Clinical Peer Review Program Self-Evaluation for US Hospitals

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, performance measurement, quality, patient safety, quality improvement, hospitals

Running Title: Clinical Peer Review

References: 17
Tables: 3
Figures: 2
Word Counts
Abstract: 149
Text: 3,517

Version: as accepted for publication by the American Journal of Medical Quality
Last Modified: 9/20/17

First Submitted: 12/24/09
Accepted for Publication: 3/9/10
Published Online: 8/23/10 **doi: 10.1177/1062860610371224**
First Appeared in Print: 10/15/10

Edwards MT. Clinical Peer Review Program Self-Evaluation for US Hospitals. Am J Med Qual. 2010;25(6):474-480.

The final definitive version is available at: http://ajm.sagepub.com/content/25/6/474

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 donates a portion of his professional time to scientific research. He has no conflicts of interest with respect to this study.

Abstract

Prior research has shown wide variation in clinical peer review program structure, process, governance and perceived effectiveness. This study sought to validate the utility o f a Peer Review Program Self-Evaluation Tool as a potential guide to physician and hospital leaders seeking greater program value. Data from 330 hospitals show that the Total Score from the Self-Evaluation Tool is strongly associated with perceived quality impact. Organizational culture also plays a significant role. When controlling for these factors, there was no evidence of benefit from a multi-specialty review process. Physicians do not generally use reliable methods to measure clinical performance. A high rate of change since 2007 has not produced much improvement. The Peer Review Program Self-Evaluation Tool reliably differentiates hospitals along a continuum of perceived program performance. The full potential of peer review as a process for improving the quality and safety of care has yet to be realized.

Background

Despite its importance, little data is available on the impact of hospital medical staff peer review on the quality and safety of care. As the key process by which physicians evaluate each other's performance, peer review also serves to maintain professional autonomy, uphold professionalism and protect the public welfare.

Only a few reports present objective measures of effectiveness for individual peer review programs (1-4). There are no data comparing program effectiveness among institutions in terms of measurable clinical outcomes.

A recent national survey identified substantial predictive value of specific practices on the level of belief that a program has a significant, ongoing impact on the quality and safety of care. (5) These practices made good sense from a quality improvement perspective. Edwards subsequently translated the survey results 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 for peer review that contrasts sharply with the oft-criticized, but still prevalent quality assurance (QA) legacy model for "weeding out the bad apples." When applied to the original study population, the distribution of Total Scores ranged from 0 to 86 with a mean of 45 and dramatized the overall improvement opportunity. Higher Total Score was strongly associated with a higher level of perceived quality impact and explained 49% of the variance. A 10 point increase in Total Score predicted a three-fold likelihood of higher quality impact.

Therefore, this study was initiated to validate the utility of the Self-Evaluation Tool and to determine whether the program factors associated with higher perceived quality impact are also associated with better objective quality performance. This report focuses on the analysis of the survey data. The comparison to objective quality data will be reported separately.

Methods

The American College of Physician Executives (ACPE), Tampa, Florida, agreed to sponsor the study. The ACPE is a membership association that has provided leadership development, educational programs, and professional networking for over 3 decades. ACPE has nearly 10,000 members, whose roles span the entire spectrum of the US healthcare system. ACPE hosts an online directory which is maintained by the membership. Approximately 20% have self-identified as holding leadership roles (such as Vice President Medical Affairs, Department Chair, Medical Director) in the hospital setting and who would, thereby, be expected to be intimately familiar with the organization's peer review process. ACPE provided a list of potential respondents from which the survey sample was constructed.

The questionnaire used to collect relevant peer review program data from this group is available for review at: http://qatoqi.com/ACPE_survey.htm. The Peer Review Program Self-Evaluation Tool, without the point scores, formed the core. The survey instrument also replicated items from the national survey, related to perceived quality impact, medical staff satisfaction, and the likelihood of program change. Open-ended questions were added to drill down on these perceptions and identify measures currently in use to evaluate peer review program effectiveness at the hospital. Because of the time lag for reporting objective measures and the high rate of expected change previously observed, it requested the fiscal year of the last major peer review program change. In addition, it included questions regarding multi-specialty review process and involvement of reviewed clinicians, which were not asked in the original survey. Special attention was given to collecting information needed to characterize whether a decision to opt out of the survey was due to lack of eligibility or other factors.

