Discrepancies Between Explicit and Implicit Review: Physician and Nurse Assessments of Complications and Quality

Saul N. Weingart, Roger B. Davis, R. Heather Palmer, Michael Cahalane, Mary Beth Hamel, Kenneth Mukamal, Russell S. Phillips, Donald T. Davies, Jr., and Lisa I. Iezzoni

Objective. To identify and characterize discrepancies between explicit and implicit medical record review of complications and quality of care.

Setting. Forty-two acute-care hospitals in California and Connecticut in 1994.

Study Design. In a retrospective chart review of 1,025 Medicare beneficiaries age ≥ 65 , we compared explicit (nurse) and implicit (physician) reviews of complications and quality in individual cases. To understand discrepancies, we calculated the kappa statistic and examined physicians' comments.

Data Collection. With Medicare discharge abstracts, we used the Complications Screening Program to identify and then select a stratified random sample of cases flagged for 1 of 15 surgical complications, 5 medical complications, and unflagged controls. Peer Review Organization nurses and physicians performed chart reviews.

Principal Findings. Agreement about complications was fair ($\kappa = 0.36$) among surgical and was moderate ($\kappa = 0.59$) among medical cases. In discordant cases, physicians said that complications were insignificant, attributable to a related diagnosis, or present on admission. Agreement about quality was poor among surgical and medical cases ($\kappa = 0.00$ and 0.13 , respectively). In discordant cases, physicians said that quality problems were unavoidable, small lapses in otherwise satisfactory care, present on admission, or resulted in no adverse outcome.

Conclusions. We identified many discrepancies between explicit and implicit review of complications and quality. Physician reviewers may not consider process problems that are ubiquitous in hospitals to represent substandard quality.

Key Words. Medical record review, complications, quality of care, peer review organization

This study compares the results of medical record reviews by nurses and physicians, aiming to identify complications and to assess quality of care. Peer

Review Organizations (PROs) and researchers frequently use medical record reviews, sometimes performed by nurses and other times by physicians, to determine substandard care. Comparing results of nurses' and physicians' reviews underscores the strengths and limitations of each method.

Despite their widespread use, medical record reviews are potentially unreliable methods for identifying substandard quality (Ashton et al. 1999; Dans, Weiner, and Otter 1985; Goldman 1992; Richardson 1972). Physician review (called "implicit" because it relies on a global impression of care) may be biased by reviewers' experience, consistency, attention to detail, and harshness of judgment (Hayward, McMahon, and Bernard 1993; Hulka, Romm, Parkerson, et al. 1979; Localio, Weaver, Landis, et al. 1996; Sanazaro and Worth 1985). Nurse review (called "explicit" because it usually involves well-defined criteria) is insensitive (Camacho and Rubin 1998). Rubin, Rogers, Kahn, et al. (1992), for example, found that PRO nurse reviewers failed to identify two of every three cases judged below standard by the research team. If explicit review is insensitive and implicit review is biased, then studies that rely on these methods are potentially flawed.

This study is part of a larger project to validate the Complication Screening Program (CSP), a computer program that screens hospital discharge abstracts for potentially substandard care. We studied the validity of the CSP from three perspectives. We validated the accuracy of *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) coding used to construct the CSP screens, validated that the screens identified quality-of-care problems, and validated that hospital-level results were replicable using

This research was supported by the Agency for Health Care Research and Quality, under grant no. R01 HS09099. The views expressed are solely those of the authors.

Address correspondence to Saul N. Weingart, M.D., Ph.D., Staff Physician, Division of General Medicine and Primary Care, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, LY 313, Boston, MA 02215. Roger B. Davis, Sc.D., is a Biostatistician, Division of General Medicine and Primary Care, Beth Israel Deaconess Medical Center, Department of Medicine, Harvard Medical School, Boston, MA. R. Heather Palmer, M.B., B.Ch., S.M., is a Professor, Department of Health Policy and Management, Harvard School of Public Health, Boston, MA. Michael Cahalane, M.D., is a General Surgeon, Department of Surgery, Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, MA. Mary Beth Hamel, M.D., M.P.H., Kenneth Mukamal, M.D., M.P.H., and Russell S. Phillips, M.D., are General Internists. Donald T. Davies, Jr., B.A., is a Research Assistant; and Lisa I. Iezzoni, M.D., M.Sc., is a Health Services Researcher, 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, Boston, MA.

different data sources. In published reports to date, we corroborated by chart review 84 percent of ICD-9-CM codes among medical cases and 89 percent among surgical cases (Lawthers, McCarthy, Davis, et al. 2000). We also showed that the CSP may not perform very well in identifying quality problems. Using explicit nurse reviews, Iezzoni, Davis, Palmer, et al. (1999) found rates of potential quality problems of at least 25 percent in 15 surgical and 2 medical screens. Using implicit physician reviews, Weingart, Iezzoni, Davis, et al. (2000) reported potential quality problems in at least 25 percent of cases identified by 9 surgical and 1 medical screen.

