute care score. The results are presented as proportions, theoretically ranging in value from 0 to 100 percent. We used the bootstrap method to estimate standard errors directly for all the aggregate scores.³⁸

Because everyone in the initial sample for the CQI study had participated in the CTS, we had a rich set of variables for assessing nonresponse. We used logistic-regression analysis to estimate the relations between individual characteristics (age, sex, race, educational level, income, self-reported level of use of physicians and hospitals, insurance status, and

health status) and participation in the study. In general, participants tended to be older than nonparticipants ($P<0.001$) and were more likely than nonparticipants to be female ($P<0.001$) and white ($P<0.001$), with higher levels of education ($P<0.001$) and income ($P<0.001$). They were also more likely to have used health care services ($P<0.001$) and to be in other than excellent health ($P=0.03$). We used the coefficients from the regression equation to adjust the scores for nonresponse, and we weighted the data for the participants to be representative of the population from which they were drawn.

RESULTS

CHARACTERISTICS OF THE PARTICIPANTS

Table 2 summarizes the characteristics of the participants; these characteristics differ from population averages but parallel the profile of persons receiving medical care. For example, the average age of patients in the National Ambulatory Medical Care Survey³⁹ is 44.7 years. Women have higher rates of visits than men (319.9 vs. 234.9 visits per 100 persons per year), and whites have higher rates of visits than blacks (293.2 vs. 210.7 visits per 100 persons per year).³⁹ Participants were well educated. Forty-three percent had one or more of the chronic conditions we assessed, and 34 percent had one or more of the acute conditions. Preventive care was assessed for all participants; in addition, participants’ care was assessed for 1.5 chronic or acute conditions, on average, for a total of 2.5 (range, 1 to 13). Participants were included in the overall denominator an average of 16 times (range, 2 to 304).

ANALYSIS OF CARE DELIVERED

Tables 3, 4, and 5 show the number of indicators included in the aggregate score, the number of persons eligible for one or more processes within the category, the number of times participants in the sample were eligible for indicators, and the weighted mean proportion (and 95 percent confidence interval) of recommended processes that were delivered.

Overall, participants received 54.9 percent of recommended care (95 percent confidence interval, 54.3 to 55.5) (Table 3). This level of performance was similar in the areas of preventive care, acute care, and care for chronic conditions. The level of performance according to the particular medical function ranged from 52.2 percent (95 percent confidence interval, 51.3 to 53.2) for screening to 58.5 percent (95 percent confidence interval, 56.6 to 60.4) for follow-up care.

“Mode” refers to the mechanism of care delivery required for the provision of the indicated process. Analysis of performance in terms of mode may identify areas in which system-wide interventions could offer solutions to problems of quality, such as improved methods for ordering, processing, and communicating laboratory results. We found greater variation among modes than among functions in adherence to the processes we studied (Table 4). Care requiring an encounter or other intervention (e.g., the annual visit recommended for patients with hypertension) had the highest rates of adherence (73.4 percent [95 percent confidence interval, 71.5 to 75.3]), and processes involving counseling or education (e.g., advising smokers with chronic obstructive pulmonary disease to quit smoking) had the lowest rates of adherence (18.3 percent [95 percent confidence interval, 16.7 to 20.0]). All pairwise differences were statistically significant at $P < 0.001$ except those between the prescribing of medication and care requiring an encounter or other intervention ($P = 0.02$), physical examination and immunization ($P = 0.001$), surgery and immunization ($P = 0.004$), and surgery and physical examination ($P = 0.05$). The difference between surgery and laboratory testing or radiography was not significant ($P = 0.39$).

PROBLEMS WITH QUALITY OF CARE

We also classified indicators according to the problem with quality that was deemed most likely to occur, and we found greater problems with underuse (46.3 percent of participants did not receive recom-

Table 2. Characteristics of the 6712 Participants.*

| Characteristic | Value |
|--|------------|
| Age (yr) | |
| Mean | 45.5±0.2 |
| Range | 18–97 |
| Female sex (%) | 59.6±0.006 |
| Nonwhite race (%) | 18.6±0.005 |
| Education (yr) | 13.7±0.03 |
| ≥1 Chronic conditions (%) | 44.7±0.006 |
| ≥1 Acute conditions (%) | 36.3±0.006 |
| No. of conditions and preventive care for which participants were eligible | |
| Mean | 2.5±0.02 |
| Range | 1–13 |
| No. of times participants eligible for indicators† | |
| Mean | 15.8±0.17 |
| Range | 2–304 |

* Plus-minus values are means or percentages ±SE.