The invitation to participate in or opt out of the survey was distributed by email to specifically named persons under a cover letter from the ACPE CEO. Two reminder notices were sent to non-respondents. Data was collected electronically via web-based forms. Form validation rules required name, title, organization, and response to the 13 Self-Evaluation Tool items. The survey period ran from August 11 through September 30, 2009.

A response was considered complete if all pages of the survey were submitted and partial if only the Self-Evaluation items were entered. Break-off (demographic information only) was treated as an opt-out for reason of personal choice. Clarification was sought as needed via email or phone contact with respondents. The author classified responses to open-ended items using empirically-developed categories. Multiple categories were allowed. A custom-developed Microsoft Access 2007 database facilitated the process. Final disposition codes for the sample frame were recorded according to 2009 AAPOR standards. Only complete responses were considered in the analyses.

Simple counts and relative frequencies of responses to survey items were tabulated. For each respondent, the Total Score for the Self-Evaluation Tool items was calculated. The reliability of the Tool was estimated with the intraclass correlation coefficient from duplicate responses (intra-rater) and from paired responses from organizations with more than 1 respondent (inter-rater) using the method of

Shrout & Fleiss. (7) In the few cases with more than 2 responses, the response of the highest ranking physician executive was paired to 1 other selected by random number assignment.

Analysis of variance and multiple regression methods served to evaluate the relationships among survey variables. Statistical Analysis was carried out using Minitab version 15 (Minitab Inc., College Station, PA).

Results

From the sample frame of 1986 members, the survey process yielded 362 complete responses , 4 partial responses, 19 break-offs, 7 refusals, 70 opt-outs for reason of ineligibility, 86 undeliverable emails, 11 duplicate responses, and 47 ineligibles via reclassification. Among the opt-outs, 41 (59%) were not in a leadership role in a hospital setting and 29 (41%) lacked of knowledge of the organization's peer review process. The 362 complete responses came from 330 facilities, including 296 acute care, 16 children's, 11 critical access, 2 long term acute care, 2 psychiatric, 1 rehabilitation, and 2 Veterans Administration hospitals. The response rate adjusted for the estimated proportion of ineligibles in the non-response group (CASRO method AAPOR RR3) was 25%.

Organizations solicit reviewed clinicians for input to the peer review process *Frequently* (59%) or *Occasionally* (29%). The input is typically requested *Following initial review, but before final scoring* (45%), *During initial review* (33%), or *Before initial review* (15%). 42% report that a multi-specialty process is the norm for peer review, and 27% note that it is used at least occasionally. While many open-ended responses used the term "multi-disciplinary", when seeking clarification, the investigator found no examples of true inter-disciplinary peer review, with physicians, nurses and others evaluating clinical performance together, as equals. A multi-specialty process is associated with perceived quality impact, but is not an independent predictor when controlling for Total Score.

Medical staff perception of the peer review process was 8% *Excellent*, 29% *Very Good*, 36% *Good*, and 26% *Fair* or *Poor*. When asked "What is the likelihood that your Peer Review Program makes a significant ongoing contribution to the quality and safety of patient care at the hospital?", 34% responded *Very Likely*, 32% *Likely*, and 23% *Somewhat Likely*.

24% made significant program changes in federal fiscal year 2009, 23% in 2008, and 13% in 2007. Program changes among the 208 who provided detail included: standardization of process (16%); improvement of administrative support or program organization (16%); the introduction of a multi-specialty review process (15%); compliance with Joint Commission standards (11%); and better integration with organizational performance improvement processes (10%). The estimated likelihood of significant program change within the next 12 months was 18% *Very Likely*, 19% *Likely*, and 19% *Somewhat Likely*.