The objective of this study is to compare the results of implicit and explicit reviews, focusing specifically on discrepancies between nurse and physician reviewers' judgments about complications and quality in individual cases.

METHODS

Description of the CSP

The CSP is a computerized algorithm that searches administrative data for 28 potential complications of adult medical and surgical hospitalizations (Iezzoni, Foley, Heeren, et al. 1992; Iezzoni, Daley, Heeren, et al. 1994a, 1994b). It uses data elements contained in standard discharge abstracts: age, sex, ICD-9-CM diagnosis and procedure codes, and number of days from admission to principal major surgeries or procedures. For example, it is unusual for surgeons to perform noncardiac surgery on patients with acute myocardial infarction (AMI) except in emergencies. If a patient had a surgery on the first or second hospital day and a secondary diagnosis of AMI, the CSP would flag the admission as a potential complication (i.e., a postoperative AMI). This strategy acknowledges that it is possible in some cases that the complication was pre-existing or caused by extenuating circumstances. The CSP logic is available on request.

Sampling of Cases

We applied the CSP to the fiscal year 1994 Medicare database containing hospital discharge abstracts in California and Connecticut. We analyzed 15 screens for major surgery and 5 screens for medical cases. Earlier studies of the CSP suggested that hospitals vary by observed-to-expected (O/E) rates of complications, and thus, we created a multivariable model to stratify hospitals

by O/E rates (Iezzoni, Daley, Heeren, et al. 1994b). We then selected hospitals randomly from each stratum in each state. We sampled admissions at random from each CSP screen and from among admissions that were not flagged by any screen (i.e., controls). Each PRO obtained photocopied medical records of sampled cases from the hospitals. The final sample included 337 surgical cases, 140 surgical controls, 208 medical cases, and 140 medical controls. The sampling schemes across CSP validation studies are presented in detail elsewhere (Lawthers, McCarthy, Davis, et al. 2000).

Chart Abstraction Instruments: Explicit Review

To determine abstraction criteria for complications of care by explicit review, we first searched the medical literature. For each complication, the lead clinician from a team of three general internists (M.B.H., K.M., and R.S.P.), a general surgeon (M.C.), and two health services researchers (L.I.I. and R.H.P.) proposed clinical criteria. Consensus was reached through discussion. We consulted informally with specialists within our academic medical center to clarify several issues. Reviewing the full listing of criteria was not feasible within our budgetary constraints, and thus, final nurse abstraction instruments included criteria supported by strong literature evidence and clinician-investigators' judgments. For example, the criteria required to confirm a postoperative myocardial infarction included electrocardiographic (EKG) evidence of a new postoperative AMI by EKG report; an elevated absolute creatinine kinase level, MB fraction, or MB index; or a physician note confirming an AMI.

We designed additional explicit review instruments to identify quality problems. We created 13 screen-specific forms addressing potentially standard processes of care. Again, using the medical literature and clinical judgment, we derived criteria to confirm process defects suggesting standard quality. For the remaining screens and control cases, it was unfeasible to conduct specific, detailed, explicit process reviews within 45 minutes. Therefore, we used a "common processes of care" form that contained aspects of care potentially relevant for all cases. Defects on the common processes of care form included absence from the medical record of a medication list, medication allergies, or vital signs within 24 hours of admission or preoperatively. Other examples included failure to monitor vital signs, EKG, or oximetry intraoperatively. We considered a complication or quality problem confirmed by explicit review if the case failed at least one explicit criterion.

Chart Abstraction Instruments: Structured Implicit Review

Structured implicit review instruments directed physicians to identify all complications considered by the CSP that occurred during the hospitalization. Physicians then identified quality problems pertinent to each complication from a list of 15 possibilities. Finally, physicians wrote open-ended comments about the case. We considered the complication confirmed by implicit review if physicians identified the complication that was flagged by the CSP. We considered the quality problem confirmed by implicit review when physicians identified at least one quality problem associated with the flagged complication. All review forms are available on request.

We pilot tested instruments with nurse and physician reviewers from the Massachusetts PRO (MassPRO). We modified explicit and implicit abstraction instruments to permit nurse reviews to be completed within 45 minutes and physician reviews within 1 hour. The Connecticut and California PROs reviewed the draft abstraction instruments and manual; their comments were incorporated in the final versions.