† The number of times a participant is eligible for an indicator is a function of the level at which the indicator is scored (participant, participant-provider dyad, or episode), the number of participants eligible for the specified process, and the number of indicators in the aggregate-score category.

mended care [95 percent confidence interval, 45.8 to 46.8]) than with overuse (11.3 percent of participants received care that was not recommended and was potentially harmful [95 percent confidence interval, 10.2 to 12.4]).

VARIATIONS IN QUALITY

Table 5 shows substantial variability in the quality-of-care scores for the 25 conditions for which at least 100 persons were eligible for analysis. Persons with senile cataracts received 78.7 percent of the recommended care (95 percent confidence interval, 73.3 to 84.2); persons with alcohol dependence received 10.5 percent of the recommended care (95 percent confidence interval, 6.8 to 14.6). The aggregate scores for individual conditions were generally not sensitive to the presence or absence of any single indicator of quality.

DISCUSSION

Overall, participants received about half of the recommended processes involved in care. These defi-

Table 3. Adherence to Quality Indicators, Overall and According to Type of Care and Function.

| Variable | No. of Indicators | No. of Participants Eligible | Total No. of Times Indicator Eligibility Was Met | Percentage of Recommended Care Received (95% CI)* |
|--------------|-------------------|------------------------------|--|---|
| Overall care | 439 | 6712 | 98,649 | 54.9 (54.3–55.5) |
| Type of care | | | | |
| Preventive | 38 | 6711 | 55,268 | 54.9 (54.2–55.6) |
| Acute | 153 | 2318 | 19,815 | 53.5 (52.0–55.0) |
| Chronic | 248 | 3387 | 23,566 | 56.1 (55.0–57.3) |
| Function | | | | |
| Screening | 41 | 6711 | 39,486 | 52.2 (51.3–53.2) |
| Diagnosis | 178 | 6217 | 29,679 | 55.7 (54.5–56.8) |
| Treatment | 173 | 6707 | 23,019 | 57.5 (56.5–58.4) |
| Follow-up | 47 | 2413 | 6,465 | 58.5 (56.6–60.4) |

* CI denotes confidence interval.

Table 4. Adherence to Quality Indicators, According to Mode.

| Mode | No. of Indicators | No. of Participants Eligible | Total No. of Times Indicator Eligibility Was Met | Percentage of Recommended Care Received (95% CI)* |
|-----------------------------------|-------------------|------------------------------|--|---|
| Encounter or other intervention | 30 | 2843 | 4,329 | 73.4 (71.5–75.3) |
| Medication | 95 | 2964 | 8,389 | 68.6 (67.0–70.3) |
| Immunization | 8 | 6700 | 9,748 | 65.7 (64.3–67.0) |
| Physical examination | 67 | 6217 | 19,428 | 62.9 (61.8–64.0) |
| Laboratory testing or radiography | 131 | 5352 | 18,605 | 61.7 (60.4–63.0) |
| Surgery | 21 | 244 | 312 | 56.9 (51.3–62.5) |
| History | 64 | 6711 | 36,032 | 43.4 (42.4–44.3) |
| Counseling or education | 23 | 2838 | 3,806 | 18.3 (16.7–20.0) |

* CI denotes confidence interval. All pairwise differences were statistically significant at $P < 0.001$ except those between medication and encounter or other intervention ($P = 0.02$), physical examination and immunization ($P = 0.001$), surgery and immunization ($P = 0.004$), and surgery and physical examination ($P = 0.05$). The difference between surgery and laboratory testing or radiography was not significant ($P = 0.39$).

cits in care have important implications for the health of the American public. For example, only 24 percent of participants in our study who had diabetes received three or more glycosylated hemoglobin tests over a two-year period. This finding parallels the finding by Saaddine and colleagues that 29 percent of adults with diabetes who participated in the nationally representative Behavioral Risk Factor Surveillance System reported having their blood sugar tested during the previous year.¹⁸ This routine monitoring is essential to the assessment of the effectiveness of treatment, to ensuring appropriate responses to poor glycemic control, and to the identification of complications of the disease at an early stage so that serious consequences may be prevented. In the United Kingdom Prospective Diabetes Study, tight blood glucose control and biannual monitoring decreased the risk of microvascular complications by 25 percent.⁴⁰

