An Audit and Feedback System for Effective Quality Improvement in Head and Neck Surgery: Can we become better surgeons?

Carol M Lewis, MD MPH, **Marcus M Monroe, MD**, **Dianna B Roberts, PhD**, **Amy C Hessel, MD**, **Stephen Y Lai, MD PhD**, and **Randal S Weber, MD**

Department of Head and Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston TX

Abstract

Background—We have established an evaluation system for measuring physician performance. We aimed to determine whether initial evaluation with surgeon feedback improved subsequent performance.

Methods—After evaluation of the initial cohort of procedures (2004-2008), surgeons were given risk-adjusted individual feedback. Procedures in the post-feedback cohort (2009-2010) were then assessed. Both groups were further stratified into high- and low-acuity procedures (HAPs and LAPs, respectively). Negative performance measures included length of peri-operative stay longer than 2 days for LAPs and 11 days for HAPs, peri-operative blood transfusion, return to the operating room within 7 days and readmission, surgical site infection, and mortality within 30 days.

Results—There were 2618 procedures in the initial cohort and 1389 procedures in the post-feedback cohort. Factors affecting performance included surgeon, procedure acuity, and patient co-morbidities. There were no significant differences in the proportion of LAPs and HAPs or in the prevalence of patient co-morbidities between the 2 assessment periods. Mean length of stay significantly decreased for LAPs from 2.1 to 1.5 days (p=0.005) and for HAPs from 10.5 to 7 days (p=0.003). The incidence of having one or more negative performance indicators decreased significantly for LAPs from 39.1% to 28.6% (p<0.001) and trended down for HAPs from 60.9% to 53.5% (p=.081).

Conclusions—Periodic assessment of performance and outcomes are essential to continual quality improvement. Significant decreases in length of stay and negative performance indicators were seen after feedback. We conclude that an audit and feedback system may be an effective means of improving quality of care as well as reducing practice variability within a surgical department.

Corresponding Author: Carol Lewis, MD, MPH, 1515 Holcombe Blvd Unit 1445, Houston, TX 77030; Phone: 713-794-4675; Fax: 713-794-4662; cmlewis@mdanderson.org.

Keywords

performance improvement; head neck cancer; head neck surgery; quality improvement; audit; feedback

Introduction

Over the past 2 decades, healthcare reform has become a national priority, largely spurred by reports from the Institute of Medicine (IOM) emphasizing the need to optimize healthcare quality.[1,2] To achieve this, the IOM proposed 6 aims: to improve healthcare timeliness, effectiveness, safety, efficiency, patient-centeredness, and equity of delivery. Specifically, priority should be placed on care with proven benefit, and ineffective practices should be limited.[1] This emphasis on identifying and implementing best practices has laid the foundation for the development of metrics indicating the quality of patient care and hospital and physician performance. Value in health care delivery is defined as the outcome achieved per unit cost. Enhanced value is achieved when the desired outcome is realized at the lowest cost. Value-based purchasing aligns the goal of seeking care from providers that achieve quality outcomes at the lowest cost.

Performance metrics are now being integrated into pay-for-performance programs, including the Affordable Care Act, with the intent of improving the value of healthcare. There are concerns, however, about the durability of the impact that such financial incentives will have on quality improvement.[3,4] Other approaches to healthcare quality improvement include audit and feedback programs, and workflow and process improvement.

Surgical performance indicators have largely relied on risk-adjusted mortality rates.[5-7] The American College of Surgeons' National Surgical Quality Improvement Program has expanded these metrics, making them applicable to general surgical procedures on institutional and national levels.[8,9] By providing institutional outcomes in the context of national standings, this program has stimulated significant quality improvement.[9,10]

We have previously reported on the program we developed for evaluating surgical performance indicators at an individual and departmental level. Having found that these metrics were significantly associated with the acuity of the procedure, patient comorbidities, and the operative surgeon, we provided feedback to each surgeon in the context of anonymized departmental data. We then collected and analyzed our surgical outcome data for the 2 years following these feedback sessions to determine whether an audit and feedback system is an effective means of motivating surgical quality improvement.

