

## Correlations among measures of quality in HIV care in the United States: cross sectional study

Ira B Wilson, professor,<sup>1</sup> Bruce E Landon, associate professor of health care policy and medicine,<sup>2</sup> Peter V Marsden, professor,<sup>3</sup> Lisa R Hirschhorn, assistant professor of medicine,<sup>4</sup> Keith McInnes, lecturer,<sup>2</sup> Lin Ding, biostatistician,<sup>2</sup> Paul D Cleary, professor and dean<sup>5</sup>

<sup>1</sup>Department of Medicine, Institute for Clinical Research and Health Care Policy, Tufts-New England Medical Center, Boston, MA 02111, USA

<sup>2</sup>Health Care Policy and Medicine, Department of Health Care Policy, Harvard Medical School, Boston, MA 02115

<sup>3</sup>Department of Sociology, 630 William James Hall, Cambridge, MA 02138

<sup>4</sup>Harvard Medical School Division of AIDS, Landmark Center 2 East, Boston, MA, 02215

<sup>5</sup>School of Public Health, Yale University, New Haven, CT 06520-8034

Correspondence to: I B Wilson  
[iwilson@tufts-nemc.org](mailto:iwilson@tufts-nemc.org)

doi:10.1136/bmj.39364.520278.55

### ABSTRACT

**Objective** To determine whether a selected set of indicators can represent a single overall quality construct.

**Design** Cross sectional study of data abstracted during an evaluation of an initiative to improve quality of care for people with HIV.

**Setting** 69 sites in 30 states.

**Data sources** Medical records of 9020 patients.

**Main outcome measures** Adjusted performance rates at site level for eight measures of quality of care specific to HIV and a site level summary performance score (the number of measures for which the site was in the top quarter of the distribution).

**Results** Of 28 site level correlations between measures, two were greater than 0.40, two were between 0.30 and 0.39, four were between 0.20 and 0.29, and the 20 remaining were all less than 0.20. One site was in the top quarter for seven measures, but no sites were in the top quarter for six or eight of the measures. Across the eight quality measures, sites were in the top quarter no more often than predicted by a chance (binomial) distribution.

**Conclusions** The quality suggested by one measured indicator cannot necessarily be generalised to unmeasured indicators, even if this might be expected for clinical or other reasons.

### INTRODUCTION

Efforts to measure and report the quality of care delivered by healthcare organisations are becoming commonplace. Publicly reported performance data are increasingly available for health plans, hospitals, nursing homes, and groups of physicians, and many providers are now being rewarded on the basis of measures of quality of care.<sup>1-3</sup> These initiatives generally rely on a small set of measures, usually of processes of care but sometimes of outcomes. Common indicators of performance of health plans and physicians focus on the provision of preventive services and the management of a small number of chronic conditions, such as diabetes and asthma.

In addition to making reported quality data more comprehensible, a rationale for using a small subset of possible quality indicators is the belief that an organisation's performance on unmeasured processes

or outcomes will be similar to that on measured ones. For example, although there are many activities involved in high quality care in diabetes, it is assumed that assessment of selected processes, such as whether a yearly retinal exam is performed or whether a test for haemoglobin A<sub>1c</sub> was ordered, provides a reasonable indication of the overall quality of an organisation's diabetes care. An extension of this logic is that monitoring care indicators for a carefully selected set of prevalent and important conditions, such as diabetes, hypertension, and heart attacks, provides valid information about the overall quality of care provided by a physician, medical group, health plan, or hospital. The use of a few indicators to assess care is consonant with systems theory, which implies that there should be relatively high correlations among quality indicators within organisations because multiple areas of performance should be influenced by common characteristics of the system.<sup>4-6</sup>

Several studies have examined the relations among quality measures for various different types of organisations, but few of these studies examined outpatient medical practices. For instance, a recent study of 11 outpatient practices that assessed measures of technical clinical quality (such as cholesterol screening), patients' satisfaction, clinic function (such as follow-up of abnormal results of laboratory tests), and compliance with treatment for diabetes and asthma found no significant correlations between these measures.<sup>7</sup> Similarly, Palmer et al found that correlations across cases seen by a given physician were low.<sup>8</sup> Other studies that examined hospitals,<sup>9,10</sup> health plans,<sup>11</sup> and communities<sup>12</sup> have found similarly low correlations among quality measures.