Chart Review Protocol and Procedures

Four nurses and six physicians with several years of medical chart abstraction and clinical experience reviewed cases at each PRO. Physician reviewers included both internists and general surgeons. We conducted a 3-day training session at each site.

A nurse read each case first. She completed an explicit review, identifying complications and possible quality problems. She also noted potential problematic aspects of the case identified during explicit review in order to facilitate physician review. The physician received the chart and implicit review materials along with nurses' notations.

The physician was blinded to the complication screen flagged by the CSP until after he or she identified complications and associated quality problems. After unblinding, physicians wrote open-ended comments about the case. Physicians were asked to reconcile differences between the complication that they identified and the CSP-flagged complication.

Physicians and nurses were expected to work together on-site to facilitate discussion and clarification of concerns raised in the nurse's explicit review. This occurred in Connecticut as planned. In California, nurses and physicians reviewed off-site; nurse reviewers were accessible to physicians by telephone.

Interrater Reliability

We compared reviewers' judgments about confirmation of the CSP-flagged complication and at least one quality problem, calculating the kappa statistic to correct for chance agreement. We generated a random reabstraction sample of 38 cases for Connecticut and 37 cases for California (5.8 percent of the total sample). Two nurses and two physicians at the same PRO reviewed the same case. Intrastate agreement among nurses was satisfactory for complications ($\kappa = 0.70$) and quality ($\kappa = 0.62$). Intrastate agreement among physicians was poor for complications ($\kappa = 0.22$) and quality ($\kappa = 0.22$).

We also generated a random reabstraction sample of 19 cases (1.5 percent of the total sample) for interstate review. Interstate agreement among nurses was satisfactory for complications ($\kappa = 0.55$) and quality ($\kappa = 0.41$). Interstate agreement among physicians was good for complications ($\kappa = 0.76$) but was poor for quality ($\kappa = 0.15$).

Data Analysis

We calculated the kappa statistic to compare nurse and physician determinations of complications and quality problems. To understand discrepancies between nurse and physician review in individual cases, we examined physicians' open-ended written comments.

Because we sampled cases flagged by the CSP rather than on the basis of known complications, calculation of the sensitivity of explicit review yielded unstable estimates. Given the rarity of most events that the CSP attempts to identify, precise calculation of sensitivity and specificity would require a larger sample of unflagged cases than was logistically practical in our study. We present the proportion of nurse-confirmed complications and quality problems that occurred among physician-confirmed cases. This approach makes no assumption about the value of physician judgment as a gold standard. We used Fisher's exact test to compare the number of process defects among cases and controls. Statistical analyses used SAS (Statistical Analysis Software, Version 6.12).

RESULTS

Complications: Measures of Concordance

Table 1 compares nurse and physician confirmation of the flagged complication. Confirmation rates varied substantially by reviewer and complication, from 6.1 percent of nurse-confirmed medication-related complications to 94.6

Table 1: Nurse and Physician Agreement about Confirmation of CSP-Flagged Complications