Methods

We previously reported specific outcomes for surgical procedures performed by faculty in the Department of Head and Neck Surgery at the University of Texas MD Anderson Cancer Center between 2004 and 2008.[11] Collected metrics included length of stay (LOS), peri-operative blood product utilization, return to the operating room within 7 days of the index

procedure, and 30-day rates of mortality, hospital re-admission, and surgical site infection. The data sources utilized were previously detailed.[11]

Data were risk-adjusted by procedure acuity and patient comorbidity. Procedures were classified by acuity; high-acuity procedures (HAPs) were tumor extirpations requiring pedicled or free flap reconstruction, whereas low-acuity procedures (LAPs) included outpatient surgeries or those normally requiring an observational hospital stay. Collected comorbidities included diabetes, cardiovascular disease, history of congestive heart failure, emphysema, liver disease, and renal disease.

Performance indicators were the same between HAPs and LAPs except as they pertained to LOS and blood transfusions. LAPs were evaluated for LOS less than 2 days and 2 or more days, whereas HAPs were evaluated for LOS less than 11 days or 11 or more days. LAPs were evaluated for blood product use of one or fewer units of packed red blood cells (PRBCs) and greater than 1 unit, and HAPs were evaluated for requiring 2 or fewer units of PRBCs and greater than 2 units.

Findings from this previous report indicated that the presence of negative performance indicators was significantly associated with procedure acuity, patient comorbidity, and operative surgeon. Each surgeon was presented with his/her individual results in the context of the department as a whole and with other department faculty anonymized. Results were discussed with the senior author (RSW) on an individual basis in a private setting.

To evaluate whether this audit and feedback intervention affected surgical performance, this study was then repeated, evaluating surgical procedures within the department between 2009 and 2010 as post-feedback cohort for comparison to the initial cohort. One difference in methodology involved blood transfusion data; we were no longer able to access the units of PRBCs transfused intra-operatively and relied on medical record review. Because of this change in data source, the period of perioperative transfusion was extended from the previous study to 24 hours (day of surgery) in this study.

The study included all patients identified whose surgical treatment and recorded information met the criteria described above. Descriptive statistics for scaled values and frequencies of study patients within the categories for each of the parameters of interest were enumerated with the assistance of commercial statistical software. Cut-off values for scaled values as possible negative performance indicators were set at the upper quartile of values for the cases in a given category. Possible differences between groups for scaled parameters were assessed by one-way ANOVA followed by the Tukey HSD for unequal N posthoc test for pairwise multivariate analysis. Correlations between categorical parameters were be assessed by Pearson's Chi-square test or, where there are fewer than ten subjects in any cell of a 2×2 grid, by the two-tailed Fisher exact test. Univariate and multivariate models were assessed by logistic regression analysis as appropriate. Statistical significance for all tests was defined as p less than 0.05. The statistical calculations were performed with the assistance of the *Statistica (*StatSoft, Inc., Tulsa, OK, www.statsoft.com) and SPSS (IBM SPSS, IBM Corporation, Somers, NY) statistical software applications.

In order to adjust changes in performance indicators by changes in a surgeon's patient population co-morbidities, we developed a model using observed to expected (O/E) ratios. Observed individual surgeon scores were calculated independently for LAPs and HAPs by dividing the percent of cases experiencing negative indicators by the percent of patients with comorbidities ($NI_{initial}/C_{initial}$); these scores represents the prevalence of negative indicators (NI) per comorbidities (C). This score from the original cohort was used as a baseline score to generate the expected score for the post feedback cohort in the following way:

$$\text{Expected score} = NI_{initial}/C_{initial} + (C_{post-feedback} - C_{initial}) * NI_{initial}/C_{initial}$$

This was then compared to the observed score for the post-feedback cohort ($NI_{post\text{-}feedback}/C_{post\text{-}feedback}$) by generating an O/E ratio. Ratios less than 1 indicate that the surgeon improved in the post-feedback cohort by having fewer negative indicators than would have been predicted based on the change in comorbidity prevalence; ratios greater than 1 indicate that performance declined.

