

The Process of Peer Review in U.S. Hospitals

Marc T. Edwards, MD, MBA, and Evan M. Benjamin, MD

Abstract

- **Objective:** To obtain information about peer review as currently practiced in U.S. hospitals.
- **Methods:** Leaders from sponsor-affiliated U.S. acute care hospitals were asked to complete a 39-item online questionnaire.
- **Results:** Data from 339 hospitals reveal wide variation in medical staff peer review structure, process, and governance. Ordinal logistical regression models for perceived reviewer participation, quality impact, and medical staff satisfaction explain a substantial proportion of the observed variance. Key drivers of the perceived impact on quality include reviewer participation, recognition of excellence, standardization and governance of process, integration with hospital performance improvement activity, reference to pertinent diagnostic studies and identification of contributory clinician-to-clinician issues during case review, turnaround time, case volume, monitoring of adverse events, and trustee involvement.
- **Conclusion:** Physician leaders should consider applying quality improvement principles to their organization's peer review processes, with particular attention to clinical performance measurement methods and program governance.

Clinical peer review, the process by which physicians evaluate each other's performance, has long been one of the activities by which the medical profession has sought to protect the quality of patient care. In U.S. acute care hospitals, peer review is also conducted to mitigate vicarious malpractice liability and comply with regulatory requirements, including accreditation, licensure, and Medicare participation [1]. Despite its importance, there are no normative data encompassing peer review program scope, structure, process, and governance. To address this knowledge gap, we conducted a survey to obtain information about current hospital medical staff peer review practices in the United States.

Methods

The survey was conducted using a Web-based utility (Survey Monkey.com). We sought the participation of hospital leaders

and administrators likely to have either detailed knowledge of their peer review program or the power to appropriately delegate the responsibility to respond: physician executives, quality/safety improvement leaders, risk managers, medical staff services heads, and CEOs or other designated organizational contacts.

To obtain our sample, we contacted all 50 state hospital associations, 3 national hospital alliances, and a physician executive association. We explained our study objectives and asked for assistance in soliciting participation. The American College of Physician Executives, the University HealthSystem Consortium, Premier Inc., and hospital associations in Arkansas, California, Florida, Michigan, Missouri, South Carolina, and Wisconsin agreed to partner with us. Each partnering organization used its own process for inviting appropriate members to participate in the survey through directed e-mail, electronic newsletter, or posting of a notice on a web portal.

We field-tested our draft survey and cover letter with hospital leaders and our partnering organizations to assess for relevance, appropriateness of scope, face validity, time to complete, and question ambiguity. The final survey contained a total of 39 questions and 5 demographic items (available at <http://QAtoQI.com/survey.htm>). Thirty-three of the questions were neither optional nor conditional on another item. A response was considered complete if at least 23 (70%) of these 33 core questions were answered.

The survey period ran from 25 October to 18 December 2007. Due to the overlap of membership among our partnering organizations and the diversity of the methods of soliciting participation, it was not possible to calculate the population of eligible hospitals or to estimate contact, refusal, cooperation, and response rates. In order to determine whether our survey sample was representative of American hospitals, we compared the demographics of our survey sample to current data on acute care facilities [2].

We used chi-square tests to measure associations among

From QA to QI Consulting, West Hartford, CT (Dr. Edwards) and Baystate Health, Springfield, MA (Dr. Benjamin).

Table 1. Demographic Data for Survey Respondents Compared with American Hospital Association Data

	Survey Sample		National Data	
	<i>n</i>	%	<i>n</i>	%
Staffed beds				
≥ 500	43	13	244	5
400–499	19	6	173	4
300–399	37	11	368	7
200–299	51	16	619	13
100–199	82	25	1129	23
50–99	38	12	1001	20
25–49	36	11	1032	21
6–24	25	8	370	7
Postgraduate training				
Major teaching	61	19	305	6
Minor teaching	73	23	901	18
Nonteaching	182	58	3730	76
Setting				
Urban	192	58	2927	59
Rural	138	42	2009	41

program characteristics. We estimated administrative support staffing levels per 100 beds based on the midpoint of the bed-size categories, using 20 and 600 beds respectively for the extremes. Then we applied analysis of variance and multiple regression methods to evaluate differences in program staffing and numbers of committees in relation to other variables, with and without high outliers removed. Finally, we developed multivariate ordinal logistic regression models for self-reported reviewer participation, perceived quality impact, and medical staff satisfaction. Hospital and program characteristics independently associated with these outcome variables were tested in the models. Demographic variables and other survey items thought to have potential relevance were also tested. Response levels were selectively collapsed guided by exploratory data analysis. To simplify presentation, we calculated the adjusted coefficient of determination (r^2) as a measure of total variance explained by each model from the equivalent multiple regression. Analysis was carried out using Minitab version 15 (Minitab, State College, PA).