27% of respondents indicated the use of a rating scale with at least 5 levels from best 2 worst. 25% indicated that *Case review is documented by rating multiple elements of performance on a template selected to match the specific type of clinical activity being reviewed, possibly including an overall score, a case analysis, etc*. Because this was unexpected, an audit was conducted. No examples of forms meeting the intended criteria were identified. For this reason, the Total Score was calculated based on the other 11 items. Also of note, for most Self-Evaluation Tool items, the *Unknown* response choice was

selected at relatively low rates, but the volume of case review in relation to hospital inpatient volume was marked *Unknown* by 27%.

Figure 1 shows that the distribution of the Total Score for 11 items (80 points maximum) approximates a normal distribution with a mean[CI] of 47.5 [45.6-49.3]. Using 27 paired ratings, the inter-rater reliability is estimated as ICC (1, 1) 0.61 [0.31-0.80]. The reliability of the mean of 2 independent ratings of the Total Score is estimated as ICC (1, 2) 0.75 [0.47-0.89]. Using 11 duplicate responses, the intra-rater reliability is estimated as ICC (1, 1) 0.88[0.63-0.97].

The Total Score was strongly associated with both perceived quality impact (R^2=45%) and medical staff perceptions (R^2=37%). A 10 point increase in Total Score predicts a 1 level increase in estimated quality impact with an OR [CI] of 2.6 [2.3-3.0]. The equivalent estimated mean [CI] Total Score from the 2007 study was 41.7 [37.1-40.5]. The estimated difference [CI] between these 2 population means is 5.8 [3.2-8.4], p<0.001.

Only 20 facilities in the current study could be identified among the 339 respondent hospitals in 2007. Among these 20, the changes in Total Score and estimated quality impact were not significant.

By and large, the open-ended questions show that physician executives use the language of quality improvement to describe beliefs about why the program may or may not be working well. Table 1 presents factors which respondents identify as supportive of the peer review program's contribution to clinical quality. Table 2 gives the major factors felt to impair that contribution. Table 3 lists factors felt to explain the degree of medical staff satisfaction with the program.

Only 105 respondents (29%) listed measures of program effectiveness currently in use, while 141 (41%) specified that none are used and 28 (8%) didn't know. These included measures of quality process (59%), patient outcomes (55%), and program efficiency (20%).

Respondent comments regarding why program change might or might not occur in the next year reflect important variation in medical staff and hospital culture:

> *We understand that our process, while pretty good, is not perfect. Our medical staff never figures that it's good enough.*
>
> *Peer review committee continuously strives to improve the process. Currently working on improving trend tracking and reporting to improve individual quality measures.*
>
> *Too much education needs to be done with a staff that is busy and doesn't want to hear it.*
>
> *Burdensome and unproductive regulatory requirements. We will comply with the rules and that will make no difference in actual patient outcomes.*
>
> *There is major mistrust between the medical staff and the board of directors at this hospital, due to lack of communication and secrecy in the administration. The medical staff does not have the*

> *will to make major changes in the process and most appear to belong to the committee out of necessity and not a desire to improve quality. It would take a major leadership initiative to get change to take place, but people are starting to see that the status quo is not good.*

Other things being equal, negative comments regarding the culture and its supports (including inertia or resistance to change, lack of leadership, lack of resources, a fault-finding orientation, or the view that peer review is irrelevant to quality) predict a lower level of estimated quality impact with OR[CI] of 0.45 [0.24-0.82]. In a sub-group analysis of select categories, which were by chance mutually exclusive, the reason for future change likelihood was strongly associated with the 11 item Total Score (R^2=27% by ANOVA, see Figure 2). A lower likelihood of future program change was associated with higher peer review process standardization, change in the current year, the lack of trustee reporting, liking the current process, and negative factors other than fault-finding (R^2=20%).

Discussion

The current study confirms and extends the findings from 2007 to a large, independent sample of physician leaders. The Total Score on the Self-Evaluation Tool is strongly associated with the belief that a peer review program has a significant ongoing impact on the quality and safety of care, as well as with perceived medical staff satisfaction.

The survey items related to structured review and rating scales proved invalid. The implications are different than some might expect. These items were not significant factors in the 2007 regression analysis. They were included in the Self-Evaluation Tool because of strong support from literature and theory. The fault is likely that the reference examples viewed by respondents in the 2007 study were omitted from the Self-Evaluation Tool design due to space constraints and were not reintroduced for purposes of this study. This does not diminish the importance of clinical performance measurement methodology.