Results

A total of 4007 procedures were evaluated; 2618 in the initial cohort and 1389 in the post-feedback cohort (Table 1). There was no significant difference in the distribution of low and high-acuity cases between the two time periods, with LAPs accounting for 85.3% of cases in the initial cohort and 85.5% of surgeries in the post-feedback cohort (p=0.925). The initial and post-feedback cohorts were likewise balanced with respect to patient comorbidities, with no significant differences noted in the prevalence of patients with 1 or 2 comorbidities (p=0.765 and p=0.241 respectively, Table 2). Physician performance relative to case volume did not reveal a statistically significant association for either LAPs (p=0.570) or HAPs (p=0.281).

HAPs were significantly more likely to be associated with negative performance indicators when compared to LAPs (Table 3). On multivariate analysis, HAPs demonstrated an OR of 2.2, (p<0.001) for 1 negative quality indicator and an OR of 3.8 (p<0.001) for 2 negative quality indicators when compared with LAPs. Individual surgeon performance was also linked to the presence of negative quality indicators. Surgeons with the highest percentages demonstrated an increased rate of negative quality indicators compared to those with the lowest percentages (OR=2.6, p<0.001 for 1 negative quality indicator and OR of 2.2, p<0.001 for 2 negative quality indicators; Table 3, Figures 1-2).

As might be expected, with increasing patient morbidity, the prevalence of negative quality indicators increased (Table 2). The presence of 1 comorbid condition was associated with increased odds of having 1 negative quality indicator on multivariate analysis (OR 1.432, p=0.002), while the presence of 2 comorbidities was significantly associated with increased odds of both 1 and 2 negative indicators (OR 1.388, p=0.031 and OR=2.693, p<0.001 respectively). When comparing the initial and post-feedback cohorts, the initial cohort demonstrated increased odds (OR 1.839, p=0.002) for the presence of 1 negative quality indicator on multivariate analysis.

When comparing the initial and post-feedback cohorts, a significant reduction in LOS was noted in both LAPs and HAPs. In patients undergoing LAPs, the percentage of patients exceeding the 2 day LOS threshold was reduced from 37.1% to 26.5% (p=<0.001), while patients exceeding the 11 day threshold for HAPs was reduced from 28.1% to 19.8% (p=0.028). For HAPs, a reduction in 30-day surgical site infections and 30-day readmission rates were noted (14.1% to 8.4%, p=0.046 and 14.2% to 7.4%, p=0.015 respectively), although the same reduction was not noted in LAPs. RBC usage above the defined threshold was significantly increased in the post-feedback cohorts in both LAPs and HAPs (2.7% to 5.8%, p<0.001 and 30.2% to 43.1%, p<0.001 respectively), although the change in collection of RBC data between cohort intervals clouds interpretation of this data.

The data in figures 1 and 2 facilitate evaluation of changes in performance indicators and of changes in the rate of co-morbid conditions, but the relationship between the two is less clear. We therefore developed an O/E ratio model in order to adjust performance indicators for changes in co-morbidities. Based on the surgeon's original performance, we generated an expected score reflective of a surgeon's expected rate of negative performance indicators given the changes in the rate of his/her patients' co-morbid conditions. An O/E ratio of less than one indicates that the surgeons performed better than expected (fewer negative performance indicators per rate of comorbid conditions) and a ratio of greater than one indicates that a surgeon performed worse than expected (greater negative performance indicators per rate of comorbid conditions). Six of 10 surgeons had improved post-feedback performance for LAPs (table 5 and figure 1) and half of surgeons had improved performance in the post-feedback cohort for HAPs (table 6 and figure 2). Four surgeons improved in the post-feedback cohort for both LAPs and HAPs; the methods of these surgeons can now be examined for best practices that can be shared with their colleagues. The range of O/E ratios was wider for LAPs than HAPs (0.163-3.016 versus 0.248-1.803); however, these data are slight skewed by one surgeon, whose LAP O/E ratio is 3.016. Excluding that surgeon, the LAP O/E ratio ranges from 0.163-1.615, which is similar to the range for HAPs.