CSP Screen ^a	CSP Flagged n	Nurse Only n (percent)	Physician Only n (percent)	Both n (percent)	Neither n (percent)	Nurse-Confirmed Among Physician Confirmed (percent)	Kappa
Major surgery							
Postoperative cerebral infarction ^b	35	3 (8.6)	1 (2.9)	29 (82.9)	2 (5.7)	96.7	0.44
Postoperative pneumonia	40	10 (25.0)	5 (12.5)	22 (55.0)	3 (7.5)	81.5	0.05
Aspiration pneumonia	32	10 (31.3)	1 (3.1)	19 (59.4)	2 (6.3)	95.0	0.14
Pulmonary compromise ^b	44	7 (15.9)	3 (6.8)	30 (68.2)	4 (9.1)	90.9	0.31
Postoperative AMI	37	0 (0.0)	2 (5.4)	33 (89.2)	2 (5.4)	94.3	0.64
Postoperative shock or cardiorespiratory arrest ^b	39	7 (18.0)	2 (5.1)	27 (69.2)	3 (7.7)	93.1	0.28
Deep vein thrombosis/pulmonary embolism	36	0 (0.0)	1 (2.8)	27 (75.0)	8 (22.2)	96.4	0.92
Postoperative gastrointestinal hemorrhage	39	5 (12.8)	4 (10.3)	25 (64.1)	5 (12.8)	86.2	0.37
Postprocedural hemorrhage or hematoma	44	11 (25.0)	2 (4.5)	24 (54.5)	7 (15.9)	92.3	0.34
Procedure-related perforation or laceration	30	6 (20.0)	0 (0.0)	21 (70.0)	3 (10.0)	100.0	0.41
Reopening of surgical site	30	9 (30.0)	2 (6.7)	17 (56.7)	2 (6.7)	89.5	0.09
Postoperative mechanical complications	42	12 (28.6)	4 (9.5)	21 (50.0)	5 (11.9)	84.0	0.14
Postoperative wound infection	41	5 (12.2)	5 (12.2)	21 (51.2)	10 (24.4)	80.8	0.47
Postoperative infection (except pneumonia and wound)	27	2 (7.4)	1 (3.7)	21 (77.8)	3 (11.1)	95.5	0.60
In-hospital hip fracture and falls	21	2 (9.5)	3 (14.3)	12 (57.1)	4 (19.0)	80.0	0.44
Totals for surgical cases	537	89 (16.6)	36 (6.7)	349 (65.0)	63 (11.7)	90.6	0.36
Surgical controls (not flagged) ^c	140	5 (3.6)	4 (2.9)	4 (2.9)	127 (90.7)	50.0	0.44
Medical							
Deep vein thrombosis/pulmonary embolism	42	9 (21.4)	1 (2.4)	19 (45.2)	13 (31.0)	95.0	0.53
Postprocedural hemorrhage or hematoma	45	10 (22.2)	5 (11.1)	23 (51.1)	7 (15.6)	82.1	0.25
In-hospital hip fracture and falls	39	1 (2.6)	1 (2.6)	6 (15.4)	31 (79.5)	85.7	0.83
Medication complications ^b	49	0 (0.0)	7 (14.3)	3 (6.1)	39 (79.6)	30.0	0.41
Totals for medical cases	175	20 (11.4)	14 (8.0)	51 (29.1)	90 (51.4)	78.5	0.59
Medical controls (not flagged) ^c	140	4 (2.9)	10 (7.1)	3 (2.1)	123 (87.9)	23.1	0.25

^aWound infection (medical patients only) had no explicit criteria and is not included in this table.
^bNo explicit criteria were developed for this screen. Nurse confirmation is based on physician note.
^cNurse and physician confirmations are based on identification of any CSP complication pertinent to the risk pool.

percent of physician-confirmed postoperative AMI. Reviewers confirmed more CSP-flagged complications among surgical than medical cases (88.3 percent vs. 48.6 percent). Among cases confirmed by a physician, the probability of a nurse-confirmed complication exceeded 0.80 for all but one screen (medication-related complications).

Reviewers disagreed frequently about the presence of the flagged complication in individual cases. Physicians confirmed complications where nurses did not in up to 14.3 percent of cases (medication-related complications and in-hospital hip fracture and falls among surgical patients). Nurses confirmed complications where physicians did not in up to 30.0 percent of cases (reopening of a surgical site). Seven surgical screens had kappas greater than or equal to 0.40. Three medical screens had kappas greater than or equal to 0.40. Overall, reviewers disagreed about complications more often in surgical than medical cases ($\kappa = 0.36$ vs. 0.59).

Complications: Physician Comments

To understand discordant reviews, we examined physicians' comments in the surgical and medical screens with poorest interrater agreement: major surgery cases with postoperative pneumonia and medical cases with postprocedural hemorrhage or hematoma.

Physicians failed to verify nurse-confirmed complications because they judged the complication clinically insignificant ("blood oozing" rather than hemorrhage), attributed it to a related diagnosis ("post-op pulmonary edema and atelectasis, not pneumonia") or found that it was present on admission. Physicians identified complications in cases where nurses found none; in several, they characterized the diagnosis as equivocal ("perhaps pneumonia. . . but no confirmation of organisms" or "a borderline call"). Some reviewers confirmed a complication on the implicit review instrument but indicated in open-ended comments that the complication did not occur—an apparent contradiction.

Quality of Care: Measures of Concordance

Table 2 compares nurse and physician confirmation of quality problems. Confirmation rates varied substantially by reviewer and complication from 2.0 percent of nurse-confirmed medication-related complications to 82.5 percent of nurse-confirmed postoperative pneumonia. Reviewers confirmed more quality problems among surgical than medical cases (65.5 percent vs. 39.4 percent). Among cases with a quality problem confirmed by a physician, the probability of a nurse-confirmed quality problem exceeded 0.80 for two

Table 2: Nurse and Physician Agreement About the Presence of at Least One Quality Problem

