The Objective Impact of Clinical Peer Review on Hospital Quality and Safety

Marc T. Edwards, MD, MBA (Principal and Corresponding Author)
QA to QI Patient Safety Organization
8905 Wildwood Links
Raleigh, NC 27613
marc@QAtoQI.com
Phone: (860) 521-8484
https://qatoqi.com

Financial Support: None

Key Words: peer review, quality, safety, improvement, hospital, evaluation

Running Title: Objective Impact of Peer Review

References:	**34**
Tables:	**5**
Figures:	**0**
Word Counts	
Abstract:	**84**
Text:	**3,308**

Version: as accepted for publication by the American Journal of Medical Quality
Last Modified: 09/20/17

First Submitted:	**3/21/10**
Accepted for Publication:	**5/10/10**
Published Online:	**12/15/10**
First Appeared in Print:	**3/15/11**

doi: 10.1177/1062860610380732

Edwards MT. The objective impact of clinical peer review on hospital quality and safety. Am J Med Qual. 2011;26(2):110-119.

The final definitive version is available at: http://ajm.sagepub.com/content/26/2/110.short

Conflict of Interest Notification

Marc T. Edwards, MD, MBA, is the President and CEO of QA to QI Consulting, West Hartford, CT. Dr. Edwards assists hospitals to improve quality, patient safety and resource management. He has special interest and expertise in peer review process. He donates a portion of his professional time to scientific research. He has no conflicts of interest with respect to this study.

Abstract

Despite its importance, the objective impact of clinical peer review on the quality and safety of care has not been studied. Data from 296 acute care hospitals show that peer review program and related organizational factors can explain up to 18% of the variation in standardized measures of quality and patient safety. The majority of programs rely on an outmoded and dysfunctional process model. Adoption of best practices informed by continuing study of peer review program effectiveness has potential to significantly improve patient outcomes.

Background

Clinical peer review has long served as the primary process through which physicians evaluate each other's performance with the aim of improving the quality and safety of patient care. Despite its importance to the profession and to society, only a few reports present measures of its effectiveness. These are limited to studies of individual peer review programs at the department level. [1-4] Three of the 4 reports came from procedure-based specialties that traditionally track complication rates. There are no data comparing overall program effectiveness among institutions in terms of objectively-measurable clinical outcomes.

A 2007 study was the first to take a comprehensive look at peer review practice in the US. [5] This study established that virtually all programs invoke committee activity and case-based retrospective review, even though the scope of what constitutes peer review varies substantially. It also identified a set of specific factors that are strongly associated with belief that a program has a significant, ongoing impact on the quality and safety of care. These factors include standardization of process, recognition of excellence, attentive program governance, integration with other hospital performance improvement activity, reviewer participation, and identification of clinician to clinician issues (as well as other process problems) during the review process.

This group of factors fits well with accepted quality improvement principles. They have been translated into a 100-point, 13-item Peer Review Program Self-Evaluation Tool designed to

support organizational improvement efforts. [6] In essence, the Self-Evaluation Tool describes a quality improvement (QI) model best practice set for clinical peer review. The imputed Total Scores for the 2007 study group ranged from 0 to 86 with a mean of 45. The low mean score and wide variation suggested marked opportunity for program improvement. No other evidence-based, hospital-level best practice models are available.

The shortcomings of the prevailing quality assurance (QA) model for peer review have been extensively explored from multiple perspectives. [7-11] It is narrowly focused on detecting grossly substandard care ("weeding out the bad apples"). This makes it unnecessarily threatening to the vast majority of otherwise competent physicians who will inevitably be connected with bad patient outcomes at some point during their careers. By focusing on the cutoff point for substandard care instead of the measurement of clinical performance, the QA model loses the ability to address marginal practice and to influence overall group performance. The associated methods of making peer review judgments have low reliability. The QA model also neglects to identify and initiate fixes for the process problems contributory to adverse events, which are far more prevalent than substandard care. Moreover, it is disconnected from other organizational activity to measure and improve performance. Ironically, even its effectiveness in restricting incompetent physicians remains in question. [12]

Despite these limitations, the QA model has dominated medicine for 30 years, largely as an outgrowth of Joint Commission standards adopted in 1979 calling for an organized program of Quality Assurance. [8] Enough time has passed for the current generation of physicians to believe that the model is sacrosanct: "it's the way we've always done it and the way it should be done." Historically, the medical profession has used other methods. Peer review has been documented as early as the 11th century and may have originated in ancient Greece. [13] Modern practice emerged from Codman's End Results System and Ponton's concept of medical audit. [14-16] In the 1950's, Butler and Quinlan described an audit project to evaluate the quality of medical records as part of a larger program to improve patient care. [17] Their report outlines methods which anticipate the QI model. The 2007 study suggested that the QI model may be emerging to replace the QA model. If so, it would be instructive to know whether it is more effective.