Discussion

Improving healthcare quality has become a national priority. There are several different approaches to quality improvement, including financial incentives for meeting specific targets, process and workflow evaluation and interventions, dissemination of information, outreach visits, and audit and feedback programs.[12,13] We previously evaluated performance metrics based on the outcomes of surgical procedures performed by departmental faculty and found a significant association between performance indicators and the operative surgeon.[11] This performance feedback was provided to individual surgeons in the context of anonymized departmental data. Having performed a follow-up audit, we have found that this program of audit and feedback is effective in improving surgical quality.

Audit and feedback programs mostly target improved compliance with a specific process.[14] Previous studies have not found this method to be consistently effective; the variability of impact may derive from a wide variability in study design.[14,15] Published reports of audit and feedback programs differ based on who performs the audit (e.g., internal or external), who provides the feedback (e.g., consultant or supervisor), and what is being measured (e.g.,

performance or outcomes).[12,14,15] Further confounding these issues is that many studies include audit and feedback as part of a larger, multi-faceted intervention.[12,14]

Without a formal audit and feedback program, physicians' self-audits use variable and possibly unreliable data sources.[16] Previous reviews have identified specific study design characteristics that enhance the success of an audit and feedback program. One systematic review found that programs were more effective if the level of baseline performance is low, the person overseeing the audit and feedback is either a colleague or supervisor, feedback is provided more than once in both writing and verbally, and there are clear targets with means of achieving them.[14] Our intervention, unlike many audit and feedback programs, did not have a single target with a specific plan of action. In addition, our baseline performance was good. However, the person responsible for overseeing the audit and delivering the feedback was a colleague (RSW) and feedback was provided both verbally and in writing. Another review that sought to emphasize a model for successful actionable feedback found that timely, individualized, non-punitive, and customizable feedback lead to optimized effects on performance.[15] Our intervention met these criteria.

In evaluating our initial and post-feedback cohorts, we found that the mix of procedure acuity and patient comorbidity was not significantly different (p=0.925 and p=0.765, respectively; Table 2). Additionally, the individual surgeon, patient comorbidities, and procedure acuity all significantly affected the prevalence of negative performance indicators in both cohorts. Overall, we found that the post-intervention cohort had a significantly shorter mean length of stay for LAPs (p=0.005) and HAPs (p=0.003). The incidence of having one or more negative performance indicators decreased significantly for LAPs from 39.1% to 28.6% (p<0.001) and trended down for HAPs from 60.9% to 53.5% (p=0.081). There was no significant difference in the rates of returns to the operating room within 7-days, or 30-day mortality. Surgical site infections and readmissions were significantly reduced in the HAPs (14.1% to 8.4%, p=0.046 and 14.2% to 7.4%, p=0.015 respectively) while no significant difference was noted for LAPs. In looking at surgeon-specific improvement, we found that 6 of 10 surgeons improved for LAPs and half of surgeons improved for HAPs; while this agrees with the net improvement seen departmentally, the implications for individual performance are less clear. While temporally these results suggest a causal link to our feedback intervention, continued monitoring and feedback sessions are needed to confirm this association.

It is possible that surgical faculty, after the feedback sessions, became aware that they were being audited and so made changes to improve their performance; this is also called the Hawthorne effect.[17] Another consideration is that the total period of data collection (2004-2010) coincided with a growing national awareness about the importance of such performance indicators as length of stay, blood transfusion rate, and re-admission rates.[18] This growing awareness could have manifested not only on an individual level but in changes in institutional processes, as well. These potential confounders make it difficult to draw a causal relationship between our intervention and subsequent improvements and these effects may continue to be present in future cycles of audit and feedback; however, performance improvement, independent of exact motivation, only serves to improve the care delivered, which is the end goal.

