

# Physician-reviewers' perceptions and judgments about quality of care

SAUL N. WEINGART<sup>1</sup>, KENNETH MUKAMAL<sup>1</sup>, ROGER B. DAVIS<sup>1</sup>, DONALD T. DAVIES, JR<sup>1</sup>, R. HEATHER PALMER<sup>2</sup>, MICHAEL CAHALANE<sup>3</sup>, MARY BETH HAMEL<sup>1</sup>, RUSSELL S. PHILLIPS<sup>1</sup> AND LISA I. IEZZONI<sup>1</sup>

<sup>1</sup>Division of General Medicine and Primary Care, Beth Israel Deaconess Medical Center, Department of Medicine, Harvard Medical School, Charles A. Dana Research Institute and the Harvard-Thorndike Library, <sup>2</sup>Department of Health Policy and Management, Harvard School of Public Health and <sup>3</sup>Department of Surgery, Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, MA, USA

---

## Abstract

**Objective.** Although Peer Review Organizations (PROs) and researchers rely on physicians to assess quality of care, little is known about what physicians think about when they judge quality. We sought to identify features of individual cases that are associated with physicians' judgments.

**Design.** Using 1994 Medicare data, we selected hospitalizations for 1134 beneficiaries in 42 acute care hospitals in California and Connecticut. The sample was enriched with 17 surgical and six medical complications identified using diagnosis and procedure codes. PRO physicians confirmed quality problems using a structured implicit chart review instrument and provided written open-ended comments about each case. We coded physicians' comments for factors presumed to influence judgments about quality.

**Results.** In crude and adjusted comparisons, reviewers questioned quality more frequently in cases with serious or fatal outcomes, technical mishaps and inadequate documentation. Among surgical (but not medical) patients, they were less likely to record poor quality among patients presenting with an acute illness.

**Conclusion.** Factors other than the adequacy of key processes of care are associated with physician-reviewers' judgments about quality.

**Keywords:** complications, organization, peer review, quality of care

---

Researchers, Peer Review Organizations (PROs) and professional colleagues often ask physicians to review individual medical records to judge the quality of care [1–4]. This global assessment is called implicit review, as it relies on the reviewer's internal metric of quality.

In theory, reviewers should judge quality as satisfactory if patients were cared for appropriately. Appropriate care is usually defined in terms of processes such as diagnosis, treatment and prevention of complications. Reviewers' judgments of quality, however, are influenced by factors other than sound processes of care. Caplan *et al.* [5], for example, asked anesthesiologists to review a set of cases which were identical except for the random assignment of clinical outcome. In half of the cases, the patient died or was permanently disabled; while the remainder had better outcomes. Reviewers judged care much more harshly among cases with serious

adverse outcomes although the care was identical in each matched case.

Aside from these findings, little is known about the factors that may influence physician-reviewers' judgments about quality. We hypothesized that a variety of factors potentially affected physicians' judgments about quality, including:

- (i) characteristics of patients, such as advanced age, the presence of comorbid illness or chronic disease and do not resuscitate (DNR) orders;
- (ii) characteristics of diseases, especially the acuity and severity;
- (iii) characteristics of complications that patients experienced, including known or anticipated risks of treatment, technical mishaps, or uncertainty regarding the occurrence of a complication;

Address reprint requests to S. Weingart, Division of General Medicine and Primary Care, Beth Israel Deaconess Medical Center, 330 Brookline Ave, LY 326, Boston, MA 02215, USA. E-mail: sweingar@caregroup.harvard.edu.

- (iv) clinical outcomes;
- (v) incomplete documentation.

These factors do not necessarily reflect deficiencies in processes of care (e.g. diagnosis, treatment and prevention of complications) that are usually thought to indicate poor quality. They represent, instead, attributes of the case that might affect physicians' judgments about quality in cases with otherwise comparable care.

In order to understand whether physicians think about these factors when they judge quality, we examined physician-reviewers' written comments from a chart review study of quality of care among inpatients.