Given the large number of initiatives for measurement and improvement of quality, many of them founded on systems theory, it is surprising that so few have reported empirical assessments of the relations among quality indicators. Examining such correlations is critical for both measurement and improvement of quality. With regard to measurement, it is important to understand whether it is appropriate to draw conclusions about the overall quality on the basis of a limited set of indicators. With regard to improvement, finding strong correlations among quality measures would

support the theory that the measures are the output of a single functional system and that efforts to improve quality should focus on characteristics of the system. Low correlations, by contrast, would suggest that multiple functionally independent systems are operating, implying that efforts to improve quality need to address these distinct systems or their integration.

We examined the relation among eight quality indicators for a single chronic medical condition in

care settings in which we expected relatively high correlations—that is, organisations that deliver HIV care to outpatients. We used data from a quality improvement initiative that included HIV care sites in 30 states to see whether we could identify organisations that were “high performers”—that is, organisations that consistently scored highly across different quality measures. High performing organisations, presumably, are organised and managed in ways that allow them to achieve high quality.

**Table 1** | Characteristics of patients and sites

Characteristic	No (%) or mean (SD)
<b>Patients (n=9020)</b>	
Mean (SD) age (years)	40.5 (8.9)
Women	2859 (31.7)
Psychiatric diagnosis	2832 (31.4)
Active substance abuse*	1470 (16.3)
Lowest CD4 cell count ( $\times 10^6/l$ ) during review period:	
0-49	929 (10.3)
50-199	1912 (21.2)
200-499	3870 (42.9)
>500	2309 (25.6)
<b>Sites (n=69)</b>	
Region:	
Midwest	14 (20)
Northeast	26 (38)
South	20 (29)
West	9 (13)
Type of clinic:	
Community based organisation	11 (16)
Community health centre	23 (33)
Health department	11 (16)
Hospital	10 (15)
University medical centre	14 (20)
Mean (SD) No of full time equivalent physicians	5.6 (5.6)
Mean (SD) No of full time equivalent nurse practitioners and physician assistants per site	2.7 (2.8)
Mean (SD) No of health professionals on staff:	
General physicians	4.9 (6.4)
Obstetrics and gynaecology	0.8 (1.7)
Fellowship trained specialists in infectious diseases	1.8 (3.4)
Mental health professionals	2.6 (4.6)
Nurses	7.2 (7.1)
Professionals routinely available on site:	
Dentist	36 (52)
Nutritionist	57 (83)
Social worker	60 (87)
Substance abuse counsellor	35 (50)
HIV specialisation:	
Specialised HIV clinic	43 (62)
General medicine with specialised HIV team	24 (35)
General medicine with no specialised HIV team	2 (3)
Mean (SD) % of budget from Ryan White funds	48.2 (34.5)
Mean (SD) % of HIV patients with case manager	70.1 (35.4)
Multidisciplinary HIV care team	55 (80)
Mean (SD) No of team meetings in past 12 months	23.5 (18.6)

\*Substance abuse noted in medical record during review period.

## METHODS

### Overview

Data for this study were collected as part of an evaluation of a quality improvement collaborative involving clinics that received funds through the Ryan Care Act.<sup>13,14</sup> We abstracted information on quality of care from medical records at two times (details below).

### Participants

**Site selection**—Of the 200 relevant sites in the United States in May 2000, 171 were eligible to participate in the study. From these, we enrolled 44 sites that were participating in the quality improvement intervention and an additional 25 sites that served as controls, giving 69 participating sites in 30 states. Details of the site selection process are described elsewhere.<sup>13,14</sup> We previously reported that changes in quality measures did not differ significantly between intervention and control sites.<sup>13</sup> We surveyed medical directors at each site to determine specific characteristics.

**Patient selection**—We randomly sampled 75 active patients from each site before the intervention and then drew a second random sample of 75 after the intervention. The intervention took place from 30 June 2000 to 31 December 2001. Patients were considered active if they visited the site at least once during the review period.