In our study, persons with hypertension received 64.7 percent of the recommended care (95 percent confidence interval, 62.6 to 66.7). We have previously demonstrated a link between blood-pressure control and adherence to process-related measures of quality of care for hypertension.⁴¹ Persons whose blood pressure is persistently above normal are at increased risk for heart disease, stroke, and death.⁴² Poor blood-pressure control contributes to more than 68,000 preventable deaths annually.⁴³

Overall, 68.0 percent (95 percent confidence interval, 64.2 to 71.8) of the recommended care for coronary artery disease was received, but only 45 percent of persons presenting with a myocardial infarction received beta-blockers, which reduce the risk of death by 13 percent during the first week of treatment and by 23 percent over the long term.⁴⁴ Only 61 percent of participants with a myocardial infarction who were appropriate candidates for aspirin therapy received aspirin, which has been shown in randomized trials to reduce the risk of death from vascular causes by 15 percent, to reduce the risk of nonfatal myocardial infarction by 30 percent, and to reduce the risk of nonfatal stroke by 40 percent.⁴⁵

Deficits in processes involved in primary and secondary preventive care are also associated with preventable deaths. Among elderly participants, only 64 percent had received or been offered a pneumococcal vaccine; nearly 10,000 deaths from pneumonia could be prevented annually by appropriate

vaccinations.⁴³ About 38 percent of participants had been screened for colorectal cancer; annual fecal occult-blood tests could prevent about 9600 deaths annually.⁴³

Nonresponse bias is a potential limitation of the study. Because the sample we analyzed included 37 percent of the eligible adults, the results are likely to be biased, but the direction of that bias is not clear. For example, because our participants were more likely to use the health care system than were eligible persons who did not participate in the study, our results may be biased toward an underestimation of deficits in quality related to underuse.

The study relied primarily on the review of medical records to score indicators, which may lead some to conclude that we have identified problems with documentation rather than quality. This issue has been examined in studies that compared process-based quality scores using standardized patients, vignettes, and abstraction of medical records⁴⁶ and studies that compared standardized patients with audiotapes of encounters.⁴⁷ Overall, the process scores among the four conditions studied were 5 percentage points lower with the use of medical records than with the use of vignettes and 10 percentage points lower with the use of medical records than with the use of standardized patients. About two thirds of the disagreement between data from standardized patients and data from audiotapes was attributable to reports by standardized patients that they received care processes that were not confirmed by audiotape. A related study reported a false positive rate of 6.4 percent in medical-record documentation, with the highest false positive rates found for physical examination and elements of the diagnostic process.⁴⁸ Thus, our scores might have been as much as 10 percentage points higher if we had used a different method of obtaining data. We used the interview about the participant's health history to partially offset this effect. For example, among elderly participants, only 15 percent had a note in any chart indicating that an influenza vaccination had been received, but 85 percent reported having received one. In general, the inclusion of self-reported data improved scores.

Our results indicate that, on average, Americans receive about half of recommended medical care processes. Although this point estimate of the size of the quality problem may continue to be debated, the gap between what we know works and what is