## Methods

### Sample selection

This project is part of a larger study to validate the Complications Screening Program (CSP), a computer program that uses hospital discharge abstracts to identify complications of care based on ICD-9-CM diagnosis and procedure codes [6,7]. We applied the CSP to Medicare's FY 1994 MEDPRO database of hospital discharge abstracts from California and Connecticut. We then stratified hospitals with high, medium and low complication rates among general surgery and medical patients. We selected hospitals at random from each stratum in each state and then selected randomly from patients whose cases were flagged by each of 17 general surgery complication screens and six medical screens and from patients whom the CSP did not flag. A list of CSP screens is included in Table 1. We sampled separately for medical and surgical cases. The final sample contained 630 surgical cases flagged by the CSP and 140 unflagged surgical cases, and 276 medical cases flagged by the CSP and 140 unflagged medical cases [8,9].

### Instrument design

We reviewed existing instruments and concluded that they did not permit us to meet the goals of a CSP validation study. We created an implicit physician review instrument based on reviews of the literature, clinician-investigator's judgment and experience and consultation with local subspecialists [9-12]. The instruments directed reviewers to select one or more quality deficiencies from a list of 13 problematic processes of care: inadequate preparation for surgery; problems with technical care during a surgical procedure, anesthesia care or medications; failure to monitor patient condition, respond to abnormal findings, provide preventive care or recognize a procedural or medication contraindication; inadequate facilities, equipment or staffing; delayed services or treatment and poor communication or coordination of care.

### Record abstraction

Each PRO selected six physicians with several years of medical chart abstraction and clinical experience. A 2-day orientation and training session was held at each site, led by a Harvard School of Public Health consultant who had

**Table 1** Complications Screening Program screens

General surgery cases
Postoperative cerebral infarction
Postoperative pneumonia
Aspiration pneumonia
Postoperative pulmonary compromise
Postoperative acute myocardial infarction
Postoperative shock or cardiorespiratory arrest
Deep vein thrombosis/pulmonary embolism
Postoperative gastrointestinal hemorrhage or hematoma
Procedure-related hemorrhage or hematoma
Procedure-related perforation or laceration
Reopening of a surgical site
Postoperative mechanical complications
Post-procedural wound infection
Postoperative infection (other than pneumonia or wound)
In-hospital hip fracture and falls
Miscellaneous complications
Iatrogenic complications
Medicine cases
Deep vein thrombosis/pulmonary embolism
Procedure-related hemorrhage or hematoma
Post-procedural wound infection
In-hospital hip fracture and falls
Medication complications
Iatrogenic complications

worked as clinical nurse and peer reviewer. The orientation included an introduction to the project objectives, an explanation of how the information would be used and an explanation of the reviewers' role. Reviewers studied the abstraction instruments and the coding manual (both available on request) and practiced sample cases. A nurse reviewer at each PRO read each case initially, using explicit review criteria to identify areas requiring special attention by the physician reviewer [8,13]. A single physician received the chart and implicit review instruments and identified quality of care problems. The reviewer was then unblinded as to the flagged complication by matching the chart abstraction and medical record materials with information about the flagged complication that had been kept separately in a sealed envelope. The reviewer was then directed to write open-ended comments about the quality of care.

### Inter-rater reliability of reviewers' judgments about quality

We considered a case to have evidence of substandard quality if reviewers identified at least one quality of care deficiency. We generated a random 5.8% reabstraction sample of 38 cases for Connecticut and 37 cases for California. Two physicians at each site reviewed the same case. To determine inter-rater reliability between the two PROs, we identified a random 1.5% sample of 19 cases that were abstracted by a reviewer at both PROs. We compared physicians' structured