This study was undertaken to determine whether the peer review program factors associated with higher subjective quality impact are also associated with better objective performance. It was part of a broader initiative that also assessed the practical utility of the Self-Evaluation Tool. That analysis has been reportedly separately. [18] It confirmed the 2007 findings of wide variation among programs, a high rate of change, a general lack of attention to program metrics, and the wholesale failure to reliably measure individual clinical performance during case-based review. Interestingly, it also showed that physician leaders use the language of quality improvement to characterize the factors that enhance or block program effectiveness, even if they have yet to systematically apply such principles to clinical peer review practice.

The Self-Evaluation Tool clearly differentiated hospitals across levels of subjective quality impact. The predictive value of the Tool was modestly enhanced with information about the

organizational culture. In particular, negative comments about the culture and its supports (i.e., resistance to change, lack of leadership and/or resources, a punitive review process, or belief that peer review is irrelevant to quality) were associated with lower performance when controlling for the Total Score.

Methods

The American College of Physician Executives (ACPE), Tampa, Florida, sponsored the study. ACPE has nearly 10,000 members, whose roles span the entire spectrum of the US healthcare system. The survey sample was constructed from a listing of those who had self-identified as holding leadership roles (such as Vice President Medical Affairs, Department Chair, Medical Director) in a hospital setting and who would, thereby, be expected to be intimately familiar with the organization's peer review process. For this part of the study, the sample frame was further restricted to non-Federal, acute care hospitals listed by CMS Hospital Compare in order to match a consistent set of objective measures.

The survey instrument and the methods for its analysis have been described. [18] In brief, the survey was based on the Peer Review Program Self-Evaluation Tool. It was designed to create a standardized picture of the critical aspects of peer review process in each organization. Because of the time lag for reporting objective measures and the high rate of expected change previously observed, the survey specifically requested the Federal fiscal year (FFY) of the last major peer review program change. The invitation to participate in the survey was distributed by email under a cover letter from the ACPE CEO. Data was collected electronically via web-based forms from August 11 through September 30, 2009. The survey instrument may be viewed at: QAtoQI.com/ACPE_survey.htm.

A response was considered complete if all 3 pages of the survey were submitted. Only complete responses were included in the analyses. One set of responses was chosen per facility. When the highest ranking physician executive was not obvious from the organizational titles provided, random number selection was used. Survey responses populated the variables given in Table 1.

Objective data for quality and patient safety were obtained from the CMS Hospital Compare website, Thomson Reuters (TR), Premier CareScience (PCS) and HealthGrades (HG). Each organization uses a different methodology to generate measures, but all are derived from MedPar and Core Measure data sets. These methodologies have been described in detail. [19-21]

TR stratifies all its measures by 5 levels of hospital size and post-graduate training. PCS and HG do not adjust for such factors. The component measures for the Patient Safety Indicators (PSI) are defined by the Agency for Healthcare Research and Quality (AHRQ), which provides the code grouping software. [22] TR provided percentile ranks. PCS provided risk-adjusted rates. HG provided z-scores, which measure the variation from the reference group mean in units of

standard deviations. Several associated hospital demographic variables were also considered: teaching status, admissions, and bed-size from MedPar; the TR hospital class; and Council of Teaching Hospitals (COTH) membership.

The Archived - September 2009 release of the Hospital Compare dataset was used. [23] Measures provided by TR and PCS were derived from Federal fiscal year (FFY) 2005-2007 reporting. The HG measures primarily came from FFY 2006-2008 data. Table 2 presents the entire list of measures studied with their respective measurement periods.