Table 5. Adherence to Quality Indicators, According to Condition.*

| Condition | No. of Indicators | No. of Participants Eligible | Total No. of Times Indicator Eligibility Was Met | Percentage of Recommended Care Received (95% CI) |
|--|-------------------|------------------------------|--|--|
| Senile cataract | 10 | 159 | 602 | 78.7 (73.3–84.2) |
| Breast cancer | 9 | 192 | 202 | 75.7 (69.9–81.4) |
| Prenatal care | 39 | 134 | 2920 | 73.0 (69.5–76.6) |
| Low back pain | 6 | 489 | 3391 | 68.5 (66.4–70.5) |
| Coronary artery disease | 37 | 410 | 2083 | 68.0 (64.2–71.8) |
| Hypertension | 27 | 1973 | 6643 | 64.7 (62.6–66.7) |
| Congestive heart failure | 36 | 104 | 1438 | 63.9 (55.4–72.4) |
| Cerebrovascular disease | 10 | 101 | 210 | 59.1 (49.7–68.4) |
| Chronic obstructive pulmonary disease | 20 | 169 | 1340 | 58.0 (51.7–64.4) |
| Depression | 14 | 770 | 3011 | 57.7 (55.2–60.2) |
| Orthopedic conditions | 10 | 302 | 590 | 57.2 (50.8–63.7) |
| Osteoarthritis | 3 | 598 | 648 | 57.3 (53.9–60.7) |
| Colorectal cancer | 12 | 231 | 329 | 53.9 (47.5–60.4) |
| Asthma | 25 | 260 | 2332 | 53.5 (50.0–57.0) |
| Benign prostatic hyperplasia | 5 | 138 | 147 | 53.0 (43.6–62.5) |
| Hyperlipidemia | 7 | 519 | 643 | 48.6 (44.1–53.2) |
| Diabetes mellitus | 13 | 488 | 2952 | 45.4 (42.7–48.3) |
| Headache | 21 | 712 | 8125 | 45.2 (43.1–47.2) |
| Urinary tract infection | 13 | 459 | 1216 | 40.7 (37.3–44.1) |
| Community-acquired pneumonia | 5 | 144 | 291 | 39.0 (32.1–45.8) |
| Sexually transmitted diseases or vaginitis | 26 | 410 | 2146 | 36.7 (33.8–39.6) |
| Dyspepsia and peptic ulcer disease | 8 | 278 | 287 | 32.7 (26.4–39.1) |
| Atrial fibrillation | 10 | 100 | 407 | 24.7 (18.4–30.9) |
| Hip fracture | 9 | 110 | 167 | 22.8 (6.2–39.5) |
| Alcohol dependence | 5 | 280 | 1036 | 10.5 (6.8–14.6) |

* Condition-specific scores are not reported for management of pain due to cancer and its palliation, management of symptoms of menopause, hysterectomy, prostate cancer, and cesarean section, because fewer than 100 people were eligible for analysis of these categories. CI denotes confidence interval.

actually done is substantial enough to warrant attention. These deficits, which pose serious threats to the health and well-being of the U.S. public, persist despite initiatives by both the federal government and private health care delivery systems to improve care.

What can we do to break through this impasse? Given the complexity and diversity of the health care system, there will be no simple solution. A key component of any solution, however, is the routine availability of information on performance at all levels. Making such information available will require a major overhaul of our current health information systems, with a focus on automating the entry and retrieval of key data for clinical decision making and for the measurement and reporting of quality.⁴⁹

Establishing a national base line for performance makes it possible to assess the effect of policy changes and to evaluate large-scale national, regional, state, or local efforts to improve quality.

Supported by the Robert Wood Johnson Foundation and by career development awards (to Drs. Asch and Kerr) from the Veterans Affairs Health Services Research and Development program.

We are indebted to Maureen Michael, James Knickman, and Robert Hughes at the Robert Wood Johnson Foundation for their support; to Paul Ginsburg at the Center for Studying Health System Change for his support of this collaboration; to Richard Strauss at Mathematica Policy Research for developing systems for passing the initial sample from the Community Tracking Study household survey to RAND for this study; to RAND's Survey Research Group (Josephine Levy and Loral Hill) and the telephone interviewers for recruiting participants; to Peggy Wallace, Karen Ricci, and Belle Griffin for their assistance in the design of the data-collection tool, for hiring and training the nurse abstractors, and for overseeing the data-collection process; to Liisa Hiatt for serving as the project manager; and to Vector Research for developing the data-collection software.