Results

We obtained a total of 401 responses. Of these, 353 were complete (at least 70% of the core questions answered). Among these, 14 represented duplicates, leaving a total of 339 complete responses. Surveys were generally completed by the organization’s head of quality, safety and/or perfor-

mance improvement (50%), or a physician executive (24%). Compared with national data, our sample overrepresents major teaching hospitals and other large hospitals in relation to nonteaching facilities with 25 to 100 beds (Table 1).

Governance

Respondents indicated that the peer review process is governed by the medical staff’s executive committee (58%), a specially designated oversight committee (29%), or the hospital’s performance improvement committee (or equivalent; 11%). In 42% of responses, governance includes regular review of data involving the process and outcomes from peer review activity with meaningful discussion related to ongoing improvement of the process (irrespective of discussions about individual performance issues). In another 23%, process and outcomes measures are regularly reviewed with little discussion. In the remainder, review of such data is irregular, if occurring at all.

Approximately 73% routinely provide their board of trustees with information regarding peer review beyond that which might otherwise surface in relation to an adverse action. Of these 240 hospitals, 73% provide high-level aggregate data, 30% offer case-specific summary information, 21% provide physician-specific aggregate data, and 11% share case-specific detailed information.

Scope

The scope of peer review program activities is highly variable. Retrospective medical record review is the only activity common to almost all programs (Table 2).

Case Identification and Review

Various approaches are used to identify cases for peer review (Table 3). Nearly all hospitals (96%) use generic screens (eg, mortality, readmissions) to identify potential adverse events or substandard care. Following case identification, 92% perform additional screening prior to peer review. This most commonly is accomplished via chart review by nonphysician clinicians (76%) and/or physicians (50%). In hospitals that do a secondary screen, 44% send fewer than 20% of cases to formal peer review, while 16% send more than 80%. In the course of case review, in addition to examining the medical record reviewers also refer to pertinent diagnostic studies (76%), case screening notes (51%), formally prepared case abstracts (35%), and pertinent operative images (34%).

Peer Review Process

Peer review is conducted in a committee setting in nearly all hospitals (98%). The number of committees varies with hospital size, ranging from a median (interquartile range [IQR]) of 1 (1–3) for hospitals under 50 beds to 10.5 (2–20)

Table 2. Scope of Peer Review Activities (*n* = 339)

	<i>n</i>	%
Retrospective medical record review	326	96
Comparative evaluation of performance measures (eg, complication rates, core measures, patient satisfaction)	251	74
Root cause analysis	224	66
Morbidity & mortality case conference	197	58
Concurrent medical record review	182	54
Comparative evaluation of aggregate data from peer review	168	50
Proctoring for new privileges	161	48
Conducting performance improvement projects	140	41
Other interventions to improve individual performance	120	36
Producing educational programs for groups of clinicians	117	35
Other forms of direct observation	80	24
Physician health program administration	51	15
Other	7	2

for those over 500 beds. 34% indicated their process is highly standardized (eg, well-defined by a policy which is followed in practice, variation from which is formally approved by whatever committee provides oversight), 28% greatly standardized (eg, there is a defined process with some variation that has not been formally approved), and 18% standardized (eg, whether or not the process has not been formally defined, there is significant commonality in approach across the organization, despite the presence of significant variation). 20% indicated their process is either somewhat or minimally standardized. Elements included in the peer review process and percentage of hospitals endorsing each are shown in **Table 4**.

Final peer review decisions are generally made via group discussion to consensus (67%) or by the department chair (17%). Infrequently they are made by majority rule (8%), the independent opinion of a single reviewer (4%), the peer review committee chair (4%), or the average of multiple independent reviews (1%).

The majority of hospitals use a rating scale to record peer review judgments. Nearly half use a multifaceted categorical "scale." Forty percent use a 3-level verbal descriptor scale (eg, acceptable, minor deviation, major deviation) and 11% use a 2-level scale (eg, yes/no) Only 3 facilities reported using a scale with 5 or more levels (eg, ranging from excellent to very poor).