Pearson correlations were used to screen for associations between the survey variables and the objective quality measures. Multiple regression methods served to further characterize these relationships by controlling for the other factors. Conservative criteria for regression significance were used at the risk of concluding that a relationship did not exist when in fact it did. A regression equation was accepted if the overall *F*-test was significant at $p<0.01$, if the data sub-setting lack of fit test (for unexplained curvature) met $p>0.1$, and if all the *t*-tests for the factor coefficients and intercept constant were significant at $p<0.05$. Outliers values were retained, unless the elimination of a few extreme values was required to resolve the model fit. The only exception was the HG DVT PSI cohort for which 11 low performing outliers with z-scores < -12 had to be removed to achieve acceptable fit.

In addition, respondent hospitals were classified into thirds with respect to the Total Score. The top third (Total Score >55) was compared to the bottom third (Total Score <40) for performance measures having a valid regression model. Variation was characterized by the mean difference using a 2 sample *t*-test. This method helps to isolate the signal (effect of Total Score) from the noise of other sources of variation. Statistical Analysis was carried out using Minitab version 15 (Minitab Inc., College Station, PA).

Study Results

From the sample frame of 1017 institutions represented by at least one ACPE member, the survey process yielded 296 complete responses, 1 partial response, 15 break-offs, 3 refusals, and 27 undeliverable emails. The response rate adjusted for the estimated proportion of ineligibles in the non-response group was 36%.

Table 3 compares respondents to non-respondents and to all US acute-care hospitals in terms of demographics and objective quality measures. The ACPE somewhat over-represents teaching hospitals and under-represents small community hospitals. There are no meaningful differences between the respondent and non-respondent hospitals, except on the PCS ALOS measure. The respondent group performed slightly better than the US average with respect to: TR PSI, TR ALOS, PCS ALOS, TR Core Measures, and TR Overall; and slightly worse for PCS Complications and PCS Morbidity.

The Self-Evaluation Tool contains 2 items to assess for the use of structured ratings and reliable measurement scales in the peer review process. In auditing positive responses to them, no hospitals could be validated as using reliable methods for measuring clinical performance via peer review. While these items reflect literature and theory [7, 24], they were not among the regression factors in the 2007 study. The Total Score calculated from the remaining 11 items (80 points maximum) had acceptable reliability for aggregate comparisons and was found to correlate strongly with perceived quality impact (R^2=45%). [18] Therefore, it was used in the analyses. The mean Total Score was marginally higher than the 2007 cohort's (mean difference [CI] of 5.8 [3.2-8.4]). On the 80-point scale, only 25 (8%) of the 296 facilities scored at or above 72 (90%, i.e., A-level). 159 (54%) scored below 52 (65%, i.e., F-level).

32 significant regression models for outcome variables are delineated in Table 4 using the program factors described in Table 1. Program factors explain as much as 18% of the variation in objective measures of hospital quality and safety. With few exceptions, the factor relationships are in the expected direction. Identification of clinician to clinician issues, standardization of process, reviewer participation, the likelihood of future program change, and organizational/cultural factors appear most frequently. The perceived quality impact contributes to a small number of models. Table 5 summarizes the significant objective performance differences associated with the Total Score found by comparing the top third peer review programs to the bottom third.

Discussion

This study shows that important differences among clinical peer review programs predict a meaningful portion of the variation in hospital quality and safety on 32 objective performance measures. It is highly unlikely that such a large number of significant relationships occurred by chance alone. The effects are fairly small, but they are comparable to those found in other contexts. [25, 26] A large effect would not have been expected. Objective measures of quality and safety are a step removed from peer review and subject to the influence of unquantified factors, including all the other activity that hospitals undertake to improve their performance.

The real effect of peer review program differences may be even larger than these results suggest. Care outcome measures like mortality rates require a 3-year frame for stable performance comparisons. During FFY 2008 and 2009, medical staffs at 47% of hospitals studied made significant peer review program changes. Since the outcome measures referenced in this study reflect performance up through FFY 2007 or 2008, it may take several more years before they fully reflect the clinical peer review processes documented for this hospital cohort.

The hospitals selected for study are broadly representative of US hospitals. They exhibit the same wide variation in performance, even if, in aggregate, they show slightly above average performance. The objective measures of quality and safety were primarily derived from

administrative data. While they have limitations, they represent the state of the art. There are no better metrics for hospital-level clinical performance. The selected measures, including risk-adjusted average length of stay, reflect clinical activity commonly subject to peer review. Only one, the TR "Top 100 Hospitals" Overall measure, includes additional factors that are not as directly linked (financial performance and patient satisfaction). Although the relatively modest survey response rate widens the potential for non-response bias, the consistency of findings across 2 large, independent national samples separated by 2 years is reassuring. Thus, these results appear to be generalizable.