CSP Screen ^d	CSP- Flagged n	Nurse Only		Physician Only		Both		Neither		Nurse-Confirmed Among Physician Confirmed (percent)		Kappa
		n (percent)	n (percent)	n (percent)	n (percent)	n (percent)	n (percent)	n (percent)	n (percent)			
Major surgery												
Postoperative cerebral infarction	35	10 (28.6)	6 (17.1)	5 (14.3)	14 (40.0)	45.5		0.03				
Postoperative pneumonia	40	31 (77.5)	0 (0.0)	2 (5.0)	7 (17.5)	100.0		0.02				
Aspiration pneumonia	32	17 (53.1)	2 (6.3)	5 (15.6)	8 (25.0)	71.4		0.02				
Pulmonary compromise	44	19 (43.2)	5 (11.4)	4 (9.1)	16 (36.4)	44.4		-0.06				
Postoperative AMI	37	14 (37.8)	5 (13.5)	3 (8.1)	15 (40.5)	37.5		-0.08				
Postoperative shock or cardiorespiratory arrest	39	14 (35.9)	4 (10.3)	3 (7.7)	18 (46.1)	42.9		-0.01				
Deep vein thrombosis/pulmonary embolism	36	11 (30.6)	4 (11.1)	15 (41.7)	6 (16.7)	78.9		0.15				
Postoperative gastrointestinal hemorrhage	39	7 (17.9)	11 (28.2)	4 (10.3)	17 (43.6)	26.7		-0.03				
Postprocedural hemorrhage or hematoma	44	16 (36.4)	4 (9.1)	13 (29.5)	11 (25.0)	76.5		0.14				
Reopening of surgical site	30	7 (23.3)	9 (30.0)	6 (20.0)	8 (26.7)	40.0		-0.07				
Postoperative mechanical complications	42	9 (21.4)	9 (21.4)	5 (11.9)	19 (45.2)	35.7		0.04				
Postoperative wound infection	41	7 (17.1)	8 (19.5)	3 (7.3)	23 (56.1)	27.3		0.04				
Postoperative infection (except pneumonia and wound)	27	7 (25.9)	7 (25.9)	5 (18.5)	8 (29.6)	41.7		-0.05				
In-hospital hip fracture and falls	21	11 (52.4)	0 (0.0)	5 (23.8)	5 (23.8)	100.0		0.18				
Totals for surgical risk pool	507	180 (35.5)	74 (14.6)	78 (15.4)	175 (34.5)	51.3		0.00				
Surgical control cases (not flagged)	140	61 (43.6)	0 (0.0)	3 (2.1)	76 (54.3)	100.0		0.05				
Medical												
Deep vein thrombosis/pulmonary embolism	42	18 (42.9)	1 (2.4)	11 (26.2)	12 (28.6)	91.7		0.22				
Postprocedural hemorrhage or hematoma	45	4 (8.9)	14 (31.1)	2 (4.4)	25 (55.6)	12.5		-0.03				
Wound infection	33	1 (3.0)	1 (3.0)	0 (0.0)	31 (93.9)	0.0		-0.03				
In-hospital hip fracture and falls	39	19 (48.7)	1 (2.6)	2 (5.1)	17 (43.6)	66.7		0.04				
Medication complications	49	0 (0.0)	7 (14.3)	1 (2.0)	41 (83.7)	12.5		0.19				
Totals for medical risk pool	208	42 (20.2)	24 (11.5)	16 (7.7)	126 (60.6)	40.0		0.13				
Medical control cases (not flagged)	140	6 (4.3)	2 (1.5)	1 (1.0)	131 (93.6)	33.3		0.18				

^aPostoperative perforation or laceration (surgical patients only) had no explicit process criteria and is not included in this table.

surgical screens (postoperative pneumonia and in-hospital hip fracture and falls) and one medical screen (pulmonary embolism/deep vein thrombosis).

Interrater agreement varied by screen but was poor overall. Physicians confirmed quality problems where nurses did not in up to 31.1 percent of cases (postprocedural hemorrhage or hematoma). Nurses confirmed quality problems where physicians did not in up to 52.4 percent of cases (in-hospital hip fracture and falls among surgical patients). No surgical screen and only one medical screen had a kappa greater than 0.20.

Quality: Physician Comments

To understand discordant reviews, we again examined physicians' written comments about cases flagged for postoperative pneumonia and postprocedural hemorrhage or hematoma. Physicians failed to verify nurse-confirmed quality problems because they judged the defects unavoidable or small compared with an overall satisfactory quality of care (as in the patient whose femoral arterial sheath precluded application of direct pressure to a bleeding site). In several cases, physicians said that no quality problem was present because the lapse did not contribute to the flagged complication. For example, a reviewer stated that "documentation of respiratory rate not adequate but not responsible for postoperative pneumonia... [therefore] no lapse in quality." This may represent the view that quality problems must lead to adverse outcomes (i.e., "no harm, no foul"). As with complications, some reviewers were internally inconsistent. They confirmed quality problems on the implicit review instrument, but their comments indicated that the quality problem occurred prior to admission.