**Table 2** Coding criteria for physician-reviewers' open-ended written comments

Factors hypothesized to influence quality judgments	Coding criteria	Examples
Advanced age	Age $\geq$ 70 Elderly Age influenced case	'Given age and primary problem [acute abdomen], I do not believe it is appropriate to code pulmonary compromise as a postoperative complication'. 'Elderly man admitted for acute arthritis of elbow'.
Co-morbid illness	Underlying chronic disease	'70 year old man with metastatic prostate cancer admitted with fever, altered mental status, pleural effusion, anemia'. '75 year old man rheumatoid arthritis and polymyalgia rheumatica treated with steroids admitted for perforated duodenal ulcer'.
DNR	Do not resuscitate No code Comfort care Care withdrawn	'Low blood pressure and hypotension in 93 year old post perforated gall bladder. No therapy because son requested no CPR'. 'Finally, at family's request treatment was withdrawn and patient arrested'.
Acuity of presentation	Life-threatening disease  Rapidly progressive downhill course	'74 year old woman with COPD admitted with 'bowel obstruction leading to surgery, resection of ischemic bowel, post-op respiratory insufficiency, re-explored, additional necrotic bowel resected, developed ARF, respiratory insufficiency, CVA, hypotension, sepsis, etc.'. 'Patient had massive CNS problem with ruptured aneurysm – critical from the start'.
Minimal complication	Minimal, trivial or insignificant complication  Not regarded as a true complication	'Elevated temperature and leukocytosis probably related to patient's underlying steroid dependency ... no real complications'. 'Small area of blistering – right buttock area post-op. Responded well to local treatment and avoidance of pressure ... There is no significant post-op complication'. 'Had some post-op hypertension requiring nipride/analgesics, transient bradycardia but really a completely unremarkable course. Nothing that can be considered an "event".'
Serious injury	Substantial injury or prolonged hospitalization	'68 year old woman admitted for carotid endarterectomy. Post-operative day 2 developed TIA ... Patient later had CVA'.
Death	Self explanatory	'Early next morning, found dead'. 'Did poorly thereafter with infection and died'.
Known risk of treatment	Known, expected or anticipated complication	'Had Swann-Gantz catheter insertion. Follow-up chest x-ray showed small pneumothorax – this is known potential complication of procedure'. 'Patient admitted for total hip replacement. Patient had temperature post-op attributed to local hematoma in hip wound. Common post-op occurrence after this type of surgery'. 'Patient experienced post-op ileus – would not consider a "complication" but rather a fairly frequent occurrence in this scenario ...'
Technical complication	Technical problems related to surgical and non-surgical procedures	'The laceration of rectum that occurred during vaginal hysterectomy on this 82 year old female is probably from a very thin recto-vaginal septum. This is also complicated by a large rectocele and adhesions'.

*continued*

Table 2 *continued*

Factors hypothesized to influence quality judgments	Coding criteria	Examples
		Patient underwent coronary artery bypass graft with repair of ... tear to ascending aorta ... This was a technical complication of surgery'. 'Three things went wrong in this case. First was a technical error when it was disclosed that the bowel stapler device had inadvertently removed part of the vagina'.
Diagnosis uncertain	Providers uncertain about the cause of a complication	'Post-op fever, question if due to infection, question of gout. Resolved'. Patient 'developed hypotension and metabolic acidosis and hepatic and renal insufficiency. Cause unknown ... a probable diagnosis of disseminated tuberculosis'.
Inadequate documentation	Important documentation missing from record	'Very poor documentation. Question of missing pages from the chart'.
	Insufficient information to judge quality	'Physical exam documentation not very good and no daily weights or I/O so am not sure if he was 'over diuresed' as implied in the discharge summary'.

ARF, acute respiratory failure. CNS, central nervous system. COPD, chronic obstructive pulmonary disease. CPR, cardiopulmonary resuscitation. CVA, cerebrovascular accident. DNR, do not resuscitate. I/O, intake and output. TIA, transient ischemic attack.

Table 3 Inter-rater reliability of coding of physicians' open-ended comments,  $n=100$ 

Factors that may influence quality judgments	Percent agreement	kappa
Patient characteristics		
Advanced age	94.0	0.75
Comorbid illness	81.0	0.41
DNR <sup>1</sup>	n/a	n/a
Disease characteristics		
Acuity of present illness	89.0	0.42
Outcome		
Minimal complication	91.0	0.43
Serious injury	89.0	0.64
Death	100.0	1.00
Nature of the complication		
Known risk of treatment	98.0	0.66
Technical complication	97.0	0.56
Diagnosis uncertain	96.0	0.31
Inadequate documentation	97.0	0.39

DNR = do not resuscitate.