This study provides significant evidence that well-designed peer review processes improve quality and patient safety. Additional studies will help bring greater detail to the emerging QI model. Although care outcome measures are essential to evaluating the overall benefits from peer review, the lack of timeliness and specificity limits their usefulness going forward to identify and validate best practice innovations. Unfortunately, normative program-level outcomes data are not available, even for measures as simple as the count of learning opportunities identified and acted on. Together with program structure and process information, such data will be critical in guiding rapid-cycle tests of peer review process change. At least one collaborative project has been launched to confidentially collect and share high-level data for peer review process and outcomes along with program parameters. [27] In addition, more organizations need to report their innovations in peer review process backed by credible measures of impact from pre-post comparisons.

Organizational culture and leadership influences quality and safety. [28-31] The presence of cultural factors in these regression models matches with prediction and thereby adds credibility to the results. Adverse culture is associated with lower quality. Physician leaders should promptly abandon a fault-finding and punitive peer review process in favor of one that is more closely aligned with QI principles.

Beyond this, it would be helpful to see confirmatory results that fit well with a peer review-specific predictive model. The prevailing model for quality improvement holds that variation can be reduced and performance enhanced via the combined effects of leadership attention and support, the identification and correction of process problems, performance measurement, and performance feedback. The frequent appearance of standardization of review process, reviewer participation, and identification of clinical process problems as factors in the regressions is encouraging, but insufficient. Despite being the best single predictor of subjective quality impact, the Total Score was not a factor in any model. The small number of quality measures on which the top third performs better than the bottom third could be a spurious finding. If the Total Score had more pronounced effects, the test of congruence with a peer review-specific predictive model would have been more closely met. This might be found among a larger number of hospitals performing peer review in full accord with the QI model.

While this study was not designed as a pre-post comparison to 2007, the minimal shift in the distribution of Total Scores despite a high rate of program change should cause physician and hospital leaders to pause for deeper reflection on their goals and methods. Peer review is a

critical pillar of medical professionalism. If the Self-Evaluation Tool scores represented a test of peer review program effectiveness, more than half of hospitals get a failing grade. Do specific peer review practices have a greater effect on quality and safety than others? Physician and hospital leaders must answer this question. Any response compels action. If this study, other work and personal experience creates confidence that certain program parameters make a difference, then there should be adequate motivation to adopt them. If there are doubts, the process is important enough to warrant cooperative efforts to obtain satisfactory answers. For those who hold that peer review can never be effective, regardless of the methodology, there is nowhere to hide: the profession will face a real dilemma in terms of the likelihood of political backlash. This has already occurred in California where a legislatively mandated study found deficiencies in program rigor and public reporting. [32] Moreover, given the recent passage of healthcare reform legislation, the history of the Joint Commission action in 1979 should not be forgotten. It was a direct result of the failure of Professional Standards Review Organizations to control the spiraling costs of the Medicare program. [33]

This study opens the door to a fresh look at peer review practices. The medical profession should be able to demonstrate that its self-regulating activity is effective, not only in protecting the public from gross outlier behavior, but also in terms of making a vital contribution to the quality and safety of care. This study shows that a well-designed peer review process can do this. If hospital and physician leaders critically examine their overall Return on Investment (ROI) for peer review in this context, they will likely find incentive to abandon the QA model.

The shift to the QI model is a process improvement challenge that can be undertaken in small steps. Based on Graber's experience, there should be quick payback from training review committees on the principles of quality improvement and the technique of root cause analysis. [2] This will promote identification of clinical process problems during case review. Review committees also need to know where to refer problems for resolution that exceed their expertise or scope of authority. Such training may produce important side benefits in fostering a culture of safety and engaging reviewer participation.

Committed physician leaders can promote standardization of the peer review process through more attentive governance oversight, by revisiting the peer review policy, and by providing supports for change. They should also track simple measures of the process and outcomes of peer review activity. In addition, they can use the Self-Evaluation Tool as a guide to other leverage points for improvement and as a metric for monitoring progress in implementing the QI model.