Of 215 respondents providing information, 19% esti-

Table 3. Methods of Peer Review Case Identification (*n* = 339)

	<i>n</i>	%
Generic screens for problematic cases (eg, mortality, readmission, operative complication)	326	96
Risk management concerns	313	92
Physician concerns	310	91
Hospital staff concerns	300	88
Patient complaints	289	85
Referrals from other committees or clinical departments	282	83
Unexplained deviation from protocols, pathways, or other standards	230	68
Quality improvement studies	191	56
Statistical monitoring of process or outcomes measures	155	46
Clinically "interesting" cases	143	42
Focused review of new privileges	135	40
Random selection	56	17
Other	11	3

mated their relative case volume as less than 1% of hospital admissions, while 21% estimated volume at 3% or greater. External peer review constituted less than 1% of total review volume for 87% of hospitals and less than 5% for another 8%. The average turnaround time for the peer review process was less than 60 days for 43% of respondents and more than 90 days for 21% of respondents.

Measurement and Feedback

Table 5 displays measures used to assess the process and outcomes of peer review activity.

About 50% of programs aggregate and analyze individual peer review data on a regular basis (eg, quarterly, semiannually). The same proportion provides feedback to clinicians either following every review (31%) or on a regular basis (20%). Only 21% provide recognition for excellent performance at least occasionally.

Organization

Medical staff peer review activity was most commonly described as either decentralized under departmental administration (37%), centralized (32%), or partially centralized-partially decentralized (21%). In contrast, the hospital's administrative support for peer review activity tends to be centralized (67%) or partially centralized (19%). Only 20% compensate peer reviewers, generally via salary (39%) or an hourly fee (30%). 30% reported that peer review is highly

Table 4. Elements Included in the Peer Review Process (*n* = 337)

	<i>n</i>	%
Categorization of an event type (eg, mortality, readmission)	273	81
Recommendations for improved performance of an individual clinician	236	70
Patient harm	219	65
Identification of clinician-to-clinician issues (gaps in communication, call coverage, supervision, coordination among clinicians)	208	62
Other recommendations (referrals for additional review, group education)	196	58
Overall quality of care rating for an individual clinician	192	57
Rating of whether an untoward event was preventable	180	53
Identification of process of care issues involving other disciplines, information systems, organizational policy/procedures	180	53
Identification of contributory factors (eg, high-risk patient or procedure)	174	52
Rating of the degree of deviation from the standard of care	143	42
Written case analysis	133	39
Structured ratings of specific elements of individual performance (legibility, quality of history and physical exam)	128	38
Rating of whether an individual clinician could have prevented an event	124	37
Other	7	2
None of the above	2	1

interdependent with hospital performance improvement activity. The same proportion reported high interdependence in relation to risk management.

Administrative staff support for peer review was quantified by 290 respondents. Roles include nurses or other non-physician clinicians (85%), administrative assistants (58%), physicians (40%), technical analysts (31%), and information systems specialists or programmers (21%). Excluding physician reviewers, a median (IQR) of 1.1 (0.6-2.1) full-time equivalent staff per 100 beds support peer review program administration. Programmers and technical analysts are uncommonly found in peer review support roles in hospitals under 100 beds. Higher total support staffing per 100 beds is best predicted by greater process standardization, broader scope of activities, and smaller bed-size, which together explain about 23% of variance. Staffing levels were not significantly associated with reviewer participation, perceived quality impact, or medical staff satisfaction.

Table 5. Process and Outcome Measures Tracked (*n* = 329)

	<i>n</i>	%
Trends in adverse event rates (either globally or by event type)	178	54
Trends in targeted clinical performance measures	168	51
Case review volume	162	49
Trends in individual or group performance on specific elements of care evaluated through the peer review process	145	44
Counts and/or patterns of recommendations for improved performance of clinicians	136	41
Counts and/or patterns of process of care improvement opportunities identified	126	38
Turnaround time for case review	59	18
We do not track and review any process or outcome measures in relation to our peer review program	59	18
Counts and/or topics of recommendations for group educational events	57	17
Case review backlog	47	14
Counts of clinicians recognized for excellent performance	21	6
Other	4	1

Perceived Participation, Impact, and Future Plans

Reviewer participation was described as excellent or very good among 53%, and fair or poor among 24%. 24% of respondents felt it is very likely that the peer review program makes a significant contribution to the quality of patient care; 33% said it is likely, and 31% somewhat likely. Medical staff perception of the peer review process was described as excellent or very good by 29%, and fair or poor by 34%. The estimated likelihood of a program change within 12 months was very likely among 29%, and likely or somewhat likely among 41%.