Quality: Surgical Controls

We sought to explain why nurses confirmed many more quality problems among surgical controls than did physicians (45.7 percent vs. 2.1 percent). We examined physicians' written comments and nurse-identified process-of-care deficiencies in discordant cases.

Of the 61 surgical controls in which nurse reviewers confirmed quality problems and physicians found none, nurses identified a single process defect in 31 cases and 2 or more in the remainder. The most frequently reported deficiencies involved failures of intraoperative monitoring: measurement of fluid intake and output (26 cases), documentation of electrocardiogram rhythm every 30 minutes (21 cases), and documentation of respiratory rate every 30 minutes (17 cases).

Physicians commented about quality in 45 of 61 cases confirmed only by a nurse. They stated explicitly that no quality problem was present in 43. In one case, the reviewer stated that a preoperative medical consultation should have been ordered. In the last case, the reviewer identified a missed diagnosis: "The patient appears hypothyroid and no one is aware."

To understand why nurses identified quality problems among surgical controls, we compared process problems in five surgical screens that shared the "common processes of care" explicit review instrument. We found a statistically significant difference between the number of process problems among surgical cases and controls for 4 of 17 processes: medication problems involving wrong dose, form, or route of administration; inadequate postoperative monitoring of the patient's temperature; infrequent intraoperative EKG monitoring; and failure to discontinue mechanical ventilation within 24 hours of surgery unless a physician indicated that weaning was begun or continued ventilatory support was required.

DISCUSSION

While examining the same medical records, nurse and physician reviewers often came to substantially different conclusions. Reviewers agreed more often about complications than quality problems, but interrater agreement varied substantially by complication. Reviewers agreed more often about complications among surgical than medical cases ($\kappa = 0.59$ vs. 0.36), but agreement about quality was little better than chance ($\kappa = 0.00$ vs. 0.13).

These findings inform the use of the CSP as a screening tool. Moderate agreement for complications among surgical cases in particular enhances the validity of the CSP for detecting complications in this population. In contrast, poor nurse and physician agreement raises our concern about the appropriate use of the CSP to screen for quality. Because nurse and physician reviewers appear to be measuring different phenomena, the validity of the CSP for measuring quality is potentially suspect, and the results must be interpreted cautiously.

Why do explicit and implicit review results diverge? Perhaps it is by design. If explicit review serves primarily as a screening tool, then we should have limited expectations about the capacity of an explicit instrument to capture the complexity of clinical care (Rubenstein, Kahn, Reinisch, et al. 1990). In support of this view, we found that extenuating circumstances (like the presence of a femoral catheter precluding the use of direct pressure over a

bleeding site) were not well captured by explicit review. Explicit review did not discriminate small lapses in the context of overall adequate or exemplary care. Discordant judgments sometimes reflected subtle distinctions among closely related diagnoses (e.g., respiratory failure rather than pneumonia) or cases in which complications were clinically insignificant or diagnoses uncertain.

A second explanation for the discrepancy between explicit and implicit review findings is that nurses and physicians attended to different phenomena. Nurses, directed by explicit review instruments, focused on specific clinical observations, laboratory studies, and clinicians' compliance with detailed process criteria. Physicians, in contrast, assessed the overall management of the case. Nurses were required by objectively defined criteria and detailed forms to categorize events as substandard care.

If nurses "missed the forest for the trees," physician often "missed the trees for the forest." Nurses frequently identified multiple process defects that physicians did not note. We do not know whether physicians judged process defects as unimportant or whether they missed them altogether. Hulka, Romm, Parkerson, et al. (1979) found that physician reviewers judged quality based on factors other than faulty processes. Among patients with high implicit ratings and poor adherence to explicit criteria, physician reviewers in Hulka, Romm, Parkerson, et al.'s study identified problematic processes only 53 percent of the time. In the remaining cases, they cited outcomes (22 percent), disease characteristics (19 percent), and patient characteristics (6 percent). Brook and Appel (1973) suggested that physicians rate "medical care in terms of conventional wisdom and not in terms of only the critical processes that would be likely to improve a patient's health."

A third explanation is that process problems among hospitalized patients are so commonplace that physicians regard substandard performance as usual care. We found that 51 percent of surgical patients and 46 percent of controls experienced at least one explicit process problem. Observational studies suggest that errors and mishaps are ubiquitous in health care (Donchin, Gopher, Olin, et al. 1995). The background prevalence of process problems decreased the power of our statistical tools to discriminate cases from controls.