We are unable to comment on the effect of our audit and feedback program on our rates of blood transfusion; unfortunately, the data source for our initial cohort, which provided intra-operative blood transfusion data, stopped storing these data for our post-feedback cohort. Therefore, these data had to be manually collected from the electronic medical record. With manual extraction, we could no longer pinpoint intra-operative transfusion and the collection period was extended to 24 hours. When we evaluated the rates of blood transfusions in our 2 cohorts, we found a significant increase in the rates of blood transfusion for both LAPs and HAPs in the post-feedback cohort ($p<0.001$ and $p=0.002$, respectively), but cannot determine whether this increase represents an actual increase in intra-operative blood transfusion or a product of an expanded collection timeframe. For future cycles of audit and feedback, however, we will be able to use these post-feedback cohort data for comparison. Lastly, because the variables we collected were categorical and not continuous, we are unable to statistically evaluate for outliers. However, the goal of this project is to motivate individual performance improvement by allowing for a comparison between an individual and his/her peers; the identification of outliers might be viewed as potentially punitive and alienating.

Our audit-and-feedback program was designed and implemented in our department of head and neck surgery, but these metrics are not unique to head and neck surgery and may be applied to any surgical field; operations may be divided into LAPs and HAPs and these metrics applied. Performance metrics are becoming increasingly more important with current on-going healthcare reforms, including the Affordable Care Act, implementing pay-for-performance programs. Most existing pay-for-performance programs incentivize the management of chronic medical conditions;[3,4] the most common publicly available surgical performance metric is peri-operative mortality.[5-7] The American College of Surgeons National Surgical Quality Improvement Program has made significant progress establishing general surgical performance indicators,[9,10] but these data are only available on institutional and national, but not individual, levels. Our data indicate that the individual surgeon affects negative performance indicators, our audit and feedback program caters to surgical improvement on an individual level, which is likely where increased transparency and public reporting of performance metrics is headed. Our program also maintains the capacity to incorporate standardized surgical performance indicators as they are established. Noteworthy was the association between negative performance indicators and comorbid conditions. This finding may have increasing significance if bundled payment schemes for cancer care services become more common. Careful identification and management of comorbid conditions will be critical as providers negotiate reimbursement within a payment bundle given the increased association between negative outcomes and comorbid conditions.

Conclusion

The recent emphasis by health care reform on performance metrics has mandated the development of programs that can measure and track these indicators. We have developed an audit and feedback program within our department that shows promise for effecting quality improvement, although future cycles of audit and feedback are necessary to support a causal link between our intervention and the improvement seen.

Acknowledgments

References

1. Institute of Medicine (IOM). Crossing the Quality Chasm: A New Health System for 21st Century. Washington, D.C: National Academy Press; 2001.

2. Institute of Medicine (IOM). To Err Is Human: Building a Safer Health System. Washington, D.C: National Academy Press; 2000.

3. Langdown C, Peckham S. The use of financial incentives to help improve health outcomes: is the quality and outcomes framework fit for purpose? A systematic review. J Public Health (Oxf). Aug 8.2013

4. Health Policy Brief: Pay-for-Performance. Health Affairs. Oct 11.2012

5. Hannan EL, Kumar D, Racz M, Siu AL, Chassin MR. New York State's Cardiac Surgery Reporting System: four years later. Ann Thorac Surg. Dec; 1994 58(6):1852–1857. [PubMed: 7979781]

6. Chinthapalli K. Surgeons' performance data to be available from July. Bmj. 2013; 346:f3795. [PubMed: 23757757]

7. Brown DL, Clarke S, Oakley J. Cardiac surgeon report cards, referral for cardiac surgery, and the ethical responsibilities of cardiologists. J Am Coll Cardiol. Jun 19; 2012 59(25):2378–2382. [PubMed: 22698492]

8. Khuri SF, Daley J, Henderson W, et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. Ann Surg. Oct; 1998 228(4):491–507. [PubMed: 9790339]

9. Khuri SF. The NSQIP: a new frontier in surgery. Surgery. Nov; 2005 138(5):837–843. [PubMed: 16291383]

10. Khuri SF, Daley J, Henderson WG. The comparative assessment and improvement of quality of surgical care in the Department of Veterans Affairs. Arch Surg. Jan; 2002 137(1):20–27. [PubMed: 11772210]

