Published in final edited form as:

Ann Surg. 2015 May ; 261(5): 920–925. doi:10.1097/SLA.0000000000001032.

Reliability of Surgeon-Specific Reporting of Complications After Colectomy

Terry Shih, MD¹, Adam I. Cole, BS¹, Paul M. Al-Attar, BS¹, Apurba Chakrabarti, BS¹, Hussein A. Fardous, BS¹, Peter F. Helvie, BS¹, Michael T. Kemp, BS¹, Chris Lee¹, Eytan Shtull-Leber, BS¹, Darrell A. Campbell Jr., MD², and Michael J. Englesbe, MD¹

¹Department of Surgery, Academic Surgeon Development Program, University Of Michigan, Ann Arbor, MI, USA

²Michigan Surgery Quality Collaborative, University of Michigan, Ann Arbor, MI, USA

Abstract

Objective—We sought to determine the reliability of surgeon-specific postoperative complication rates after colectomy.

Background—Conventional measures of surgeon-specific performance fail to acknowledge variation attributed to statistical noise, risking unreliable assessment of quality.

Methods—We examined all patients who underwent segmental colectomy with anastomosis from 2008 through 2010 participating in the Michigan Surgical Quality Collaborative (MSQC) Colectomy Project. Surgeon-specific complication rates were risk-adjusted according to patient characteristics with multiple logistic regression. Hierarchical modeling techniques were used to determine the reliability of surgeon-specific risk-adjusted complication rates. We then adjusted these rates for reliability. To evaluate the extent to which surgeon-level variation was reduced, surgeons were placed into quartiles based on performance and complication rates were compared before and after reliability adjustment

Results—A total of 5,033 patients (n=345 surgeons) undergoing partial colectomy reported a risk-adjusted complication rate of 24.5%. Approximately 86% of the variability of complication rates across surgeons was explained by measurement noise, while the remaining 14% represented true signal. Risk-adjusted complication rates varied from 0% to 55.1% across quartiles prior to adjusting for reliability. Reliability adjustment greatly diminished this variation, generating a 1.2 fold difference (21.4%-25.6%). A caseload of 168 colectomies across three years was required to achieve a reliability of >0.7, which is considered a proficient level. Only one surgeon surpassed this volume threshold.

Conclusions—The vast majority of surgeons do not perform enough colectomies to generate a reliable surgeon-specific complication rate. Risk-adjusted complication rates should be viewed with caution when evaluating surgeons with low operative volume, as statistical noise is a large determinant in estimating their surgeon-specific complication rates.

Keywords

colectomy; outcomes; health services research; surgeon-specific reporting

Introduction

Provider profiling is increasingly being used by public and private payers to measure performance, increase accountability, and improve healthcare quality.^{1, 2} Most recently, these profiling methods are being applied to individual surgeons and surgical outcomes.³ These profiling methods should be approached with caution, as inaccurate assessment of surgeon quality could have implications for willingness to perform higher risk operations on severely ill patients, loss of referrals, risk of liability, and issues with public perception.^{4, 5} The ability to confidently compare surgeon-specific outcomes relies on the measures being highly reliable, i.e. high quality surgeons can confidently be distinguished from poor quality surgeons, with minimal chance of misclassification.

Importantly, there is little evidence that traditional risk-adjusted profiling systems can reliably distinguish actual differences in surgeon quality (signal) from measurement error (noise).⁶⁻⁸ Reliability for outcome rates is primarily determined by case volume. Low surgeon case volumes reduce the reliability of outcome measurement. Even at the hospital level, most hospitals do not have high enough case volume to achieve reliabilities above the standard level of 0.7,^{9, 10} leading to high levels of misclassification of hospital performance quality.¹¹⁻¹³ It is therefore questionable whether or not individual caseloads will be large enough to reliably report quality on the individual-surgeon level.

In this context, we sought to determine the reliability of complication rates as a measure of surgeon-specific quality in patients undergoing partial colectomy with anastomosis in the Michigan Surgical Quality Collaborative (MSQC) Colectomy Project. First, we estimated the amount of variation in surgeon complication rates that could be attributed to true differences in surgeon quality and chance variation. We then used hierarchical modeling techniques to adjust complication rates for reliability and analyzed its effect on variation in surgeon-specific complication rates.