This study has several limitations. First, our explicit review instruments may not adequately capture the complexity of the clinical setting. This was particularly worrisome for the CSP screens whose review criteria became so complex and lengthy that reviewers could not complete the instrument within the allotted time. We then developed a "common processes of care" instrument for these complications, recognizing that the use of a generic

instrument may diminish reviewers' ability to determine specific quality shortfalls.

Second, we may have overestimated the prevalence of quality problems by explicit review. We judged a quality problem to be confirmed by explicit review when a nurse confirmed at least one process defect. Because a case with adequate care overall may include specific lapses, this approach may overestimate the prevalence of quality problems. In addition, this approach contrasts with methods that require multiple process defects to define a quality problem (Ashton et al. 1999; Weissman, Ayanian, Chasan-Taber, et al. 1999). Lowering the bar may inflate the number of quality problems confirmed by explicit review.

Third, compliance with the abstraction protocol was variable. We found several cases of internally inconsistent physician reviews. In addition, some physicians failed to complete abstraction instruments; we returned 166 forms for reabstraction. This raises the possibility that our training session and abstraction manual were inadequate or that reviewers misunderstood the instructions. Differences in chart abstraction protocols in California and Connecticut (i.e., off-site physician review at the former) may have introduced bias. Finally, the use of a single physician reviewer per case may contribute to unreliable judgments of quality.

Indeed, poor interrater reliability on implicit review of quality judgments is an important concern. Geraci (2000) suggested several strategies by which we could have applied implicit review more rigorously, including the screening and exclusion of reviewers with excessively harsh or lenient judgments and the use of multiple implicit reviewers per case. Other strategies include the use of an elaborately structured implicit review instrument and more stringent reviewer recruitment and extensive training. Poor interrater reliability is common in studies of quality, prompting some investigators to recommend more widespread use of explicit criteria for quality of care research (Ashton et al. 1999).

Nevertheless, our results sound a cautionary note. Discrepancies between explicit and implicit review are not necessarily caused by the failure of explicit methods to account for the complexity of clinical practice. Skilled and experienced physician reviewers may not attend to important aspects of care. This may be due to the inherent difficulty of chart review and the responsibility of rendering judgments about quality. It may reflect slow dissemination of research results and evidence-based practice guidelines into clinical practice. However, it may also reflect reviewers' difficulty dealing with process problems

that are ubiquitous in health care. Problematic processes may be so commonplace that physicians regard these lapses as standard care.

These findings call for a re-examination of our approach to quality measurement. The PRO program in particular must be wary of its standard operating procedures. Experienced PRO reviewers may share informal biases about the conduct of chart review and the types of lapses that constitute substandard care. An interactive approach to chart review may be feasible—one that challenges physician reviewers either to acknowledge process deficiencies or to proposed improved explicit review criteria.

The findings should also capture the attention of risk managers. If process defects are ubiquitous in health care and often unrecognized by physicians, there is a reservoir of potentially discoverable data pertinent to liability claims in hospital records. Accreditation agencies, in addition, may find such records useful for conducting audits of high-risk processes.

To understand the discrepancies between explicit and implicit review, we need a clearer understanding of how physicians judge quality. What factors weigh most heavily in reviewers' minds? Which factors are discounted or ignored? Do features of the case other than processes like diagnosis, treatment, and prevention figure into their calculus? Perhaps we could judge better the adequacy of implicit review if we understood what reviewers' were thinking as they rendered these assessments. Eliciting reviewers' comments in writing or debriefing them verbally could generate data from which we could discern reviewers' intent, assess its value, avoid its biases, and perhaps, capture its wisdom.

REFERENCES

- Ashton, C. M., D. H. Kuykendall, M. L. Johnson, and N. P. Wray. 1999. "An Empirical Assessment of the Validity of Explicit and Implicit Process-of-Care Criteria for Quality Assessment." *Medical Care* 37 (8): 798–808.
- Brook, R. H., and F. A. Appel. 1973. "Quality-of-Care Assessment: Choosing a Method for Peer Review." *The New England Journal of Medicine* 288 (25): 1323–9.
- Camacho, L. A., and H. R. Rubin. 1998. "Assessment of the Validity and Reliability of Three Systems of Medical Record Screening for Quality of Care Assessment." *Medical Care* 36 (5): 748–51.
- Dans, P. E., J. P. Weiner, and S. E. Otter. 1985. "Peer Review Organizations: Promises and Potential Pitfalls." *The New England Journal of Medicine* 313 (18): 1131–7.