The 2007 observation that hospitals are using unreliable methods to “score” the findings from case review still holds true. Moreover, most organizations do not appear to be monitoring and managing the peer review process. Thus, it would seem that, even though physician leaders talk the language of quality improvement, they have yet to apply the fundamental principle of performance measurement to peer review process and program governance. This issue has already been discussed in depth. (8) Physician and hospital leaders who intend to use the original Self-Evaluation Tool need to appreciate what a reliable structured rating form would look like. (9)

The effect of organizational culture on the process and perceived effectiveness of peer review is a new finding, albeit one that may not be surprising. It likely has much in common with the effect of leadership in quality improvement revealed by studies with the Healthcare Leadership Assessment Tool (HLQAT). (10, 11) Further research would be required to fully characterize this factor. Nevertheless, the management literature, both popular and academic, regarding the effects of culture on organizational performance is robust enough to warrant taking this finding seriously. (12-14) Physician and hospital leaders who work in organizations with adverse culture would appear to have much to gain in addressing their problem sooner rather than later.

Despite the high rate of change among peer review programs, it's not clear that it has generated aggregate improvement. The increase in Total Scores from 2007 to 2009 is quite small, even if statistically significant. There are few, if any, A-players. Huge opportunity for improvement remains.

Much of the recent change has concentrated on the replication of a multi-specialty review process. In fact, it seems to be something of a fad. Only 1 published article provides testimonial support for the concept. (15) While there may be merit to this design in terms of reviewer participation, standardization and the ability to address clinician to clinician issues, multi-specialty review is not of itself sufficient to close the gap in program performance.

For those physician and hospital leaders who might desire to improve peer review program effectiveness, this study may raise more questions than it answers. The Self-Evaluation Tool provides a high-level view of a best practice, QI model for peer review. It is the only evidence-based general model available. This needs to be supplemented with additional study to further define and evolve best practice.

The incongruity between physician leader use of the language of quality and the current state of peer review practice should not be minimized. The QA focus on outliers is structurally powerless to substantially affect group performance. In general, improvement science seeks to "shift the curve" of performance and reduce variation by leveraging performance measurement, performance feedback and process improvement. It's time we thought more deeply about what this might mean in the context of peer review.

Implicit in the concept of a QI model for peer review is the possibility of balancing system of care concerns with individual accountability. This is complicated by the recent recognition within the patient safety movement that the pendulum toward "no blame" has swung too far. (16) Peer review has been handicapped by the exclusive focus on individual fault: the presence or absence of substandard care. The judgment of substandard care implies possible incompetence. The question of competence puts professional livelihood at risk, and thereby creates a high stakes game which tends to play out only in the most egregious of cases. As a result, we set the "standard of care" ridiculously low and the large grey zone of borderline performance goes unaddressed.

From a QI perspective, the standard of care judgment, by itself, has little utility. Peer review is not intended to be the court of adjudication for alleged malpractice. Clinical performance is multi-dimensional and subject to variation due to circumstantial factors affecting the performers, including faults in the system of care. In contrast, competence is an enduring quality of the individual, which is unlikely to acutely deteriorate in the absence of a major health problem. The competence evaluation has traditionally come under the purview of the credentials committee and is subject to fairly explicit bylaws provisions and external requirements. Peer review contributes important performance information to the credentialing process, but should not usurp it.

To resolve the tension between "no blame" and accountability, the key question driving case-based peer review should be, "What can we learn from this case to improve clinical performance?" This question opens the door to the exploration of all avenues for improvement at the individual, group and system level. It also makes it easier to enter a collegial dialogue about how things might have been done differently to prevent a recurrence. The dialogue can be enhanced with references to applicable clinical

evidence and guidelines. None of this precludes referral for disciplinary action for those rare instances of willful disregard of patient safety or repeated failure to respond to constructive feedback, particularly when there is strong rationale for the recommendations (e.g., hand washing, Central Line bundles, etc.).