Methods

Data Source and Study Population

This study was performed using data from the Michigan Surgical Quality Collaborative (MSQC). This is a regional coalition of 52 teaching and community hospitals across the state of Michigan that uses an audit and feedback system as well as regular meetings and site visits for quality improvement. The program is financially supported by the Blue Cross Blue Shield of Michigan/Blue Care Network. Trained data abstractors prospectively collect patient characteristics, intraoperative processes of care, and 30-day postoperative outcomes from general and vascular surgery patients in accordance with established policies and procedures. The core data available for the present study were based on the American College of Surgeons National Surgical Quality Improvement Program. The data were

collected via the MSQC's Colectomy Project, a subset of the larger MSQC initiative that was started in 2007 to measure colectomy-specific processes and outcomes.¹⁴ Twenty-four hospitals agreed to participate and captured all defined segmental colectomy cases with anastomosis.

Using the MSQC Colectomy Project Data file, we identified all patients 18 years and older who underwent segmental colectomy with anastomosis from January 1, 2008 through December 31, 2010. This sample includes both open and laparoscopic colectomy patients with the following four CPT codes: 44140 (partial colectomy with anastomosis), 44160 (partial colectomy with removal of terminal ileum with ileocolostomy), 44204 (partial colectomy with anastomosis, laparoscopic), and 44205 (partial colectomy with removal of terminal ileum with ileocolostomy, laparoscopic). This procedure was chosen for analysis because it is an intermediately complex procedure performed by both general and colorectal surgeons, thus providing a wide-range of skill sets, complication rates, and caseloads.

Complication Rates

A "complication" was defined as one or more of the following 30-day post-operative occurrences as defined by the MSQC protocol: surgical site infection (superficial, deep, and organ space), wound disruption, urinary tract infection, pneumonia, unplanned intubation, intraoperative or post-operative transfusion, stroke or cerebrovascular accident (CVA), acute myocardial infarction, cardiac arrest requiring cardio pulmonary resuscitation (CPR), pulmonary embolism, deep venous thrombosis (DVT) requiring therapy, acute renal failure, progressive renal insufficiency, coma lasting longer than 24 hours, mechanical ventilation lasting longer than 48 hours, graft/prosthesis failure, peripheral nerve injury, sepsis, septic shock, systemic inflammatory response syndrome (SIRS). In addition, the Colectomy Project recorded postoperative bowel obstruction and anastomotic leak identified via medical chart review by trained data abstractors as additional complications. Anastomotic leak was considered present if it occurred within 30 days of the procedure at the site of the colectomy, requiring antibiotics or more aggressive intervention. MSQC also records reoperation and 30-day postoperative mortality rates. For this present study, we did not include reoperation or 30-day mortality in our definition of a post-operative complication per MSQC protocol. Prolonged ileus was also not included as a post-operative complication.

Complication rates were adjusted using detailed data on patient characteristics. For this risk adjustment, we entered all patient-level variables into a forward step-wise logistic regression where all variables with a p-value of less than 0.05 were retained. Patient variables included age, gender, race, emergency status, procedure type, indication for procedure, body mass index, and other patient co-morbidities. The probability of suffering a complication for each patient is estimated from this model and the probabilities for all patients of each physician are summed to estimate an expected number of complications for each surgeon. The observed number of complications for each surgeon is then divided by the expected number to yield a surgeon-specific observed/expected ratio. This ratio is then multiplied by the overall state-wide average complication rate to yield a risk-adjusted complication rate for each surgeon.

Estimating Reliability

We next determined the reliability of risk-adjusted complication rates for each surgeon. Reliability is analogous to the precision or reproducibility of a point measurement. In our case, reliability describes the proportion of observed variation in partial colectomy complication rates that can be explained by true differences in surgeon quality.

For example, imagine two general surgeons with vastly different case volumes. The low volume surgeon performed 5 colectomies last year with zero complications, while the high volume surgeon performed 100 colectomies last year with 10 complications. Though the low volume surgeon has a lower complication rate (0% vs. 10%), because he/she has only performed 5 colectomies, it is difficult to predict how many complications would arise after the next 5-10 colectomies performed because there is a low sample size to evaluate his/her complications. In contrast, there is more reliability for the high volume surgeon, and we can more reliably predict his/her complication rate over the next 100 colectomies.