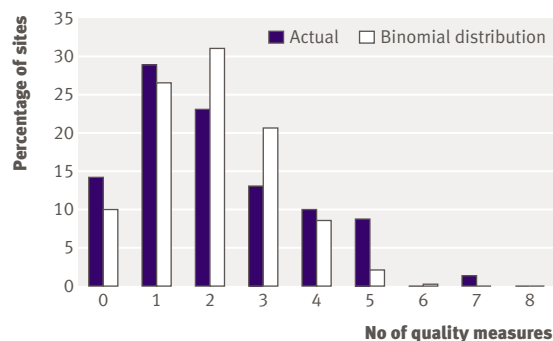
### Data collection

We collected data from the medical records of each sampled patient over one year of care for the two review periods. Data abstracted included age, sex, history of HIV related illnesses, comorbid medical or psychiatric conditions including current substance abuse or psychiatric illness, screening and prophylaxis for HIV related conditions, number and timing of visits, CD4 cell counts, viral loads, and antiretroviral medications. Reviewers specified whether each visit was to a physician, a nurse practitioner or physician assistant, a nurse, or some “other” clinician (such as a nutritionist). The first review covered the year before the intervention (1 June 1999 to 31 May 2000), and the second covered the year beginning six months after the start of the intervention and ending three months after the end of the intervention (1 January 2001 to 31 December 2001).

**Table 2** | Measures of quality (proportions, adjusted for patients' characteristics) at the 69 sites studied

Quality measure	Mean (range)	Median (IQR*)
HAART use	0.81 (0.57-0.93)	0.81 (0.77-0.86)
Non-detectable HIV viral load	0.38 (0.01-0.60)	0.41 (0.30-0.48)
<i>P carinii</i> pneumonia prophylaxis	0.70 (0.25-1.00)	0.75 (0.61-0.84)
Tuberculosis screening	0.52 (0.06-0.91)	0.53 (0.35-0.69)
Hepatitis C screening	0.81 (0.25-1.00)	0.87 (0.73-0.93)
Cervical cancer screening	0.60 (0.27-0.98)	0.62 (0.50-0.71)
Influenza vaccination	0.51 (0.02-0.82)	0.53 (0.45-0.61)
Visits in three quarters	0.67 (0.47-0.82)	0.68 (0.62-0.75)

\*Interquartile range.



Percentage of sites performing in top quarter for specified number of quality measures (from 0 to 8); comparison of actual with binomial distribution

### Quality of care measures

The eight measures of quality of care were based on guidelines that did not change over this time period.<sup>15-18</sup> Our primary measures were proportion of use of highly active antiretroviral therapy (HAART) at the time of the last visit during the review period and control of HIV viral load for appropriate patients. Patients included in the denominator for the proportion of HAART use were those with CD4 cell counts less than  $500 \times 10^6/l$  or viral loads greater than 20 000 copies/ml, and patients already receiving HAART, as per guidelines in effect at the time.<sup>15</sup> Viral load was considered as controlled if it was undetectable or if the total viral load was less than 400 copies/ml. We also assessed whether screening for tuberculosis, hepatitis C, and cervical cancer (for women only), appropriate prophylaxis against *Pneumocystis carinii* pneumonia, and influenza vaccinations were provided during the review period. For hepatitis C, we accepted documentation of a previous positive result of a hepatitis C test. We defined appropriate access to outpatient care as actually visiting the site during at least three of four quarters. All measures at the patient level were dichotomous.

### Analyses

Our unit of analysis was the care site. The site level value for each measure of quality was the proportion of patients for whom the quality indicator was documented in the reviewed medical records. Our goal was to review 75 records at each of 69 sites to give a total of 10 350 records, though the final number of records reviewed was 9986 (97%). We initially analysed the first and second review periods separately but because results were similar we aggregated data from the two periods. In 966 cases, a patient's medical record was selected for review in both periods. For these cases we dropped the data from the second period, leaving a total of 9020 unique patients in care at 69 sites. We present descriptive statistics for the 69 sites. Because characteristics of patients vary among sites, we examined adjusted means for each of the eight quality measures. We used the GLIMMIX

macro in SAS (version 8.2) to produce least square means for sites using the logit link function. We adjusted for patients' characteristics that might be related to the quality measures, including age, sex, stage of disease based on lowest recorded CD4 cell count over the period of care, active psychiatric or substance abuse problems, history of HIV related diagnoses, number of comorbid medical conditions, and review period. Adjusted least square means were then converted to proportions.