Measurement of clinical performance may be the most difficult challenge of the QI model, simply because of the change in mindset that's required. It's not just a matter of redesigning forms. The profession has long ignored the truth about the poor reliability of QA-style categorical judgments about the standard of care. [7] Structured review methods can produce subjective performance measures that are adequately reliable for aggregate reporting and feedback. [11, 34] Nevertheless, such measures will never have the precision of common clinical lab tests. Physicians need to get comfortable about working within this limitation.

In conclusion, we are on the threshold of a major paradigm shift. A non-punitive peer review process infused with QI principles appears to be more effective than the "traditional" QA model. Given the wide variation in programs and the gap in application of QI principles to peer review processes, much work lays ahead before we can be satisfied that peer review is making its full contribution to the quality and safety of care. Adoption of best practices informed by continuing study of program effectiveness has potential to significantly improve patient outcomes.

Acknowledgments

This research was conducted without grant-funding or outside financial contributions. The author gratefully acknowledges the sponsorship provided by the American College of Physician Executives; the contributions of data, data interpretation and manuscript review provided by David Foster, PhD, of Thomson Reuters, Eugene Kroch, PhD, of Premier CareScience, and Harold Taylor, PhD, of HealthGrades; and the critical reviews offered by Azita Hamedani, MD, University of Wisconsin Hospital, Keith Starke, MD, St. John's Mercy Medical Center, and Rachel Sorokin, MD, Thomas Jefferson University Hospital.

Reprint Requests

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

References

1) Vitez TS. A model for quality assurance in anesthesiology. J Clin Anesth. 1990;2(4):280-287.

2) Graber ML. Physician participation in quality management: Expanding the goals of peer review to detect both practitioner and system error. Jt Comm J Qual Improv. 1999;25(8):396-407.

3) Olcott C IV, Mitchell RS, Steinberg GK, Zarins CK. Institutional peer review can reduce the risk and cost of carotid endarterectomy. Arch Surg. 2000;135(8):939-942.

4) Antonacci AC, Lam A, Lavarias V, Homel P, Eavey RA. A report card system using error profile analysis and concurrent morbidity and mortality review: surgical outcome analysis, part II. J Surg Res. 2009;153:95-104.

5) Edwards MT, Benjamin EM. The process of peer review in US hospitals. J Clin Outcomes Manage. 2009(Oct);16(10):461-467.

6) Edwards MT. Peer review: a new tool for quality improvement. Phys Exec. 2009;35(5):54-59.

7) Goldman RL. The reliability of peer assessments: A meta-analysis. Eval Health Prof 1994;17(1):3-21.

8) Sanazaro PJ, Mills DH. A critique of the use of generic screening in quality assessment. JAMA. 1991;265(15):1977-1981.

9) Berwick DM. Peer review and quality management: are they compatible? Qual Rev Bull. 1990;16(7):246-51.

10) Dans P.E., Clinical peer review: burnishing a tarnished image. Ann Int Med. 1993;118(7):566-568.

11) Edwards MT. Measuring clinical performance. Phys Exec. 2009;35(6):40-43.

12) National Practitioner Data Bank: Major Improvements Are Needed to Enhance Data Bank's Reliability. Washington DC: General Accounting Office; 2000. GAO-01-130. http://www.gao.gov/new.items/d01130.pdf

13) Ajlouni KM, Al-Khalidi U. Medical records, patients outcome, and peer review in eleventh-century Arab medicine. Ann Saudi Med. 1997;17(3):326-327.

14) Codman EA. A Study in Hospital Efficiency. Boston, MA: T Todd Company; 1917.

15) Ponton TR, Gauging efficiency of hospital and its staff. Mod Hosp. 1928;31(Aug):64-68.

16) Lembbcke PA. Evolution of the medical audit. JAMA. 1967;199(8):111-118.

17) Butler JJ, Quinlan JW. Internal audit in the department of medicine of a community hospital: Two years' experience. JAMA. 1958;167(5):567-572.

18) Edwards MT. Clinical peer review program self-evaluation for US hospitals. Am J Med Qual. 2010; Published online before print August 23, doi: 10.1177/1062860610371224 http://ajm.sagepub.com/content/early/2010/08/19/1062860610371224.abstract?papetoc

19) Griffith JR, Pattullo A, Alexander JA, Jelinek RC, Foster DA. Is anybody managing the store? National trends in hospital performance. J Healthcare Manage. 2006;51(6):392-406.