To determine reliability, we used the following formula: $\text{reliability} = \text{signal} / (\text{signal} + \text{noise})$.^{9, 15} Reliability ranges from 0 to 1; 0 means that all the variation in surgeon-specific complication rates is the result of measurement error and is not reproducible from year to year, and 1 means that all of the variation is due to true differences in performance and is reproducible. Using hierarchical logistic regression, the signal was defined as the variance of the surgeon random effect. Noise was calculated using standard techniques for determining the standard error of a proportion.

After determining the reliability for individual surgeons, we separated them into 4 roughly equal-sized groups (quartiles) based on caseload and examined the variance in complication rate attributable to noise in each quartile. We then calculated the number of colon resections that must be performed to reach a reliability of 0.7, a standard of acceptable reliability.^{9, 10} To do this, we used the signal that was defined previously as the variance of the surgeon random effect. We determined the maximum noise level, and corresponding sample size, allowed to achieve a reliability of 0.7 given the signal. We then determined the percent of Michigan surgeons whose observed complication rates achieved this minimum level of reliability.

Reliability Adjustment

As described previously,^{11, 12} we next adjusted surgeon complications rates for reliability using empirical Bayes' techniques. Reliability adjustment involves shrinking observed rates toward the overall mean. This is done by weighting the observed rate based on reliability of the measure, and attributing the remaining weight to the overall mean. The magnitude of shrinkage is greater for small surgeons' caseloads.

Going back to our previous example, imagine that the overall complication rate for all surgeons is 15%. The low volume surgeon with only 5 colectomies and a 0% complication rate has very poor reliability, and we know very little about his/her true quality. Therefore, his complication rate after reliability adjustment shrinks greatly towards the mean complication rate for all surgeons (15%). In contrast, the high volume surgeon with 100

colectomies and a 10% complication rate has a higher reliability, and thus his complication rate shrinks very little, and stays closer to 10%.

We compared the variation in surgeon partial colectomy complication rates before and after reliability adjustment in two different ways. First, we graphed the partial colectomy complication rate before and after reliability adjustment. To better graphically illustrate the effect of reliability adjustment, we chose to focus on a subset of 30 randomly sampled surgeons rather than all 345 surgeons. Second, we sorted surgeons, from lowest to highest, into 4 equal sized groups (quartiles) of surgeons based on partial colectomy complication rate and compared the mean complication rates for the highest and lowest quartiles before and after reliability adjustment. All statistical analysis was done using Stata 12.0 (College Station, TX).

Results

We analyzed the 5,033 patients undergoing partial colectomy with anastomosis performed by 345 surgeons. Baseline patient demographics, functional status, and co-morbidities are shown in Table 1. The major indication for surgery was neoplasm (58.0%), and the majority of patients received open surgery (59.0%) [CPT codes 44140 and 44160].

The average complication rate was 24.5% and the reliability of complication rates based on caseload can be found in Table 2. The reliability increased as caseload increased, with the top quartile having an average reliability of 0.36. Caseload for the top quartile of physicians ranged from 18-171 cases. The number of cases needed to reach a proficient level of reliability of 0.70 in the three-year span we analyzed is 168 cases. Only one surgeon performed enough cases to reach this reliability threshold. The relationship between surgeon caseload and reliability is displayed in Figure 1.

There was dramatic variation in risk-adjusted complication rates across surgeons. After reliability adjustment, variation in the complication rates across surgeons was greatly diminished. The impact of reliability adjustment on complication rates for a randomly selected 30 individual surgeons is shown in Figure 2. After reliability adjustment, complication rates for each surgeon converge toward the mean (24.5%).

After stratifying surgeons into quartiles based on complication rates, reliability adjustment reduced variation. Complication rates varied from 0% to 55.1% across quartiles of performance prior to adjusting for reliability. Reliability adjustment greatly diminished this variation, generating a 1.2 fold difference (21.4%-25.6%).

Discussion

The rate of postoperative complications is often used as an indicator of the quality of a surgeon. In this study, we demonstrate that the use of postoperative complication rates is a poor measure of surgical quality at the surgeon level. A large fraction of the surgeon level variation was attributed to noise, which is often not accounted for during risk-adjusted analysis. In order to reduce the noise to an acceptable rate, a high volume of cases must be performed, which only one surgeon in our cohort managed to achieve. One countermeasure

to account for noise in measuring surgeon-level complication rates is reliability-adjustment. When compared with risk-adjusted complication rates, reliability adjustment rates reduced surgeon wide variation.