To effectively evaluate clinical performance, peer review needs to include repeated measurement of multiple aspects of clinical performance over time. Isolated point measures are insufficient. Longitudinal measurement enables control-charting at the group level and comparative profiling for individuals. Clinical performance measures can be either explicitly-defined and objective (e.g., complication rates) or implicitly-defined and subjective (e.g., a quality rating given to a physician's admitting assessment). There are well established principles for making subjective performance ratings that could be adapted to the traditional case-based peer review process. (17) Even if we would prefer explicit measures, only a fraction of care delivered falls within their scope. We need both. They provide complementary information. While explicit and implicit measures are made by different mechanisms, the data need to be connected and synthesized at a well- defined point of responsibility.

The scope of peer review activity may need to be adjusted in relation to other organizational processes. For example, even though many organizations have a utilization management committee to meet CMS conditions of participation, the argument could be made to routinely assess resource use during peer review: one function looking at high-level trends and the other at case-specific details. Waste is antithetical to the concept of quality and hospitals are under renewed pressure to control costs. On the other hand, disruptive behavior fits better with the physician health construct and requires specialized expertise for effective management. Only 15% of hospitals included physician health within the scope of peer review in 2007 (5). It would not seem as sensible to attempt to merge these two functions.

This research affirms the importance of reviewer participation to the perceived effectiveness of peer review. Further study will be required to drill-down on the important elements of this factor. Is it really a proxy measure for the rigor of the review process? What should be the qualifications for a reviewer? What training might be required? What should be the definition of a peer? Regardless of its efficacy, the multi-specialty review committee model appears to have successfully challenged the assumption of need for same-specialty review. It is not clear, however, that this boundary can or should be pushed across disciplines. Some hospitals are experimenting with 360-degree evaluations for their medical staff. That experience might serve as a guide.

Given the long history of the QA model, the transition to a QI model may engender unexpected resistance from both reviewers and support staff. Like it or not, physicians are accustomed to judgments about standard of care. Systems thinking is not yet routine. Clinical performance measurement is unfamiliar territory. Unlike acute care medicine, in which the greatest challenge is often in making the diagnosis, for organizational change, it is implementation. Leadership, training and support will be imperative. Sharing stories of both failure and success will be helpful to those in transition.

Readers should appreciate that the data collected for this study were self-reported with limited external validation. The potential for non-response bias could not be directly controlled. Even so, the consistency of findings across 2 large independent populations provides reassurance that the results are generalizable to US hospitals.

The Peer Review Program Self-Evaluation Tool reliably differentiates hospitals along a continuum of program performance, particularly when two or more independent ratings are averaged. If the items related to review form structure and scale are rated with a clear appreciation of best practice, the Tool can serve as a guide to and measuring rod for program improvement. The full potential of peer review as a process for improving the quality and safety of care needs to be explored.

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 and critical reviews of the draft manuscript by Robert Klugman, MD and Mark Keroack, MD.

Reprint Requests

Marc T. Edwards, MD, MBA
QA to QI Patient Safety Organization
8905 Wildwood Links
Raleigh, NC 27613
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 et al. A report card system using error profile analysis and concurrent morbidity and mortality review: surgical outcome analysis, part II. J Surg Res. 2008 Mar 31 [Epub ahead of print PMID: 18511079].

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) Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psych Bull. 1979;86(2)420-428.

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

9) Rubin HR, Rubenstein LV, Kahn KL, Sherwood M. *Guidelines for Structured Implicit Review of Diverse Medical and Surgical Conditions*. Santa Monica, CA: RAND, N-30066-HCFA; 1990.

10) 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.

11) 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.

12) Kotter JP, Heskett JL. *Corporate Culture and Performance*. New York, NY: The Free Press; 1992.

13) Senge PM. *The Fifth Discipline. The Art and Practice of the Learning Organization*. New York, NY: Doubleday; 2006.

14) Gregory BT, Harris SG, Armenakis AA, Shook CL. Organizational culture and effectiveness: A study of values, attitudes, and organizational outcomes. J Business Res. 2009;62:673-679.

15) Agee C. Professional review committee improves the peer review process. Phys Exec. 2007;33(1):52-55.