<sup>1</sup>None present in sample.

implicit review judgments of whether any quality problem was present. Agreement was slight to fair. Physicians within a given state agreed 85.7% of the time ( $\kappa=0.22$ ) and physicians in different states agreed 72.2% of the time ( $\kappa=0.15$ ) about the presence of at least one quality problem.

### Physicians' comments

Our instrument asked reviewers to write open-ended comments about the case, explaining the basis for their judgments about quality on the structured review instrument. The instrument provided two half-page spaces for physicians' comments. Reviewers wrote a total of 1078 comments for 741 surgical (96.2%) and 393 medical patients (94.5%). The reviewer wrote two comments per case in 264 cases and one comment per case in the remainder.

A research assistant (D.T.D.) transcribed verbatim and a board-certified internist (S.N.W.) proofread and coded each comment, using the criteria outlined in Table 2. The internist did not supply an independent judgment or second-guess the reviewers' attributions, but instead, identified factors that influenced reviewers' judgments about quality. For example, one reviewer described a small pneumothorax as a 'known potential complication' of central venous catheterization. The reviewer's phrase was classified as a 'known risk of treatment'. Another reviewer described a 'tear to [the] ascending aorta' during coronary artery bypass graft surgery as a 'technical complication of surgery'. This comment was classified as a 'technical complication'.

To assess the reliability of free text coding, a second investigator (K.M.) read and coded a random sample of 100 open-ended written comments independently. The kappa statistic (used to correct for chance association) comparing codes assigned by S.N.W. and K.M. ranged from 0.31 to 1.00 (Table 3). There was moderate agreement ( $0.40 > \kappa \geq 0.60$ ) in four out of 10 factors and substantial agreement or better ( $\kappa > 0.60$ ) in four out of 10 factors. Review instruments and the CSP logic are available upon request.

**Table 4** Physician-identified quality problems (by structured implicit review)

Type of quality problem	Surgery ( <i>n</i> = 741 <sup>1</sup> )		Medicine ( <i>n</i> = 393 <sup>1</sup> )	
	<i>n</i>	%	<i>n</i>	%
Problems with technical care during a surgical procedure	62	8.4	13	3.3
Failure to provide preventive care (e.g. prophylactic antibiotic or anticoagulation)	35	4.7	11	2.8
Problem with medications administered	30	4.0	10	2.5
Failure to monitor patient condition or medications	24	3.2	5	1.3
Failure to respond to abnormal findings	19	2.6	4	1.0
Inadequate preparation for surgery	12	1.6	4	1.0
Delay in services or treatment	9	1.2	2	0.5
Problem with anesthesia care prior to or during a surgical procedure	7	0.9	0	0.0
Failure to recognize procedure contraindication	5	0.7	1	0.3
Inadequate or inappropriate equipment or facilities	4	0.5	0	0.0
Failure to recognize medication contraindication	3	0.4	3	0.8
Poor care communication or coordination of care	1	0.1	2	0.5
Inappropriate or inadequate staffing	1	0.1	0	0.0

<sup>1</sup>Percentages exceed 100% because cases may have more than one quality problem.

### Statistical analyses

We defined poor quality (the outcome variable) based on reviewers' selection of at least one of 13 quality deficiencies listed in Table 4. In contrast, we drew independent variables exclusively from reviewers' open-ended written comments about each case.

We used Fisher's exact test to assess whether particular physician comments occurred more often among patients with quality problems than among those without quality problems. We created a univariable logistic regression model to assess whether severity of adverse outcome (none = 0, minimal = 1, serious injury = 2, or death = 3) affected quality judgments. We analyzed major surgery and medicine patients separately to assess whether reviewers' judgments differed by clinical service.