Previous studies have recognized the utility of reliability-adjusted rates for quality measurement. Dimick et al examined the importance of reliability adjustment when comparing mortality rates across hospitals for coronary artery bypass grafting.¹³ Reliability adjustment decreased variation greatly, from an 8-fold difference between “best” and “worst” hospitals before reliability adjustment to a 2-fold difference between “best” and “worst” hospitals after adjustment.¹³ Moreover, reliability adjustment was also shown to be a better predictor of future performance than traditional risk-adjustment for uncommon operations, such as abdominal aortic aneurysm repair and pancreatic resection.¹³ Osborne et al demonstrated that after reliability adjustment for vascular procedures, 43% of hospitals were misclassified in the wrong quartile of performance using traditional risk-adjustment.¹¹ Dimick et al also found similar results in a separate analysis for hospital mortality rates after colon resection – 44% of the “best” hospitals were reclassified after reliability adjustment.¹² Although important in their findings, most of the previous work in reliability adjustment in surgery has been at the hospital-level and not the surgeon-level.^{11-13, 16}

The concept of reliability has been evaluated for physician specific reporting in the primary care setting. These studies have shown similar results for a variety of different outcome measures. Hofer et al evaluated the reliability of a set of physician performance measures for diabetic care and also discovered that the majority of the variation (60%) between individual physicians was due to statistical noise.⁶ Scholle et al examined the reliability of a set of 27 commonly measured quality indicators for a sample of primary care physicians.¹⁷ They demonstrated that most physicians do not have adequate numbers of quality events to support reliable quality measurement. Cost measures have been explored as well on the provider level. In a study on the reliability of physician cost profiling involving 28 different specialties, 7 of them being surgical subspecialties, Adams et al. found that the majority (59%) of physicians had cost-profile scores with a reliability of less than 0.70.⁹ This suggested that traditional methods of profiling physicians with respect to costs and services may lead to misleading results. Additionally, the majority of surgical specialties had reliabilities that were on average lower than the medical specialties, with vascular surgery and cardiothoracic surgery being the lowest.

Our study adds to the existing body of work by analyzing the reliability of an outcome measure commonly evaluated in surgery at the provider level. Despite current surgical quality initiatives like the American College of Surgeons National Surgical Quality Improvement Program, these data sources often report only group performances as opposed to physician-specific data. Parsing data further at the individual provider level may have potential limitations. Qualitative research has demonstrated concerns that surgeon-specific reporting of outcomes may fail to acknowledge limitations due to insufficient sample size.⁵ Moreover, literature regarding physician-specific reporting has stressed the need for adjustment of quality measures in all aspects of medicine in order to produce more statistically valid data.^{17, 18} This is the first study to our knowledge that explores the reliability of surgeon-specific outcomes rates.

Our study has several limitations. First, MSQC is a statewide collaborative, and we are unable to assure that our data are representative of a national cohort. Second, as with any registry data, there may be errors in data abstraction and reporting. There may be additional complications that are not captured in MSQC data. The complications captured in MSQC mirror those collected and analyzed by the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP).¹⁹ We utilized CPT codes to determine laparoscopic and open cases. It is possible that some of the cases with laparoscopic CPT codes were converted to open and were thus misclassified. Third, no risk adjustment is perfect. There may be additional risk factors not collected that influence the development of complications, including socioeconomic status that are not included in the MSQC dataset. Though enhanced risk adjustment may improve the signal for reliability, this effect would likely be minimal given the large noise component for reliability given the low case volumes for the majority of surgeons. Fourth, our study only examines patients who had undergone partial colectomy with primary anastomosis, and not a broader surgical population. Although broadening our cohort to all colectomy patients would have yielded a larger sample size, we chose a more homogenous population to decrease variation in outcome rates due to patient factors. Including other procedures in our patient cohort would likely further decrease the reliability of complication rates, which we have already demonstrated to be unreliable. Finally, an evaluation of the reliability of post-operative mortality was not included in the study. Statistical reliability is determined by event rate and sample size. The event rate of mortality and individual surgeon volumes were too low for any individual surgeon to achieve a reliable mortality rate.