20) Pauly MV, Brailer DJ, Kroch G, Even-Shoshan O. Measuring hospital outcomes from a buyer's perspective. Am J Med Qual. 1996;11(3):112-122.

21) *The Twelfth Annual HealthGrades Hospital Quality in America Study*. Golden, CO: Health Grades, Inc.;2009 http://www.healthgrades.com/media/DMS/pdf/HealthGradesTwelfthAnnualHospitalQualityStudy2009.pdf .

22) *Guide to Patient Safety Indicators version 3.1*. Washington, DC: Agency for Healthcare Research and Quality; 2007 (http://www.qualityindicators.ahrq.gov/psi_download.htm).

23) http://www.cms.hhs.gov/HospitalQualityInits/11_HospitalCompare.asp#TopOfPage, accessed 10/27/2009.

24) Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to their Development and Use. 3rd ed. New York, NY: Oxford University Press; 2003.

25) Shortell SM, Zimmerman JE, Rouseau DM, Gillies RR, Wagner DP, Draper EA, Knaus WA, Duffy J. The performance of intensive care units: does good management make a difference? Med Care. 1994;32(5):508-525.

26) Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. JAMA. 2002;288(16):1987-1993.

27) www.qatoqi.com/proj_normative_database.htm

28) Bradley EH, Holomboe ES, Mattera JA, Roumanis SA, Radford MJ, Krumholz HM. The roles of senior management in quality improvement efforts: what are the key components? J Healthcare Manage. 2003;48(1):15-29.

29) Keroack MA, Younberg BJ, Cerese JL, Krsek C, Prellwitz LW, Trevelyan EW. Organizational factors associated with high performance in quality and safety in academic medical centers. Acad Med. 2007;82(12):1178-1186.

30) Vaughn T, Koepke M, Kroch E, Lehrman W, Sinha S, Levey S. Engagement of leadership in quality improvement initiatives: executive quality improvement survey results. J Patient Saf. 2006;2(1):2-9.

31) Kroch E, Vaughn T, Keopke M, Roman S, Foster D, Sinha S, Levey S. Hospital boards and quality dashboards. J Patient Saf. 2006;2(1):10-19.

32) Seago JA, Benroth R, McDougall B, Fernandez M, Giberson A, Daniel P. Comprehensive Study of Peer Review in California: Final Report. San Francisco, CA: Lumetra; 2008.

33) Dershewitz RA, Gross RJ. Why medical audits are in disfavor. Arch Int Med. 1980;140:168-169.

34) 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? JAMA. 1992;267(17):2349-2354.

Tables

Table 1. Peer Review Program Variables Studied and Their Descriptors

Variable Name	Descriptor
Peer Review Program Self-Evaluation Tool Items	
Total Score	Total Score derived from the following 11 variables
Standardization	Degree of standardization of peer review process within the hospital
Recognition	Frequency of giving recognition for outstanding clinical performance
Governance	Active governance of peer review process
Reviewer	Level of participation of reviewers in the peer review process
Integration	Integration of peer review with the hospital performance improvement process
Identification	Identification of process improvement opportunities, including clinician to clinician issues during peer review
Board	Board of Trustees involvement with peer review program performance
Feedback	Feedback of important review findings to clinicians in a timely manner
Volume	Volume of cases reviewed estimated as greater than 1% of hospital volume
Diagnostics	Whether pertinent diagnostic studies are examined in the case review process
Trends	Trends in adverse event rates monitored as an outcome measure of peer review
Other Survey Items	
Quality	Perceived impact of the peer review program on quality and safety
Satisfaction	Reported Medical Staff satisfaction with the peer review program
Change	Likelihood of significant program change in the next 12 months
Last	Federal Fiscal Year of the last significant peer review program change
Multispecialty	The extent to which a multispecialty peer review process might be used
Input	Likelihood of a request to a reviewed physician for input to the process
Stage	The point at which a request to a reviewed physician is commonly made
Organizational and Cultural Factors Reported to Explain the Likelihood of Future Change	
Accountability	Desire for more accountability for clinical performance
Compliance	Need to comply with accreditation or regulatory requirements
Fault	Desire to replace fault-finding with a focus on improvement opportunity
Inertia	Resistance to change
Involvement	Seeking greater participation in or commitment to the process
Leadership	Lack of physician and/or hospital leadership
Like	Like current program the way it is
Resources	Lack of resources (staff support, money, information systems, etc.)