Multivariate Analysis

Multivariate models for perceived reviewer participation, quality impact, and medical staff perception of the process explain a substantial proportion of the observed variances. **Table 6** summarizes the program features that appear in the models. Controlling for other factors, major teaching hospital status predicts a 1-level change in the perceived impact on quality with an odds ratio (OR; 95% confidence interval [CI]) of 2.0 (1.1-3.8). In bivariate analysis, this relationship is not significant (*P* = 0.1). With this exception, respondent and hospital characteristics are not significant independent predictors in the models. None of the survey items in combination can explain more than 8% of the variance in the

Table 6. Multivariate Models Predicting Key Peer Review Program Outcome Variables*

Reported reviewer participation ($r^2 = 34\%$)

- Standardization of peer review process
- Regular feedback to clinicians of the results of peer review
- Integration with hospital performance improvement activity
- Recognition of excellence
- Identification of clinician-to-clinician issues
- Centralized or partially centralized review activity

Perceived impact on quality and safety

Best model ($r^2 = 53\%$)

- Reported reviewer participation
- Access to pertinent diagnostic studies during case review
- Governance of peer review process
- Sharing peer review information with the board
- Case review volume $\geq 1\%$ of hospital admissions

Best alternative model ($r^2 = 43\%$)

- Standardization of peer review process
- Integration with hospital performance improvement activity
- Active governance of peer review process
- Turnaround time for peer review under 90 days
- Identification of clinician-to-clinician issues
- Recognition of excellence
- Monitoring adverse event trends
- Major teaching hospital

Reported medical staff satisfaction

Best model ($r^2 = 52\%$)

- Reported reviewer participation
- Perceived quality impact

Best alternative model ($r^2 = 22\%$)

- Standardization of peer review process
- Integration with hospital performance improvement activity
- Turnaround time for peer review under 90 days
- Recognition of excellence
- Prereview case screening by physicians

*Complete details of model parameters available on request from author.

estimated likelihood of program change. Lack of standardization of process is the major associated factor. In contrast to programs with decentralized review activity, those with centralized or partially centralized activity were less likely to report review volume over 1%, with an OR (CI) of 0.46 (0.24–0.87). Centralized or partially centralized review activity was associated with greater perceived quality when controlling for case volume but not when also controlling for the degree of process standardization.

Discussion

Few studies have looked at peer review practices in aggregate [3–5]. Our survey sought to obtain information about current hospital medical staff peer review practices in the United States. Our results revealed wide variation in peer review program scope, structure, process and governance across a sample of 339 acute care hospitals (7% of U.S. total), along with associations between program features and perceived reviewer participation, medical staff satisfaction, and impact on quality of care.

The program features associated with perceived impact on quality of care included recognition for outstanding clinical performance; standardization and governance of peer review process; integration with hospital performance improvement activity; timeliness of review; and identification of contributory clinician-to-clinician issues during case review. These factors have face validity as elements of better process and may account for the greater perceived program impact.

The perception that medical staff had a positive opinion about the peer review process was associated with perceived program impact on quality. This is an encouraging finding and suggests that the more a peer review program relies on quality improvement principles and systems thinking, the more likely it will be embraced by the medical staff, potentially enhancing the quality of care.

We found widespread inclination toward program change. 70% of respondents estimated that modifications of peer review program structure, process, or governance would be at least somewhat likely within the next year.

Even so, we could not develop an adequate multivariate predictive model. Our results could have been confounded by the effect of Joint Commission standards for ongoing and focused professional practice evaluation (OPPE/FPPE), which took effect January 2008. Indeed, these standards were referenced in several comments to the survey. Nonetheless, we received even more comments reflecting internal organizational drivers of change and dissatisfaction with current practice.

The significance of case volume in the regression model ($\geq 1\%$ of admissions reviewed associated with perceived impact on quality) may seem counter to quality improvement principles; however, in general, clinical care is not a well-controlled process. Weinberg and Stason found a 6% rate of quality improvement opportunities among admissions to a primary care service in a community hospital [6]. Thus, low peer review case volume may result in neglect of many potential opportunities for improvement and might also hamper efforts to provide timely performance feedback to medical staff.