11. Weber RS, Lewis CM, Eastman SD, et al. Quality and performance indicators in an academic department of head and neck surgery. Arch Otolaryngol Head Neck Surg. Dec; 2010 136(12): 1212–1218. [PubMed: 21173370]

12. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. Jama. Sep 6; 1995 274(9):700–705. [PubMed: 7650822]

13. Oxman AD, Thomson MA, Davis DA, Haynes RB. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. Cmaj. Nov 15; 1995 153(10):1423–1431. [PubMed: 7585368]

14. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2012; (6):CD000259. [PubMed: 22696318]

15. Larson EL, Patel SJ, Evans D, Saiman L. Feedback as a strategy to change behaviour: the devil is in the details. J Eval Clin Pract. Apr; 2013 19(2):230–234. [PubMed: 22128773]

16. Lockyer J, Armson H, Chesluk B, et al. Feedback data sources that inform physician self-assessment. Med Teach. 2011; 33(2):e113–120. [PubMed: 21275533]

17. McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. BMC medical research methodology. 2007; 7:30. [PubMed: 17608932]

18. Weber RS. Improving the quality of head and neck cancer care. Archives of otolaryngology--head & neck surgery. Dec; 2007 133(12):1188–1192. [PubMed: 18086958]

Figure 1.
Physician performance for low acuity procedures as measured by patient comorbidity and the presence of negative indicators for the initial (red) and post-feedback (blue) cohorts. Letters represent individual physicians to track changes in performance.

Comorbidity and Negative Indicators in High Acuity Procedures by Physician
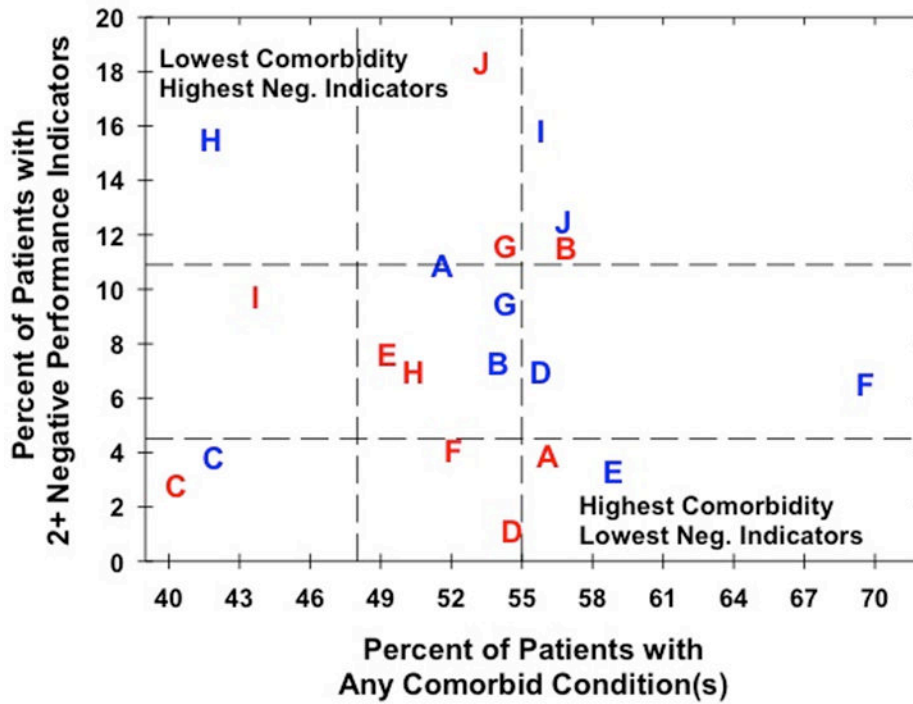### FY 2004-08 & 2009-10
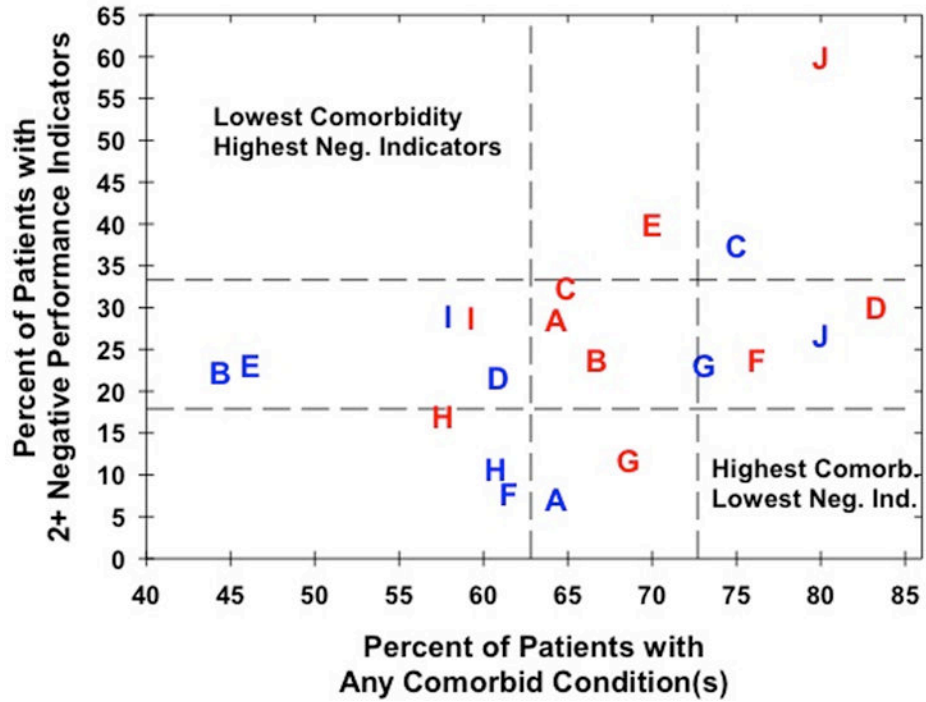
Figure 2.
Physician performance for high acuity procedures as measured by patient comorbidity and the presence of negative indicators for the initial (red) and post-feedback (blue) cohorts. Letters represent individual physicians to track changes in performance.