Our study has important implications. Our analysis suggests evaluating surgeons individually for their outcomes may be statistically unreliable, particularly for surgeons with low case volumes. Nonetheless, in the current health care climate promoting increased transparency and quality measurement, hospitals and individual surgeons are increasingly being scrutinized for their outcomes. There are possible solutions to reconcile this tension. First, surgeons that fall below a certain volume threshold could be compared only with other surgeons with similar case volumes. Another countermeasure to combat poor reliability would be to adjust surgeon-specific rates using hierarchical modeling and empirical Bayes' techniques as we have done in our analysis. Third, a potential solution would be to develop composite measures that incorporate multiple outcome and process measures to evaluate surgeons.

In conclusion, we have demonstrated that risk-adjusted rates fail to serve as reliable measure of surgical quality in clinicians with low operative volume, as statistical noise is a large determinant in surgeon specific complication rates. Utilizing surgeon-specific complication rates to evaluate individual performance is unreliable and can result in unwarranted repercussions including avoidance of high-risk procedures in severely ill patients, loss of referrals, and issues with public perception.

Acknowledgments

This study was supported by a grant to Dr. Shih from the National Institutes of Health (5T32HL07612309). The views expressed herein do not necessarily represent the views of the United States Government.

We would also like to thank Greta L. Krapohl, PhD, RN for her assistance in the acquisition of data necessary to perform this study.

Funding: This study was supported by a grant to Dr. Shih from the National Institutes of Health (5T32HL07612309). The views expressed herein do not necessarily represent the views of the United States Government.

REFERENCES

1. Guidelines for Reporting Physician Data. American Medical Association; 2012.
2. Longo DR, Daigird AJ. Measuring the quality of care: reforming the health care system. *Am J Med Qual.* 1994 Fall;9(3):104–115. [PubMed: 7950482]
3. Birkmeyer NJ, Birkmeyer JD. Strategies for improving surgical quality--should payers reward excellence or effort? *N Engl J Med.* Feb 23; 2006 354(8):864–870. [PubMed: 16495401]
4. Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care. A survey of cardiovascular specialists. *N Engl J Med.* Jul 25; 1996 335(4):251–256. [PubMed: 8657242]
5. Yi SG, Wray NP, Jones SL, et al. Surgeon-Specific Performance Reports in General Surgery: An Observational Study of Initial Implementation and Adoption. *J Am Coll Surg.* Jun 29; 2013
6. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician “report cards” for assessing the costs and quality of care of a chronic disease. *JAMA.* Jun 9; 1999 281(22):2098–2105. [PubMed: 10367820]
7. Green J, Wintfeld N. Report cards on cardiac surgeons. Assessing New York State’s approach. *N Engl J Med.* May 4; 1995 332(18):1229–1232. [PubMed: 7700321]
8. Iezzoni LI. ‘Black box’ medical information systems. A technology needing assessment. *JAMA.* Jun 12; 1991 265(22):3006–3007. [PubMed: 2033775]
9. Adams JL, Mehrotra A, Thomas JW, McGlynn EA. Physician cost profiling--reliability and risk of misclassification. *N Engl J Med.* Mar 18; 2010 362(11):1014–1021. [PubMed: 20237347]
10. Safran DG, Karp M, Coltin K, et al. Measuring patients’ experiences with individual primary care physicians. Results of a statewide demonstration project. *J Gen Intern Med.* Jan; 2006 21(1):13–21. [PubMed: 16423118]
11. Osborne NH, Ko CY, Upchurch GR, Dimick JB. The impact of adjusting for reliability on hospital quality rankings in vascular surgery. *J Vasc Surg.* 2011; 53(1):1–5. [PubMed: 21093202]
12. Dimick JB, Ghaferi AA, Osborne NH, Ko CY, Hall BL. Reliability adjustment for reporting hospital outcomes with surgery. *Ann Surg.* 2012; 255(4):703–707. [PubMed: 22388108]
13. Dimick JB, Staiger DO, Birkmeyer JD. Ranking hospitals on surgical mortality: the importance of reliability adjustment. *Health Serv Res.* 2010; 45(6 Pt 1):1614–1629. [PubMed: 20722747]
14. Englesbe MJ, Brooks L, Kubus J, et al. A statewide assessment of surgical site infection following colectomy: the role of oral antibiotics. *Ann Surg.* Sep; 2010 252(3):514–519. discussion 519-520. [PubMed: 20739852]
15. Adams, JL. The Reliability of Provider Profiling. RAND Corporation; 2009.
16. Kao LS, Ghaferi AA, Ko CY, Dimick JB. Reliability of superficial surgical site infections as a hospital quality measure. *J Am Coll Surg.* 2011; 213(2):231–235. [PubMed: 21622011]
17. Scholle SH, Roski J, Adams JL, et al. Benchmarking physician performance: reliability of individual and composite measures. *Am J Manag Care.* Dec; 2008 14(12):833–838. [PubMed: 19067500]
18. Weifeng W, Hess BJ, Lynn LA, Holmboe ES, Lipner RS. Measuring physicians’ performance in clinical practice: reliability, classification accuracy, and validity. *Eval Health Prof.* Sep; 2010 33(3):302–320. [PubMed: 20801974]
19. Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. *Ann Surg.* Sep; 2009 250(3):363–376. [PubMed: 19644350]