REFERENCES

1. McGlynn EA, Brook RH. Keeping quality on the policy agenda. *Health Aff (Millwood)* 2001;20(3):82-90.
2. Institute of Medicine. *Crossing the quality chasm: a new health system for the 21st century*. Washington, D.C.: National Academy Press, 2001.
3. Schuster MA, McGlynn EA, Brook RH. How good is the quality of health care in the United States? *Milbank Q* 1998;76:517-63.
4. Miller RH, Luft HS. Managed care plan performance since 1980: a literature analysis. *JAMA* 1994;271:1512-9.
5. Jencks SM, Cuedon T, Burwen DR, et al. Quality of medical care delivered to Medicare beneficiaries: a profile at state and national levels. *JAMA* 2000;284:1670-6.
6. Ellerbeck EF, Jencks SF, Radford MJ, et al. Quality of care for Medicare patients with acute myocardial infarction: a four-state pilot study from the Cooperative Cardiovascular Project. *JAMA* 1995;273:1509-14.
7. Murata PJ, McGlynn EA, Siu AL, et al. Quality measures for prenatal care: a comparison of care in six health care plans. *Arch Fam Med* 1994;3:41-9.
8. Krumholz HM, Radford MJ, Ellerbeck EF, et al. Aspirin in the treatment of acute myocardial infarction in elderly Medicare beneficiaries: patterns of use and outcomes. *Circulation* 1995;92:2841-7.
9. Brechner RJ, Cowie CC, Howie LJ, Herman WH, Will JC, Harris MI. Ophthalmic examination among adults with diagnosed diabetes mellitus. *JAMA* 1993;270:1714-8.
10. Starfield B, Powe NR, Weiner JR, et al. Costs vs quality in different types of primary care settings. *JAMA* 1994;272:1903-8.
11. Payne SM, Donahue C, Rappo P, et al. Variations in pediatric pneumonia and bronchitis/asthma admission rates: is appropriateness a factor? *Arch Pediatr Adolesc Med* 1995;149:162-9.
12. Udvarhelyi IS, Jennison K, Phillips RS, Epstein AM. Comparison of the quality of ambulatory care for fee-for-service and prepaid patients. *Ann Intern Med* 1991;115:394-400.
13. Jencks SF, Huff ED, Cuedon T. Change in the quality of care delivered to Medicare beneficiaries, 1998-1999 to 2000-2001. *JAMA* 2003;289:305-12.
14. Asch SM, Sloss EM, Hogan C, Brook RH, Kravitz RL. Measuring underuse of necessary care among elderly Medicare beneficiaries using inpatient and outpatient claims. *JAMA* 2000;284:2325-33.
15. Kahn KL, Keeler EB, Sherwood MJ, et al. Comparing outcomes of care before and after implementation of the DRG-based prospective payment system. *JAMA* 1990;264:1984-8.
16. The state of health care quality, 2002. Washington, D.C.: National Committee for Quality Assurance. (Accessed May 30, 2003, at <http://www.ncqa.org/sohc2002/>.)
17. Nelson DE, Bland S, Powell-Griner E, et al. State trends in health risk factors and receipt of clinical preventive services among US adults during the 1990s. *JAMA* 2002;287:2659-67.
18. Saaddine JB, Engelgau MM, Beckles GL, Gregg EW, Thompson TJ, Narayan KM. A diabetes report card for the United States: quality of care in the 1990s. *Ann Intern Med* 2002;136:565-74.
19. Asch SM, Gifford AL, Bozzette SA, et al. Underuse of primary Mycobacterium avium complex and Pneumocystis carinii prophylaxis in the United States. *J Acquir Immune Defic Syndr* 2001;28:340-4.
20. Hyman DJ, Pavlik VN. Characteristics of patients with uncontrolled hypertension in the United States. *N Engl J Med* 2001;345:479-86. [Erratum, *N Engl J Med* 2002;346:544.]
21. Kemper PD, Blumenthal D, Corrigan JM, et al. The design of the Community Tracking Study: a longitudinal study of health system change and its effects on people. *Inquiry* 1996;33:195-206.
22. Strunk BC, Ginsburg PB, Gabel JR. Tracking health care costs: hospital care surpasses drugs as the key cost driver. *Bethesda, Md.: Health Affairs*, September 2001. (Accessed May 30, 2003, at http://www.healthaffairs.org/1110_web_exclusives.php.)
23. Cunningham PJ. Declining employer-sponsored coverage: the role of public programs and implications for access to care. *Med Care Res Rev* 2002;59:79-98.
24. Draper DA, Hurley RE, Lesser CS, Strunk BC. The changing face of managed care. *Health Aff (Millwood)* 2002;21(1):11-23.
25. McGlynn EA, Kerr EA, Asch SM. New approach to assessing the clinical quality of care for women: the QA Tool system. *Women's Health Issues* 1999;9:184-92.
26. McGlynn EA. Choosing and evaluating clinical performance measures. *Jt Comm J Qual Improv* 1998;24:470-9.
27. Brook RH. The RAND/UCLA appropriateness method. In: McCormick KA, Moore SR, Siegel RA, eds. *Clinical practice guideline development: methodology perspectives*. Rockville, Md.: Agency for Health Care Policy and Research, November 1994:59-70. (AHCPR publication no. 95-0009).
28. Shekelle PG, Kahan JP, Bernstein SJ, Leape LL, Kamberg CJ, Park RE. The reproducibility of a method to identify the overuse and underuse of medical procedures. *N Engl J Med* 1998;338:1888-95.
29. Shekelle PG, Chassin MR, Park RE. Assessing the predictive validity of the RAND/UCLA appropriateness method criteria for performing carotid endarterectomy. *Int J Technol Assess Health Care* 1998;14:707-27.
30. Kravitz RL, Park RE, Kahan JP. Measuring the clinical consistency of panelists' appropriateness ratings: the case of coronary