Table 1

Procedure Type and Acuity for Each Cohort

Procedure Acuity	Initial Cohort (2004-08)	Post-Feedback Cohort (2009-10)	Total
Low Acuity	2234 (85.3%)	1187(85.5%)	3421
High Acuity	384 (14.7%)	202 (14.5%)	586
Total	2618	1389	4007

Percentages were tallied within each column.

Table 2

Comparison of case mix and patient comorbidity between initial and post-feedback cohorts

	Initial Cohort	Post-Feedback Cohort	p
Low Procedure acuity	2234/2618 (85.3%)	1187/1389 (85.5%)	0.925
1 Comorbidity	1372/2618 (52.4%)	735/1389 (52.9%)	0.765
2 Comorbidity	451/2618 (17.2%)	260/1389 (18.7%)	0.241

Table 3

Factors Affecting the Prevalence of Negative Indicators

Univariate Analysis	1 Negative Indicator			2 Negative Indicators		
Factor	**OR**	**95% CI**	**p**	**OR**	**95% CI**	**p**
Procedure (High vs. Low Acuity)	2.555	2.137-3.054	**< 0.001**	4.225	3.364-5.307	**< 0.001**
Surgeon (Highest vs. Lowest Percentages)	2.396	1.945-2.952	**< 0.001**	3.686	2.639-5.147	**< 0.001**
Cohort (Initial vs. Post-Feedback)	1.544	1.347-1.770	**< 0.001**	0.962	0.774-1.196	0.728
1 Comorbid Condition	1.609	1.415-1.830	**< 0.001**	2.318	1.848-2.907	**< 0.001**
2 Comorbid Conditions	1.781	1.512-2.097	**< 0.001**	2.522	2.006-3.170	**< 0.001**