Next, we calculated the correlations among the eight adjusted quality measures. Finally, we examined the degree to which high performance on one indicator was related to high performance on the other seven. To do this, we dichotomised each quality measure at the 75th centile of the distribution across all 69 sites, and called those sites in the top quarter "high performers." When the top fifth was used, similar results were obtained. We then examined the distribution of the number of times sites performed in the top quarter across the eight quality measures and compared it to a binomial distribution for eight independent trials with a probability of success (being a high performer) of 0.25. If the actual distribution differs from the binomial distribution, there are more high performers than expected by chance; if the two distributions are similar, this suggests that site's performance on one quality measure is independent of its performance on others. The distributions were compared with  $\chi^2$  test.

## RESULTS

**Characteristics of patients and sites**—Thirty two per cent of patients were female, 16% reported active substance abuse, and 32% had CD4 cell counts below  $200 \times 10^6/l$  (table 1). The 69 sites that we studied were in 30 states in all regions of the US, representing the full spectrum of types of organisation that provide HIV care. Most sites described themselves as having HIV expertise, including 62% that were specialised HIV clinics and 35% that were general medicine clinics with specialised HIV care teams. Only 3% were general medicine clinics with no specialised HIV team. Most (80%) had multidisciplinary HIV care teams that met about twice a month.

**Site level quality measures**—Clinic performance on the quality measures ranged from 0.38 of patients with non-detectable viral loads (table 2) to 0.81 of eligible

patients on HAART and 0.81 of patients with documented hepatitis C status. The greatest variation across clinics was for the proportion of patients who received tuberculosis screening (interquartile range of 0.035-0.69), and the least variation was seen for the proportion of eligible patients who received HAART (interquartile range of 0.77-0.86).

**Correlations among quality measures**—Of the 28 correlations between measures at the clinic level (table 3), the highest was the relation between proportions of HAART therapy and *P carinii* pneumonia prophylaxis at 0.42 ( $P<0.001$ ). The correlation between the proportion receiving cervical cancer screening and tuberculosis screening was nearly as high at 0.40 ( $P<0.001$ ). Two other correlations were greater than 0.30, those between the proportion receiving hepatitis C screening and tuberculosis screening (0.32,  $P<0.01$ ), and between influenza vaccination and non-detectable viral loads (0.30,  $P<0.05$ ). Four additional correlations were between 0.20 and 0.29, and the 20 remaining correlations were all less than 0.20.

**Distribution of number of high performance areas**—The number of times sites were in the top quarter (a "high performer") for the eight quality measures ranged from none (never in the top quarter) to seven (in the top quarter for all but one measure). The figure shows the actual and expected distribution under an assumption that "high performance" on different measures occurs at random (according to a binomial distribution in which the probability of success on each trial is 0.25 and the eight trials are independent). The actual and the binomial distributions are not statistically different ( $P=0.49$ ).

## DISCUSSION

We found relatively weak associations between the assessed indicators of quality of HIV care. Of the 28 possible correlations between the eight quality measures, only two (7%) were greater than 0.40, two were between 0.30 and 0.39, and 20 (71%) were less than 0.20. This was particularly surprising because we assessed quality of care for a single chronic medical condition in sites that were specialised HIV clinics or had specialised HIV care teams. Furthermore, there were no more "high performing" organisations than were predicted by chance. Only one site was in the top quarter for seven measures, and no sites were in

**Table 3 | Correlations among adjusted quality measures (n=69)\***

	1	2	3	4	5	6	7	8
1 HAART use	1.0	—	—	—	—	—	—	—
2 Non-detectable HIV viral load	0.20	1.0	—	—	—	—	—	—
3 <i>Pneumocystis carinii</i> pneumonia prophylaxis	0.42***	0.02	1.0	—	—	—	—	—
4 Tuberculosis screening	0.13	0.02	0.18	1.0	—	—	—	—
5 Hepatitis C screening	-0.02	0.10	0.06	0.32**	1.0	—	—	—
6 Cervical cancer screening	-0.09	0.06	0.14	0.40***	0.05	1.0	—	—
7 Influenza vaccination	-0.002	0.30*	0.12	0.18	0.24*	0.07	1.0	—
8 Visits in three quarters	0.03	0.26*	0.04	0.08	0.009	0.21	0.14	1.0