Table 2. Outcome Variables Studied and Their Measurement Periods

Measure	Description	Measurement Period
Thomson Reuters	Measures given as percentile ranks	
Morbidity	In-hospital risk-adjusted complications index	10/05-9/07
Mortality	In-hospital risk-adjusted mortality index	10/05-9/07
PSI	Patient Safety Index derived from 15 AHRQ-defined indicators	10/05-9/07
ALOS	Severity-adjusted average length of stay weighted by R-DRG	10/05-9/07
Core Measures	An index based on CMS Core Measures	10/05-9/07
Overall	Top 100 Hospitals combined measure	10/05-9/07
Premier CareScience	Measures, except Overall Index, given as risk-adjusted rates	
Complications	In-hospital risk-adjusted complications rate	10/05-9/07
Morbidity	In-hospital risk-adjusted serious morbidity rate	10/05-9/07
Mortality	In-hospital risk-adjusted mortality rate	10/05-9/07
ALOS	Risk-adjusted average length of stay	10/05-9/07
Quality Index	Overall quality index rating	10/05-9/07
HealthGrades	Measures given as z-scores	
Morbidity	Risk-adjusted complications cohort roll-up	10/06-9/08
Mortality	Risk-adjusted overall mortality	10/06-9/08
Mortality Cohorts		
AMI	Heart attack	10/06-9/08
CHF	Heart failure	10/06-9/08
Pneumonia		10/06-9/08
Stroke		10/06-9/08
GI Surgery	Gastrointestinal surgical procedures	10/06-9/08
CABG	Coronary artery bypass surgery	10/06-9/08
PCTA/Stent	Coronary artery angioplasty procedures	10/06-9/08
PSI	Patient Safety Index derived from 15 AHRQ-defined indicators	10/05-9/07
PSI Cohorts		
Accidental Puncture/Laceration	Accidental cut, puncture, or perforation complicating a surgical procedure	10/06-9/08
DVT	Postoperative deep vein thrombosis or pulmonary embolism	10/06-9/08
FTR	Failure to rescue – a measure of preventable mortality among surgical inpatients	10/06-9/08
Hemorrhage	Postoperative hemorrhage or hematoma	10/06-9/08
Infection	Catheter-related bloodstream infection	10/06-9/08
Metabolic Derangements	Postoperative acute renal failure or uncontrolled diabetes	10/06-9/08
Respiratory Failure	Postoperative reintubation or mechanical ventilation	10/06-9/08
Sepsis	Postoperative bloodstream infection	10/06-9/08
CMS Measures	Measures given as rates	
Outcomes of care	Readmissions and 30-day risk-adjusted mortality	7/05-6/08
Process of care	Core Measure standards compliance rates	1/08-12/08

Table 3. Comparison of Class Distribution and Means for Performance Measures among Respondent, Non-Respondent, and All US Acute-Care Hospitals

TR Hospital Class	N (%)	N (%)	N (%)
Major Teaching	29 (9.8)	91 (13.7)	174 (5.8)
Teaching	101 (34.1)	157 (23.6)	426 (14.2)
Large Community	41 (13.9)	114 (17.2)	333 (11.1)
Medium Community	92 (31.1)	211 (31.8)	1103 (37.8)
Small Community	33 (11.2)	91 (13.7)	964 (32.1)

Measure	Respondent	Non-respondent	All US Acute-Care
	Mean [95% CI]	Mean [95% CI]	Mean [95% CI]
Thomson Reuters			
Morbidity	46.9 [43.4-50.3]	47.2 [44.9-49.4]	50.0*
Mortality	52.6 [49.1-56.0]	51.5 [49.3-53.7]	50.0*
PSI	55.0 [51.7-58.3]	51.7 [49.5-54.0]	50.0*
ALOS	54.0 [50.7-57.2]	50.1 [47.9-52.3]	50.0*
Core Measures	57.0 [53.8-60.1]	52.8 [50.6-55.0]	50.0*
Overall	54.9 [51.6-58.2]	51.1 [48.9-53.3]	50.0*
Premier CareScience			
Complications	42.2 [41.8-42.6]	41.5 [41.2-41.8]	40.9 [40.6-41.1]
Morbidity	12.9 [12.8-13.0]	12.9 [12.8-13.0]	12.7 [12.6-12.8]
Mortality	3.55 [3.44-3.66]	3.70 [3.62-3.78]	3.74 [3.68-3.80]
ALOS	3.92 [3.87-3.97]	4.10 [4.05-4.15]	4.00 [3.98-4.02]
Quality Index	100.2 [99.4-100.9]	99.8 [99.4-100.2]	100.0 [99.5-100.5]