Figure 1.

Relationship between reliability of surgeon-specific complication rates after segmental colectomy and surgeon case volume based on Michigan Surgical Quality Collaborative data for 2008 to 2010. Each black circle represents one surgeon. As shown, only one surgeon performed enough cases to reach the reliability threshold of 0.7.

Figure 2.

Risk-adjusted complication rates for colon resection, before and after adjustment for reliability, among 30 individual surgeons (sampled randomly) from the Michigan Surgical Quality Collaborative. Each pair of circles represents one surgeon. Reliability adjustment diminished the variation in complication rates among surgeons.

Figure 3. Risk-adjusted complication rates across surgeon quartiles before and after adjustment for reliability. Surgeons were stratified into quartiles based on risk-adjust complication rates.

TABLE 1

Characteristics of patients undergoing partial colectomy in a three-year span between January 1, 2008 – December 31, 2010.

Baseline Demographics	Number (%)
n = 5,033	
Age category, yrs	
50 or younger	921 (18.3%)
51-65	1571 (31.2%)
66-80	1743 (34.6%)
81 or older	798 (15.9%)
Race	
White	3938 (78.2%)
Black	555 (11.0%)
Other	540 (10.7%)
Male	2376 (47.2%)
Body Mass Index, mean (SD)	28.4 (7.7)
Emergency Operation	522 (10.4%)
Laparoscopic	2064 (41.0%)
Primary indication for operation	
Neoplasm	2920 (58.0%)
Diverticular Disease/Ischemia	1390 (27.6%)
Inflammatory Bowel Disease	198 (3.9%)
Obstruction/Volvulus	306 (6.1%)
Other	219 (4.4%)
Functional Status	
Independent	4675 (92.9%)
Partially Dependent	297 (5.9%)
Totally Dependent	61 (1.2%)
ASA * Class	
1 or 2	2491 (49.6%)
3	2189 (43.5%)
4 or 5	350 (7.0%)
Coronary Artery Disease	691 (13.7%)
Congestive Heart Failure	65 (1.3%)
Hypertension	2849 (56.6%)
Diabetes	
None	4160 (82.7%)
Non-insulin dependent diabetes mellitus	603 (12.0%)
Insulin dependent diabetes mellitus	270 (5.4%)
Cerebrovascular Disease	453 (9.0%)
Smoker	987 (19.6%)
COPD **	328 (6.5%)

Baseline Demographics	Number (%)
Preoperative steroid therapy	209 (4.2%)
Renal Failure	67 (1.3%)
Ascites	56 (1.1%)
Weight loss of >10% in previous 6 months	197 (3.9%)
Disseminated Cancer	165 (3.3%)
Preoperative Chemotherapy	41 (0.8%)
Transfusion >4 units pRBC*** within 72 hours of surgery	60 (1.2%)
Preoperative Sepsis	
None	4576 (90.6%)
SIRS	285 (5.7%)
Sepsis	139 (2.8%)
Septic Shock	48 (1.0%)

* American Society of Anesthesiologists

** Chronic Obstructive Pulmonary Disease

*** Packed red blood cells

TABLE 2

Reliability and noise values of physicians based on quartiles of caseload

Caseload Quartile	Number of Cases	Number of Surgeons	Total Number of Patients	Risk-adjusted Complications (mean 24.5%)	
				Reliability*	Noise*
1 st	1-2	108	143	0.02	98%
2 nd	3-6	75	320	0.07	93%
3 rd	7-17	80	902	0.15	85%
4 th	18-171	83	3668	0.36	64%
Overall	1-171	345	5033	0.142	86%

* Reliability and Noise are average values for physicians in each quartile