We also examined whether the following variables, derived exclusively from physicians' comments, were associated with physician-identified quality deficiencies: advanced age, comorbid illness, DNR, acuity of illness, clinical outcome, known risk of treatment, technical complication, uncertain diagnosis and poor documentation. We completed this analysis by using a multivariable logistic regression model with backward elimination of variables with *P* values  $\geq 0.05$ . Statistical analyses used Stata Version 4.0 (StataCorp, College Station, Texas).

### Results

Reviewers identified at least one quality of care deficiency in 171 (23.1%) of 741 surgical patients and in 47 (12.0%) of 393 medical patients. Reviewers most often identified quality problems involving the following processes: technical care

during a procedure, failure to provide preventive care (e.g. prophylactic antibiotics or anticoagulation) and problems with medication administration. Table 4 lists the 13 quality of care deficiencies that we offered physician reviewers on the implicit review forms. Next, we examined reviewers' comments among cases with and without quality deficiencies.

#### General surgery cases

Among general surgery patients (Table 5), reviewers' comments about age, comorbid illness, DNR status and acuity of illness were not associated with quality deficiencies. Reviewers noted serious injury or death more often among patients with quality problems (36.9 versus 23.9%) than among those with minimal consequences (2.9 versus 8.1%, *P* for trend = 0.02).

Reviewers' notations about known risks of treatment and uncertain diagnoses were not associated with assessments of quality. Reviewers identified more technical complications (22.8 versus 4.4%, *P* < 0.01) and poor documentation (14.6 versus 7.5%, *P* = 0.01) among patients with quality problems than among those without.

In the multivariable model for general surgery, acuity of illness (OR = 0.43, 95% CI = 0.22–0.86) was associated with decreased odds that reviewers found quality of care deficiencies. In contrast, reviewers' comments about adverse outcome (OR = 1.37, 95% CI = 1.14–1.65), technical complications (OR = 6.52, 95% CI = 3.77–11.28) and inadequate documentation (OR = 2.13, 95% CI = 1.22–3.72) increased the odds that reviewers identified quality deficiencies.

#### Medicine cases

Medicine cases (Table 6) mirrored surgery cases in that reviewers were more likely to mention technical complications

**Table 5** Factors hypothesized to influence physician-reviewers' judgments about quality deficiencies among surgical cases (n = 741)

	Quality deficiency present		Quality deficiency absent		P value	Multivariable model	
	n = 171	%	n = 570	%		Odds ratio <sup>1</sup>	95% CI
Inadequate documentation	25	14.6	43	7.5	0.01	6.5	3.8– 11.3
Characteristics of the complication							
Known risk of treatment	7	4.1	33	5.8	0.45	n.a.	
Technical complication	39	22.8	25	4.4	<0.01	2.1	1.2– 3.7
Diagnosis uncertain	11	6.4	42	7.4	0.74	n.a.	
Clinical outcome							
Minimal complication	5	2.9	46	8.1	0.02 <sup>2</sup>	1.4	1.1– 1.7
Serious injury	49	28.7	94	16.5			
Death	14	8.2	42	7.4			
Disease characteristics							
Acuity of present illness	14	8.2	70	12.3	0.17	0.4	0.2–0.9
Patient characteristics							
Advanced age	23	13.5	86	15.1	0.71	n.a.	
Co-morbid illness	40	23.4	126	22.1	0.75	n.a.	
DNR	2	1.2	20	3.5	0.13	n.a.	

DNR = do not resuscitate.

<sup>1</sup> n.a. indicates variable dropped from model. <sup>2</sup> Test for trend.

Table 6 Factors hypothesized to influence physician-reviewers' judgments about quality deficiencies among medical cases ( $n=393$ )