Multivariate Logistic Regression	1 Negative Indicator			2 Negative Indicators		
Factor	**OR**	**95% CI**	**p**	**OR**	**95% CI**	**p**
Procedure (High vs. Low Acuity)	2.214	1.638-2.993	**< 0.001**	3.834	2.607-5.638	**< 0.001**
Surgeon (Highest vs. Lowest Percentages)	2.574	2.088-3.173	**< 0.001**	2.192	1.539-3.121	**< 0.001**
Cohort (Initial vs. Post-Feedback)	1.839	1.471-2.297	**< 0.001**	NS	NS	NS
1 Comorbid Condition	1.432	1.14-1.798	**0.002**	NS	NS	NS
2 Comorbid Conditions	1.388	1.03-1.870	**0.031**	2.693	1.843-3.935	**< 0.001**

Table 4

Comparison of negative performance indicators between initial and post-feedback cohorts

Negative Quality Indicators	Low Acuity			High Acuity		
	Initial (%)	Post-Feedback (%)	p*	Initial (%)	Post-Feedback (%)	p*
Length of Stay > Threshold	829/2234 (37.1%)	315/1187 (26.5%)	<0.001	108/384 (28.1%)	40/202 (19.8%)	0.028
Any RBC Use	61/2234 (2.7%)	69/1187 (5.8%)	<0.001	156/384 (40.6%)	111/202 (55.0%)	0.001
RBC Use > Threshold	61/2234 (2.7%)	69/1187 (5.8%)	<0.001	116/384 (30.2%)	87/202 (43.1%)	0.002
7-Day Return to OR	54/2234 (2.4%)	32/1187 (2.7%)	0.620	35/384 (9.1%)	10/202 (5.0%)	0.075
Died before Discharge	3/1544 (0.2%)	2/462 (0.4%)	0.326	6/384 (1.6%)	0/202 (0.0%)	0.098
30-Day Readmission	83/1308 (6.3%)	30/460 (6.5%)	0.894	54/379 (14.2%)	15/202 (7.4%)	0.015
30-Day Surgical Site Infection	26/2234 (1.2%)	21/1187 (1.8%)	0.148	54/384 (14.1%)	17/202 (8.4%)	0.046
30-Day Mortality	7/2234 (0.3%)	2/1187 (0.2%)	0.728	6/384 (1.6%)	2/202 (1.0%)	0.721
1 Negative Quality Indicator	873/2234 (39.1%)	339/1187 (28.6%)	<0.001	234/384 (60.9%)	108/202 (53.5%)	0.081
2 Negative Quality Indicators	151/2234 (6.8%)	96/1187 (8.1%)	0.153	102/384 (26.6%)	43/202 (21.3%)	0.160

*
Pearson Chi Square test or, if 10 or fewer cases in any cell, 2-tailed Fisher Exact Test

Table 5

Expected and observed individual surgeon score for low-acuity procedures.

Surgeon	Observed Score Initial Cohort	Expected Score Post-Feedback Cohort	Observed Score Post-Feedback Cohort	Observed: Expected Ratio
A	0.070	0.066	0.211	0.314
B	0.202	0.196	0.135	1.452
C	0.069	0.071	0.091	0.778
D	0.020	0.020	0.125	0.163
E	0.154	0.169	0.056	3.016
F	0.079	0.092	0.093	0.990
G	0.214	0.214	0.175	1.221
H	0.139	0.127	0.371	0.342
I	0.222	0.249	0.283	0.879
J	0.343	0.355	0.220	1.615

Table 6

Expected and observed individual surgeon score for high-acuity procedures.

Surgeon	Observed Score Initial Cohort	Expected Score Post-Feedback Cohort	Observed Score Post-Feedback Cohort	Observed: Expected Radtio
A	0.445	0.445	0.110	0.248
B	0.357	0.277	0.500	1.803
C	0.499	0.550	0.500	0.910
D	0.360	0.279	0.356	1.275
E	0.571	0.435	0.500	1.148
F	0.312	0.266	0.125	0.470
G	0.172	0.180	0.316	1.758
H	0.295	0.304	0.176	0.579
I	0.486	0.479	0.501	1.046
J	0.750	0.750	0.334	0.445