\* $P<0.05$ ; \*\* $P<0.01$ ; \*\*\* $P<0.001$ .

the top quarter for six or eight measures. We expected that their focus on HIV care would lead some of these sites to develop systems and procedures that would positively affect multiple aspects of care. Moreover, guidelines for HIV care had been widely disseminated when these data were collected.<sup>15-17</sup>

We thought that in HIV care sites the preconditions would exist for high correlations among measures, including focus on a single condition, relatively high proportions of specialisation (97%), and the presence of multidisciplinary HIV care teams (at 80% of sites). Our results suggest that specialisation and focus on a specific condition may not be sufficient to produce high quality in multiple aspects of care. Consistency may require the coordination of multiple processes and procedures. Consider, for the sake of argument, two contrasting models of the clinical processes related to the eight quality measures we assessed. In the first model, a common system connects all eight measures. Elements in this common system include the physical space, clinic staff, a phone and messaging system, medical records, regular group meetings, and shared (specialised) clinical knowledge. In the second model, each quality measure can be thought of as the outcome of an independent chain of linked processes; failure of any single process in the chain causes the desired quality event not to occur. Because each chain of processes is independent of the others, success or failure of one chain has little impact on the success or failure of a simultaneously operating or parallel chain.

For example, starting and maintaining a patient on HAART may require preparatory visits with several different providers (for example, physicians, pharmacists, and case managers) and access to these providers during the initial phases of treatment. Doing tests to screen for tuberculosis requires a provider, usually a nurse, who carries out the test, ensures that it is appropriately read 24-48 hours later, and documents the results. Ensuring regular cervical cancer screening, on the other hand, may require the cooperation of a nearby gynaecology practice. Each of these examples involves largely independent chains of processes. *P carinii* pneumonia prophylaxis and HAART may have been more highly correlated than most other pairs of measures because the chains of processes that produce these outcomes have several shared elements (that is, both are prescriptions given by physicians, and both are guided by CD4 cell counts).

Adequate coordination among processes is probably more difficult to achieve when quality measures assess care given by different providers at multiple care sites (such as different services in a hospital). One potential reason that studies of quality<sup>19,20</sup> and quality improvement efforts<sup>13,21-23</sup> have yielded less impressive results than many expected may be the difficulty of simultaneously improving and coordinating multiple systems.

During the study period there were no specific incentives in place (financial or otherwise) for the practices we studied to meet specific quality targets, nor were there any centralised or public processes to

measure quality. Such measurement processes and incentives may increase correlations among quality measures, even in the absence of effective and coordinated systems and processes.

One interpretation of these data could be that providers recognise that they cannot provide uniformly high quality care and that they therefore prioritise. For example, few would debate that HAART use is clinically the most consequential of the eight measures we assessed, and the median proportion for use was among the highest the eight proportions at 0.81. Furthermore, the correlation between HAART use and *P carinii* pneumonia prophylaxis (also consequential clinically) was relatively high (0.42). On the other hand, prioritisation would not explain why the median proportion for hepatitis C screening was higher than for HAART use. High proportions of hepatitis C screening may be observed because it involves only ordering a blood test, and because once a positive result is found the test does not need to be repeated. If providers know that they have to trade off some goals of care against others, however, this is further proof that performance on one measure might imply little about performance on others, even in the setting of specialty care for a single disease. While no one advocates a healthcare system in which one measure of good care competes against another, limited resources and difficult choices are a reality in all healthcare settings.

#### Study limitations

We examined quality measures that could be assessed by reviewing medical records. All but one (non-detectable viral load) were measures of process. Our findings might have differed if we had been able to assess mortality, rates of admission to hospital, appropriate management of opportunistic infections, changes in health status, adherence to medication, or patients' reports about care.<sup>24</sup> We think that measurement of other processes or outcomes, however, would yield even lower correlations. Four of the care processes we assessed (screening for tuberculosis, hepatitis C, and cervical cancer, and influenza vaccinations) are simple to implement for anyone with basic clinical training, and the remaining four (HAART therapy, viral load control, *P carinii* pneumonia prophylaxis, and frequency of visits) are the subject of detailed guidelines for clinical practice.<sup>25,26</sup> Guidelines for tuberculosis skin testing do not suggest yearly testing, but rather that annual repeat testing (after an initial negative test result) should be considered in populations with a "substantial risk" of exposure (such as prison inmates),<sup>18</sup> which may have reduced the proportion who received tuberculosis screening. Another limitation of reviewing medical record is that processes may have been completed, but not documented, biasing our estimates downwards.