16) Wachter RM, Pronovost PJ. Balancing “no blame” with accountability in patient safety. N Eng J Med. 2009;361(14):1401-1406.

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

Tables

Table 1. Factors Felt to Support Program Contribution to Quality	
Factor	**n (%)***
Use of clinical performance data	40(14)
Leadership	39(14)
Participation of reviewers	38(13)
Integration with hospital performance improvement	33(12)
Multispecialty review process	33(12)
System issue identification	32(11)
Focus on improvement opportunity, not fault	30(11)
Timely, useful, or balanced feedback	29(10)
Visible program achievements	29(10)
Rigor of evaluation	25(9)
Administrative support	23(8)
Governance of program	20(7)
group learning	17(6)
Standardization of process	15(5)
Culture of excellence/safety	13(5)

*Proportion of 283 responses in which factor was identified

Table 2. Factors Felt to Impair Program Contribution to Quality	
Factor	**n (%)***
Poor participation	44(16)
Culture of resistance to improvement	40(14)
Fault-finding, fear-inducing, or punitive process	21(7)
Lack of standardization or transparency of process	21(7)
Excessive turn-around-time for review	20(7)
Lack of rigor in evaluation	19(7)
Clinical data quality or timeliness issues	18(6)
Failure to address clinical performance issues	16(6)
Lack of administrative support	15(5)
Lack of legal protections	14(5)
Failures of leadership	13(5)
Inadequate integration with hospital performance improvement	13(5)

*Proportion of 282 responses in which factor was identified

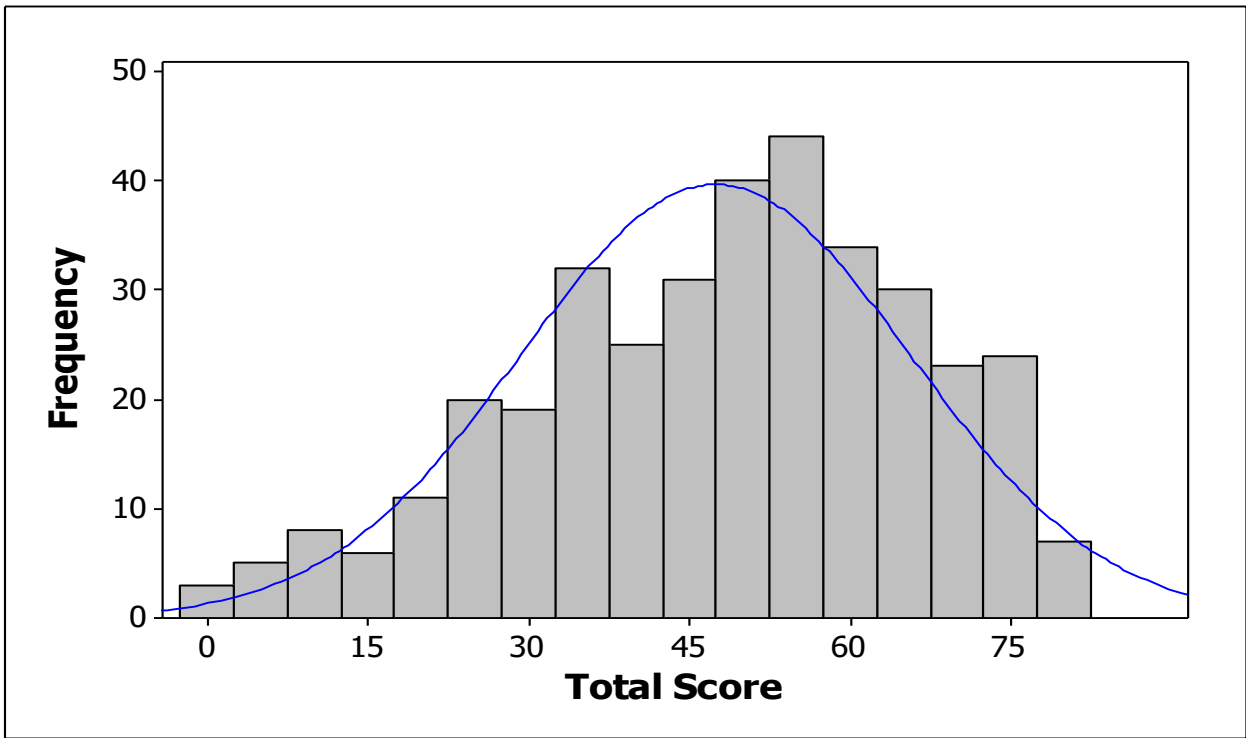
Table 3. Factors Felt to Explain the Level of Medical Staff Satisfaction with the Peer Review Program