- Donchin, Y., D. Gopher, M. Olin, Y. Badihi, M. Biesky, C. L. Sprung, R. Pizov, and S. Coter. 1995. "A Look into the Nature and Causes of Human Errors in the Intensive Care Unit." *Critical Care Medicine* 23 (2): 294-300.
- Geraci, J. M. 2000. "In-Hospital Complication Occurrence as a Screen for Quality-of-Care Problems: What's Next?" *Medical Care* 38 (8): 777-80.
- Goldman, R. L. 1992. "The Reliability of Peer Assessments of Quality of Care." *Journal of the American Medical Association* 267 (7): 958-60.
- Hayward, R. A., L. F. McMahon, Jr., and A. M. Bernard. 1993. "Evaluating the Care of General Medicine Inpatients: How Good is Implicit Review?" *Annals of Internal Medicine* 118 (7): 550-6.
- Hulka, B. S., F. J. Romm, G. R. Parkerson, I. T. Russell, N. E. Clapp, and F. S. Johnson. 1979. "Peer Review in Ambulatory Care: Use of Explicit Criteria and Implicit Judgments." *Medical Care* 17 (Suppl): i-v, 1-73.
- Iezzoni, L. I., R. B. Davis, R. H. Palmer, M. Cahalane, M. B. Hamel, K. Mukamal, R. S. Phillips, N. J. Banks, and D. T. Davies, Jr. 1999. "Does the Complication Screening Program Flag Cases with Process of Care Problems? Using Explicit Criteria to Judge Processes." *International Journal of Quality in Health Care* 11 (2): 107-18.
- Iezzoni, L. I., J. Daley, T. Heeren, S. M. Foley, E. S. Fisher, C. Duncan, J. S. Hughes, and G. A. Coffman. 1994a. "Identifying Complications of Care Using Administrative Data." *Medical Care* 32 (7): 700-15.
- Iezzoni, L. I., J. Daley, T. Heeren, S. M. Foley, J. S. Hughes, E. S. Fisher, C. C. Duncan, and G. A. Coffman. 1994b. "Using Administrative Data to Screen Hospitals for High Complication Rates." *Inquiry* 31 (1): 40-55.
- Iezzoni, L. I., S. M. Foley, J. S. T. Heeren, J. Daley, C. C. Duncan, E. S. Fisher, and J. Hughes. 1992. "A Method for Screening the Quality of Hospital Care Using Administrative Data: Preliminary Validation Results." *Quality Review Bulletin* 18 (11): 361-71.
- Lawthers, A. G., E. P. McCarthy, R. B. Davis, L. E. Peterson, R. H. Palmer, and L. I. Iezzoni. 2000. "Identifying In-Hospital Complications from Claims Data: Is It Valid?" *Medical Care* 38 (8): 785-95.
- Localio, A. R., S. L. Weaver, R. Landis, A. G. Lawthers, T. A. Brennan, L. Hebert, and T. J. Sharp. 1996. "Identifying Adverse Events Caused by Medical Care: Degree of Physician Agreement in a Retrospective Chart Review." *Annals of Internal Medicine* 125 (6): 457-64.
- Richardson, F. M. 1972. "Peer Review of Medical Care." *Medical Care* 10 (1): 29-39.
- Rubenstein, L. V., K. L. Kahn, E. J. Reinisch, J. M. Sherwood, W. H. Rogers, C. Kamberg, D. Draper, and R. H. Brook. 1990. "Changes in Quality of Care for Five Diseases Measured by Implicit Review, 1981 to 1986." *Journal of the American Medical Association* 264 (15): 1974-9.
- Rubin, H. R., W. H. Rogers, K. L. Kahn, L. V. Rubenstein, and R. H. Brook. 1992. "Watching the Doctor-Watchers: How Well Do Peer Review Organization Methods Detect Hospital Care Quality Problems?" *Journal of the American Medical Association* 267 (17): 2349-54.

- Sanazaro, P. J., and R. M. Worth. 1985. "Measuring Clinical Performance of Individual Internists in Office and Hospital Practice." *Medical Care* 23 (9): 1097-114.
- Weingart, S. N., L. I. Iezzoni, R. B. Davis, R. H. Palmer, M. Cahalane, M. B. Hamel, K. Mukamal, R. S. Phillips, D. T. Davies, and N. J. Banks. 2000. "Using Administrative Data to Find Substandard Care: Validation of the Complications Screening Program." *Medical Care* 38 (8): 796-806.
- Weissman, J. S., J. Z. Ayanian, S. Chasan-Taber, M. J. Sherwood, C. Roth, and A. M. Epstein. 1999. "Hospital Readmissions and Quality of Care." *Medical Care* 37 (5): 490-501.