Overall, our survey findings point to the need to improve the peer review process. If hospitals and their medical staffs have come to embrace the quality improvement movement over the past decade, it is less clear that peer review is routinely conducted within that framework. Few frequently recognize excellent performance. For the majority, there is significant room for improvement in terms of the degree of integration with hospital performance improvement activity and the timeliness with which data is aggregated, analyzed, and shared with clinicians.

There is also significant opportunity to better standardize the peer review process and improve the tracking of outcome measures for peer review. It appears that few have thought it important to do this. We believe this stems from an underlying failure of program governance. It would be difficult to systematically improve and standardize peer review processes without diligent governance, leadership support, and attention to program metrics.

The interrater reliability of peer review methods has received considerable scrutiny [7,8]. Greater interrater reliability has been seen with the use of structured ratings of specific elements of performance using a 5-level scale ranging from excellent to very poor [9]. Only 38% of hospitals in our sample employed structured ratings for specific performance elements, and virtually none used a 5-level rating scale. Unstructured ratings and scales with fewer than 5 levels have unacceptably low reliability [7].

A more reliable approach to performance measurement also would support the OPPE/FPPE process. Sanazaro and Worth have shown that even relatively small sample sizes may suffice to reliably compare performance when

structured ratings are used [10]. Such changes are advisable if peer review is to be an ongoing exercise in clinical performance measurement and improvement rather than an isolated, binary judgment of competence.

We observed a median of 1.1 full-time equivalent staff committed to peer review program support per 100 beds. This represents only 0.2% of the national average of 495 staff per 100 beds [11]. Further staffing research is recommended to better understand infrastructure needs in relation to peer review program design and related processes (eg, performance improvement, risk management).

Our study has significant limitations. The data are self-reported, without independent validation. We did not compare perceptions of program impact on quality to objective clinical quality and outcomes measures. Our sampling method was unbiased only with respect to each partnering organization and, in aggregate, overrepresented major teaching hospitals. There is also the potential for nonresponse bias.

Medical staff peer review in the United States appears to be evolving from a model focused primarily on the lowest performers to one more closely aligned with contemporary quality improvement philosophy. Our study offers a framework for thinking about the range of potential options for improvement and highlights those likely to have the greatest impact. Medical staffs should consider applying quality improvement principles to their peer review programs, with particular attention to enhancing clinical performance measurement methods and program governance. Additional research is necessary to further define best practice and more directly demonstrate the value of peer review.

Acknowledgments: In addition to the support provided by our organizational partners, we gratefully acknowledge the contributions of Deborah Naglieri-Prescod, PhD, who managed the interface to Survey Monkey throughout the study period; Penelope Pekow, PhD, who provided advice regarding statistical analysis and presentation; and Mark Keroack, MD, MPH, and Robert Klugman, MD, who critically reviewed our draft manuscript.

Corresponding author: Marc T. Edwards, MD, MBA, QA to QI Consulting, 111 Hunter Dr., West Hartford, CT 06107, marc@QAtoQI.com.

References

1. Haines S. Hospital peer review systems: an overview. *Health Matrix* 1984;5:2:30-2.
2. American Hospital Association data. Accessed 30 Nov 2007 at www.thirdwaveresearch.com/aha_wizard/.
3. Ross K. Peer review in Texas: a survey of medical staffs. *Tex Med* 1987;83:91-3.
4. Rosen HM, Feigin W. Quality assurance and data feedback. *Health Care Manage Rev* 1983;8:67-74.
5. Lichtmacher A. Quality assessment tools: ACOG Voluntary Review of Quality Program, peer review reporting system.

- Obstet Gynecol Clin North Am 2008;35:147–62.
6. Weinberg NS, Stason WB. Managing quality in hospital practice. *Int J Qual Health Care* 1998;10:295–302.
 7. Goldman RL. The reliability of peer assessments: a meta-analysis. *Eval Health Prof* 1994;17:3–21.
 8. Rubin HR, Rogers WH, Kahn KL, et al. Watching the doctor-watchers: How well do peer review organization methods detect hospital care quality problems? *JAMA* 1992;267:2349–54.
 9. Rubenstein LV, Kahn KL, Reinisch EJ, et al. Changes in quality of care for five diseases measured by implicit review, 1981 to 1986. *JAMA* 1990;264:1974–9.
 10. Sanazaro PJ, Worth RM. Measuring clinical performance of internists in office and hospital practice. *Med Care* 1985;23:1097–114.
 11. AHA Hospital statistics. Chicago: Health Forum; 2008: 11.

Copyright 2009 by Turner White Communications Inc., Wayne, PA. All rights reserved.