Factor	n (%)*
A fair, credible, respected, or consistent process	79(28)
Culture of excellence/safety	69(24)
Focus on improvement opportunity, not fault	66(23)
Participation in, time for, or commitment to program	50(18)
Visible program achievements	24(9)
Unaware or not impacted	22(8)
Leadership active or respected	20(7)
Timely, useful, or balanced feedback	18(6)
Group learning	15(5)
Clinical data quality or presentation	13(5)

* Proportion of 282 responses in which factor was identified

Figures

Figure 1: Distribution of Total Scores for 11 Self-Evaluation Tool Items



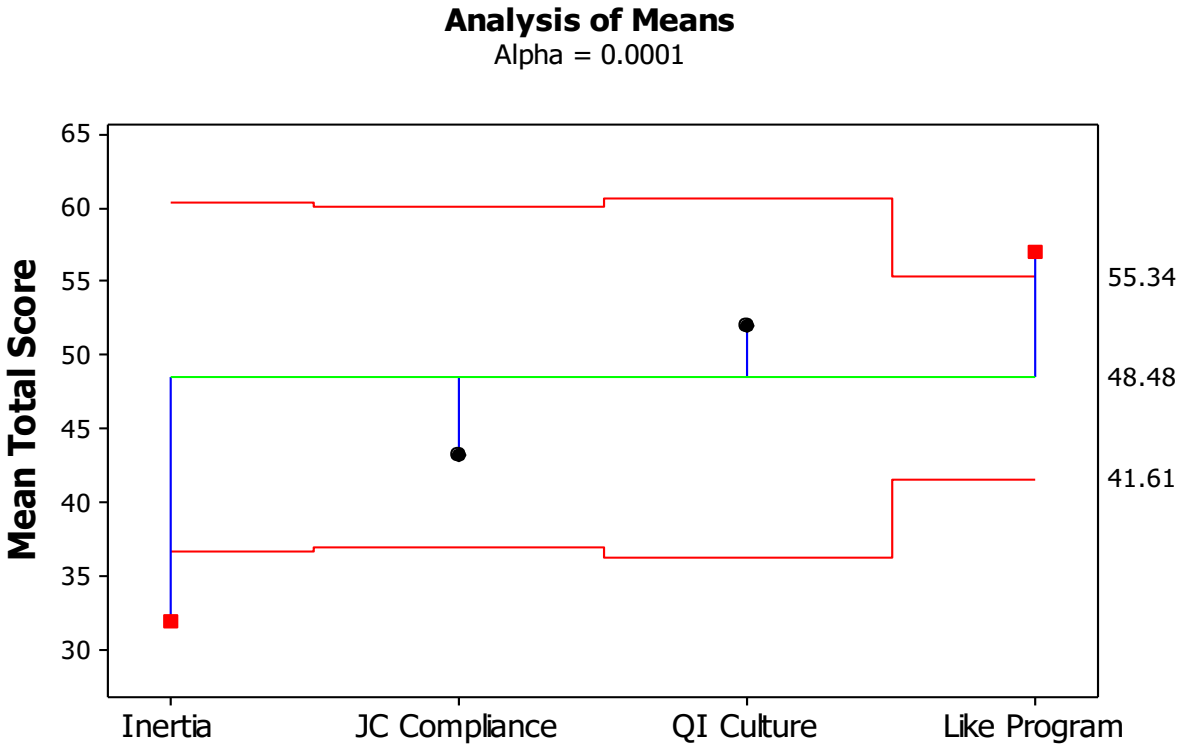


Figure 2: Analysis of Means Graph for 11-Item Total Score by Select, Exclusive Change Likelihood Reason Categories (N=133)