- artery bypass surgery. *Health Policy* 1997; 42:135-43.
31. Selby JV, Fireman BH, Lundstrom RJ, et al. Variation among hospitals in coronary angiography practices and outcomes after myocardial infarction in a large health maintenance organization. *N Engl J Med* 1996; 335:1888-96.
32. Hemingway H, Crook AM, Feder G, et al. Underuse of coronary revascularization procedures in patients considered appropriate candidates for revascularization. *N Engl J Med* 2001;344:645-54.
33. McGlynn EA, Kerr EA, Damberg CL, Asch SM, eds. *Quality of care for women: a review of selected clinical conditions and quality indicators*. Santa Monica, Calif.: RAND, 2000.
34. Kerr EA, Asch SM, Hamilton EG, McGlynn EA, eds. *Quality of care for cardiopulmonary conditions: a review of the literature and quality indicators*. Santa Monica, Calif.: RAND, 2000.
35. *Idem*. *Quality of care for general medical conditions: a review of the literature and quality indicators*. Santa Monica, Calif.: RAND, 2000.
36. Asch SM, Kerr EA, Hamilton EG, Reifel JL, McGlynn EA, eds. *Quality of care for oncologic conditions and HIV: a review of the literature and quality indicators*. Santa Monica, Calif.: RAND, 2000.
37. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
38. Efron B, Tibshirani R. *An introduction to the bootstrap*. New York: Chapman & Hall, 1993.
39. Cherry DK, Burt CW, Woodwell DA. National ambulatory medical care survey: 1999 summary. Advance data from vital and health statistics. No. 322. Hyattsville, Md.: National Center for Health Statistics, July 2001. (DHHS publication no. (PHS) 2001-1250 01-0383.)
40. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998;352:837-53. [Erratum, *Lancet* 1999;354:602.]
41. Asch SM, Kerr EA, Lapuerta P, Law A, McGlynn EA. A new approach for measuring quality of care for women with hypertension. *Arch Intern Med* 2001;161:1329-35.
42. Collins R, Peto R, MacMahon S, et al. Blood pressure, stroke, and coronary heart disease. 2. Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* 1990;335:827-38.
43. Woolf SH. The need for perspective in evidence-based medicine. *JAMA* 1999;282: 2358-65.
44. Hennekens CH, Albert CM, Godfried SL, Gaziano JM, Buring JE. Adjunctive drug therapy of acute myocardial infarction — evidence from clinical trials. *N Engl J Med* 1996;335:1660-7.
45. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatments for myocardial infarction. *JAMA* 1992;268:240-8.
46. 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.
47. Luck J, Peabody JW. Using standardised patients to measure physicians' practice: validation study using audio recordings. *BMJ* 2002;325:679.
48. Dresselhaus TR, Luck J, Peabody JW. The ethical problem of false positives: a prospective evaluation of physician reporting in the medical record. *J Med Ethics* 2002;28: 291-4.
49. Berwick DM, James B, Coye MJ. Connections between quality measurement and reporting. *Med Care* 2003;41:Suppl:I-30-I-38.

Copyright © 2003 Massachusetts Medical Society.

PERSONAL ARCHIVES IN THE JOURNAL ONLINE

Individual subscribers can store articles and searches using a new feature on the Journal's Web site (www.nejm.org) called "Personal Archive." Each article and search result links to this feature. Users can create personal folders and move articles into them for convenient retrieval later.