	Quality deficiency present		Quality deficiency absent		P value	Multivariable model	
	$n=47$	%	$n=346$	%		Odds ratio <sup>1</sup>	95% CI
Inadequate documentation	10	21.3	13	3.8	<0.01	7.0	2.8–17.6
Characteristics of the complication							
Known risk of treatment	2	4.3	17	4.9	1.00	n.a.	
Technical complication	3	6.4	2	0.6	0.01	12.0	1.8–78.2
Diagnosis uncertain	3	6.4	11	3.2	0.23	n.a.	
Clinical outcome							
Minimal complication	4	8.5	19	5.5	<0.01 <sup>2</sup>	1.7	1.2–2.3
Serious injury	8	17.0	9	2.6			
Death	3	6.4	12	3.5			
Disease characteristics							
Acuity of present illness	5	10.6	13	3.8	0.05	n.a.	
Patient characteristics							
Advanced age	13	27.7	49	14.2	0.03	n.a.	
Co-morbid illness	13	27.7	47	13.6	0.02	n.a.	
DNR	3	6.4	5	1.4	0.06	n.a.	

DNR = do not resuscitate.

<sup>1</sup>n.a. indicates variable dropped from model. <sup>2</sup>Test for trend.

(6.4 versus 0.6%,  $P=0.01$ ), poor documentation (21.3 versus 3.8%,  $P<0.01$ ) and adverse clinical outcome (23.4 versus 6.1%,  $P<0.01$ ) among patients with quality problems. In addition, reviewers identified the potentially mitigating factors such as advanced age (27.7 versus 14.2%,  $P=0.03$ ) and comorbid illness (27.7 versus 13.6%,  $P=0.02$ ) more often among patients with quality deficiencies than those without.

Age and comorbid illness, however, were not independently associated with quality judgments in the multivariable analysis. In the model for medical patients, reviewers' comments about adverse outcome (OR = 1.66, 95% CI = 1.18–2.34), technical complications (OR = 11.98, 95% CI = 1.84–78.17) and inadequate documentation (OR = 7.02, 95% CI = 2.81–17.57) increased the odds that the case included quality deficiencies.

## Discussion

Although implicit physician reviews of the medical record are often used to evaluate quality, we know little about the factors that influence physician judgments. This study used the open-ended comments of experienced PRO medical record reviewers to elucidate factors that were associated with physicians' judgments about quality. In crude and adjusted comparisons among surgical and medical cases, reviewers judged poor quality in cases where they made special note of serious or fatal consequences, technical complications and inadequate documentation. Among surgical (but not medical) cases, reviewers also identified acuity of the presenting illness as a mitigating factor. Our findings confirm the relationship between adverse clinical outcome and physician judgments of quality that Caplan *et al.* [5] described.

We hypothesized that reviewers would forgive complications of a technical nature, given a presumed reluctance on their part to second guess the operator in the setting of a surgical or non-operative procedure. In fact, reviewers judged technical mishaps harshly. Perhaps reviewers felt they could distinguish between acceptable and unacceptable technical glitches or identify quality problems in cases that involved technical mishaps. We did not find that appropriate management of comorbid illness leads to favorable judgments of care, a result that Hulka *et al.* [1] reported in a study of quality among ambulatory medical practices. Interestingly, poor documentation increased the odds of finding substandard quality. Reviewers may have thought that poor documentation indicated sloppy care, inadequate attention to detail, or poor clinical judgment, rather than inadequate information from which to draw conclusions.

This study has several limitations, including poor inter-rater reliability of quality judgments by PRO reviewers and a small abstraction sample. This is a problem found in many studies of health care quality [14–18]. Inter-rater agreement about negligent care was poor ( $\kappa=0.24$ ), for example, in the Harvard Medical Practice Study [4]. Second, we considered substandard care present when reviewers identified at least one quality deficiency. Because specific quality deficiencies do not necessarily imply overall substandard care, this may overestimate the number of cases with poor quality. Third,

a small sample size may limit our power to detect significant differences. Fourth, reviewers may have differed in their interpretation, enthusiasm or completeness of open-ended comments for each case. Furthermore, we cannot be certain whether reviewers' perceptions about quality in fact informed their implicit quality judgments or simply rationalized them; we argue only that an association exists. Finally, a single reviewer coded all physicians' written comments. While this insured consistency and showed sound inter-rater agreement, the utility of the approach requires further study.