Finally, we studied patients at clinics receiving specific funding, and our findings may not generalise to other HIV care settings. Because this specific funding goes to rural and urban underserved care sites, our

**WHAT IS ALREADY KNOWN ON THIS TOPIC**

Selected indicators are often used as measures of overall quality of care

Few studies have published correlations between indicators of care quality, and none has done so for outpatient specialty care

**WHAT THIS STUDY ADDS**

There were low correlations among quality indicators for people with HIV disease

It might be hazardous to generalise from performance on a small number of quality indicators to performance on other indicators that are not measured

findings may not generalise to sites that care for patients with, for example, higher incomes, more education, and better health insurance. The sites studied, however, receive considerable scrutiny as a condition of participation in the programme, and quality levels there might be higher than at some other HIV care sites. To the extent that our study design excludes sites with consistently low quality scores, the correlations that we report are lower than they would be in a broader sample.

**Implications**

Our findings have implications for efforts to monitor quality and improvement. Current policy initiatives that seek to pay physicians for their performance on a small selected set of indicators or that create tiers of physicians or hospitals may not improve quality across a broad spectrum of care or conditions. Indeed, such programmes could prompt physicians or physician organisations to channel efforts into affecting the indicators being assessed to the detriment of other aspects of quality.<sup>27</sup> More empirical studies are needed on the impact of pay-for-performance initiatives and other improvement strategies on overall quality.<sup>28 29</sup>

Our results suggest that none of the sites we studied had the kinds of administrative, clinical, and human resources systems in place that are necessary to produce consistently high care quality. Continued and concerted efforts to improve healthcare systems may yield such patterns of high performance, but that goal has remained elusive to date. This should stimulate us to redouble our efforts to identify and implement the kinds of system changes that will allow us to cross the “quality chasm.”<sup>30</sup> Focusing on the improvement and coordination of multiple systems within organisations may be a useful direction to pursue.

We thank Carol Cosenza and Patricia Gallagher of the Center for Survey Research who assisted with instrument development and survey administration and our colleagues at the Health Resources and Services Administration and at the Institute for Healthcare Improvement who participated in and facilitated the EQHIV study.

**Contributors:** All authors made substantial contribution to conception, design, analysis, and interpretation of data, drafting the article, and revising it critically for important intellectual content and final approval. IBW is guarantor.

**Funding:** Agency for Healthcare Research and Quality (R-01HS10227), and the Lifespan/Tufts/Brown Center for AIDS Research (grant No P30AI142853). IBW was supported in part by a mid-career investigator award in patient oriented research from the National Center for Research Resources (K24 RR020300).