*Reference value by measure design (CI not applicable)

Table 4. Program and Hospital Characteristics Predictive of Outcome Measure Performance*

Measure	Model R^2 (%)	Factor Influence on Measure	
		Favorable	**Unfavorable**
CMS AMI Mortality	12.0	Quality Admits	Change
CMS CHF Mortality	9.6	Identification Standardization Class	
CMS Pneumonia Mortality	8.7	Admits	Change
CMS AMI Readmits	4.7	Quality	COTH
CMS CHF Readmits	11.2	Identification Admits	Class
CMS Pneumonia readmits	9.2	Input	Class
TR Mortality	7.7	Standardization Identification Admits per Bed	Fault
PCS Mortality	4.1	Identification Multispecialty	
HG Morbidity	15.4	Standardization Identification Admits	Change
TR Complications	8.7	*Change*	Involvement Compliance *Multispecialty*
PCS Complications	4.6	Trends	*Quality*
TR PSI	10.2	Standardization Admits	Change
HG PSI	12.0	Reviewer Admits per Bed	Accountability
CMS Core Measures Overall Un-weighted Average	10.1	Reviewer Volume Class	Change
TR Core Measures	7.9	Reviewer	Change
TR Overall	7.0	Identification Standardization Admits	
TR ALOS	17.9	Identification Standardization Last Admits per Bed	*Volume* Compliance
PCS ALOS	17.7	Identification Standardization	*Volume* Beds
HG CHF Mortality	7.3	Identification *Accountability*	Change
HG Pneumonia Mortality	10.4	Identification Admits	Fault

Table 4. Program and Hospital Characteristics Predictive of Outcome Measure Performance*

Measure	Model R^2 (%)	Factor Influence on Measure	
		Favorable	**Unfavorable**
HG Stroke Mortality	5.6	Identification Admits	Fault
HG GI Surgery Mortality	6.9	Quality Standardization Admits	COTH
HG FTR PSI Mortality	3.2	Standardization	
HG DVT PSI Complications	9.2	Governance Last	Class
Heart Attack AMI-2 Aspirin at discharge	8.3	Reviewer COTH	
Heart Attack AMI-5 ß-blocker at discharge	4.4	Reviewer COTH	
Heart Failure HF-2 Left Ventricular Systolic function assessed	17.0	Reviewer Class	*Input*
Pneumonia PN-5 Antibiotic <6 hr.	6.5	Reviewer	Class
Surgery SCIP-INF-1 Antibiotic 1 hr prior	6.0	Reviewer Class	
Surgery SCIP-INF-2 Antibiotic choice	5.5	Identification	*Governance*
Surgery SCIP-VTE-1 Venous Thromboembolism Prophylactic orders	11.7	Identification Class	
Surgery SCIP-VTE-2 Treatment delivery	8.3	Identification Class	

* Paradoxical relations with program factors highlighted with italics

Table 5. Comparison of Total Score Top Third to Bottom Third Performance on Specific Measures of Quality and Safety

Measure	Top Third [CI]	Bottom Third [CI]	Difference [CI]	*P*
TR Overall[a]	61.5 [56.3-66.7]	53.2 [47.4-59.0]	8.2 [0.1-16.0]	0.04
TR Core Measures[a]	61.7 [56.5-66.9]	52.3 [46.5-58.1]	9.3 [1.6-17.0]	0.02
TR ALOS[a]	59.4 [53.6-65.2]	50.8 [45.2-56.4]	8.6 [0.7-16.5]	0.03
CMS AMI-5[b]	98.2 [97.6-98.7]	97.1 [96.7-97.5]	1.0 [0.7-2.0]	0.04

[a] Percentile rank
[b] Percent compliance with standard