Despite these limitations, we have identified several credible relationships between physicians' written comments and their judgments about quality. Our findings suggest that factors other than the adequacy of key processes like diagnosis, treatment and prevention of complications may influence reviewers' judgments about quality. Perhaps researchers could train implicit reviewers to recognize and avoid these biases, or to use explicit process criteria to complement implicit review. Further study of physicians' judgments may help us understand differences among highly experienced and well trained clinician reviewers [14–18] and inform the design of instruments that control for factors that influence reviewers' judgments. Understanding the nature of physicians' judgments about quality may clarify the basis for these judgments and improve their reliability.

## Acknowledgements

This research was supported by the Agency for Health Care Policy and Research, under grant no. R01 H509099.

## References

1. Hulka BS, Romm FJ, Parkerson GR *et al.* Peer review in ambulatory care: Use of explicit criteria and implicit judgments. *Med Care* 1979; **17**: supplement, i–vi, 1–73.
2. Dans PE, Weiner JP, Otter SE. Peer review organizations: promises and potential pitfalls. *N Engl J. Med* 1985; **313**: 1131–1137.
3. Draper D, Kahn KL, Reinisch EJ *et al.* Studying the effects of the DRG-based prospective payment system on quality of care: Design, sampling and fieldwork. *J Am Med Assoc* 1990; **264**: 1956–1961.
4. Brennan TA, Leape LL, Laird NL *et al.* Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J. Med* 1991; **324**: 370–376.
5. Caplan RA, Posner KL, Cheney FW. Effect of outcome on physician judgments of appropriateness of care. *J Am Med Assoc* 1991; **265**: 1957–1960.
6. Iezzoni LI, Daley J, Heeren T *et al.* Using administrative data to screen hospitals for high complication rates. *Inquiry* 1994; **31**: 30–55.
7. Iezzoni LI, Daley J, Heeren T *et al.* Identifying complications

- of care using administrative data. *Med Care* 1994; **32**: 700–715.
8. Iezzoni LI, Davis RB, Palmer RH *et al.* Does the Complication Screening Program flag cases with process of care problems? Using explicit criteria to judge processes. *Int J Qual Health Care* 1999; **11**: 107–118.
  9. Weingart SN, Iezzoni LI, Davis RB *et al.* Using administrative data to find substandard care: Validation of the Complications Screening Program. *Med Care* 2000; **38**: 796–806.
  10. Rubin HR, Rogers WH, Kahn KL, Rubenstein LV, Brook RH. Watching the doctor-watchers: how well do peer review organization methods detect hospital care quality problems? *J Am Med Assoc* 1992; **267**: 2349–2354.
  11. Hayward RA, McMahon LF Jr, Bernard AM. Evaluating the care of general medicine inpatients: How good is implicit review? *Ann Intern Med* 1993; **118**: 550–556.
  12. Rubenstein LV, Kahn KL, Reinisch EJ *et al.* Changes in quality of care for five diseases measured by implicit review, 1981 to 1986. *J Am Med Assoc* 1990; **264**: 1974–1979.
  13. Iezzoni LE, Lawthers A, Davis RB *et al.* *Screening Quality of Care using Administrative Data: Final Report to the Agency for Health Care Policy and Research, R01 HS09099*. Boston: Beth Israel Deaconess Medical Center, 1999.
  14. Richardson FM. Peer review of medical care. *Med Care* 1972; **10**: 29–39.
  15. Brennan TA, Localio RJ, Laird NL. Reliability and validity of judgments concerning adverse events suffered by hospitalized patients. *Med Care* 1989; **27**: 1148–1158.
  16. Goldman RL. The reliability of peer assessments of quality of care. *J Am Med Assoc* 1992; **267**: 958–960.
  17. Localio AR, Weaver SL, Landis R *et al.* Identifying adverse events caused by medical care: Degree of physician agreement in a retrospective chart review. *Ann Intern Med* 1996; **125**: 457–464.
  18. Smith MA, Atherly AJ, Kane RL, Pacala JT. Peer review of the quality of care. Reliability and sources of variability for outcome and process assessments. *J Am Med Assoc* 1997; **278**: 1573–1578.

Accepted for publication 29 May 2001