**Competing interests:** None declared.

**Ethical approval:** The committee on human studies of Harvard Medical School approved the study protocol.

**Provenance and peer review:** Not commissioned; externally peer reviewed.

- Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM. Pay for performance in commercial HMOs. *N Engl J Med* 2006;355:1895-902.
- Rowe JW. Pay-for-performance and accountability: related themes in improving health care. *Ann Intern Med* 2006;145:695-9.
- Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006;355:375-84.
- Berwick DM. A primer on leading the improvement of systems. *BMJ* 1996;312:619-22.
- Berwick DM, Nolan TW. Physicians as leaders in improving health care: a new series in Annals of Internal Medicine. *Ann Intern Med* 1998;128:289-92.
- Brook RH, McGlynn EA, Cleary PD. Quality of health care. Part 2: Measuring quality of care. *N Engl J Med* 1996;335:966-70.
- Gandhi TK, Francis EC, Puopolo AL, Burstin HR, Haas JS, Brennan TA. Inconsistent report cards: assessing the comparability of various measures of the quality of ambulatory care. *Med Care* 2002;40:155-65.
- Palmer RH, Wright EA, Orav EJ, Hargraves JL, Louis TA. Consistency in performance among primary care practitioners. *Med Care* 1996;34(9 suppl):S552-66.
- Jha AK, Li Z, Orav EJ, Epstein AM. Care in US hospitals—the Hospital Quality Alliance program. *N Engl J Med* 2005;353:265-74.
- Landrum MB, Bronskill SE, Normand SL. Analytic methods for constructing cross-sectional profiles of health care providers. *Health Serv Outcomes Res Methodol* 2000;1:23-47.
- Schneider EC, Zaslavsky AM, Landon BE, Lied TR, Sheingold S, Cleary PD. National quality monitoring of Medicare health plans: the relationship between enrollees' reports and the quality of clinical care. *Med Care* 2001;39:1313-25.
- Kerr EA, McGlynn EA, Adams J, Keeseey J, Asch SM. Profiling the quality of care in twelve communities: results from the CQI study. *Health Aff (Millwood)* 2004;23:247-56.
- Landon BE, Wilson IB, McInnes K, Landrum MB, Hirschhorn L, Marsden PV, et al. Effects of a quality improvement collaborative on the outcome of care of patients with HIV infection: the EQHIV study. *Ann Intern Med* 2004;140:887-96.
- Wilson IB, Landon BE, Hirschhorn LR, McInnes k, Ding L, Marsden PV, et al. Quality of HIV care provided by nurse practitioners, physician assistants, and physicians. *Ann Intern Med* 2005;143:729-36.
- Carpenter CC, Cooper DA, Fischl MA, Gatell JM, Gazzard BG, Hammer SM, et al. Antiretroviral therapy in adults: updated recommendations of the International AIDS Society-USA Panel. *JAMA* 2000;283:381-90.
- Yeni PG, Hammer SM, Carpenter CC, Cooper DA, Fischl MA, Gatell JM, et al. Antiretroviral treatment for adult HIV infection in 2002: updated recommendations of the International AIDS Society-USA Panel. *JAMA* 2002;288:222-35.
- Dybul M, Fauci AS, Bartlett JG, Kaplan JE, Pau AK. Guidelines for using antiretroviral agents among HIV-infected adults and adolescents. *Ann Intern Med* 2002;137:381-433.
- Masur H, Kaplan JE, Holmes KK. Guidelines for preventing opportunistic infections among HIV-infected persons—2002. Recommendations of the US Public Health Service and the Infectious Diseases Society of America. *Ann Intern Med* 2002;137:435-78.
- McGlynn EA, Asch SM, Adams J, Keeseey J, Hicks J, DeChristofaro A, et al. The quality of health care delivered to adults in the United States. *N Engl J Med* 2003;348:2635-45.
- Committee of Quality of Health in America, Institute of Medicine. *To err is human: building a safer health system*. Washington, DC: National Academy Press, 2000.
- Homer CJ, Forbes P, Horvitz L, Peterson LE, Wypij D, Heinrich P. Impact of a quality improvement program on care and outcomes for children with asthma. *Arch Pediatr Adolesc Med* 2005;159:464-9.
- Schonlau M, Mangione-Smith R, Chan KS, Keeseey J, Rosen M, Louis TA, et al. Evaluation of a quality improvement collaborative in asthma care: does it improve processes and outcomes of care? *Ann Fam Med* 2005;3:200-8.
- Asch SM, Baker DW, Keeseey JW, Broder M, Schonlau M, Rosen M, et al. Does the collaborative model improve care for chronic heart failure? *Med Care* 2005;43:667-75.
- Wilson IB, Ding L, Hays RD, Shapiro MF, Bozzette SA, Cleary PD. HIV patients' experiences with inpatient and outpatient care: results of a national survey. *Med Care* 2002;40:1149-60.

- 25 *Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents*. Washington, DC: US Department of Health and Human Services, 2004.
- 26 Yeni PG, Hammer SM, Hirsch MS, Saag MS, Schechter M, Carpenter CC, et al. Treatment for adult HIV infection: 2004 recommendations of the International AIDS Society-USA panel. *JAMA* 2004;292:251-65.
- 27 Fisher ES. Paying for performance—risks and recommendations. *N Engl J Med* 2006;355:1845-7.
- 28 Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance: from concept to practice. *JAMA* 2005;294:1788-93.
- 29 Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med* 2006;145:265-72.
- 30 Committee on Quality of Health Care in America. *Crossing the quality chasm: a new health system for the 21st century*. Washington, DC: National Academy Press, 2001.

**Accepted:** 3 September 2007