

# Accuracy of National Surgery Quality Improvement Program Models in Predicting Postoperative Morbidity in Patients Undergoing Colectomy

Jeffrey A Neale, MD, FACS, FASCRS; Craig Reickert, MD, FACS, FASCRS; Andrew Swartz; Subhash Reddy, MBBS; Maher A Abbas, MD, FACS, FASCRS; Ilan Rubinfeld, MD, MBA, FACS

Perm J 2014 Winter;18(1):14-18

<http://dx.doi.org/10.7812/TPP/12-133>

## Abstract

**Background:** The National Surgery Quality Improvement Program (NSQIP) is the standard for assessment of acuity-adjusted outcomes in surgery. The validity of NSQIP has not been well established in colorectal surgery. Technical and process variables, which NSQIP may not consider, affect morbidity rate.

**Objective:** A retrospective observational study was undertaken to determine the accuracy of NSQIP models in predicting morbidity for patients undergoing laparoscopic or open colectomy.

**Methods:** NSQIP participant use files for 2005 to 2008 were obtained. Data were selected using *Current Procedural Terminology* coding for open or laparoscopic colectomy. NSQIP-generated predicted morbidities were used to create area under the receiver operator curves (AUROCs).

**Results:** AUROCs demonstrated an accurate predictive model if the value was above 0.8 and indicated a marginal predictor mode if below 0.7. The AUROC for the general NSQIP model was 0.817 (confidence interval [CI] = 0.815-0.819,  $p < 0.001$ ). AUROC for the combined laparoscopic and open colectomy group was 0.703 (CI = 0.698-0.709,  $p$  value  $< 0.001$ ). AUROCs for the individual laparoscopic and open colectomy groups were 0.627 (CI = 0.615-0.640,  $p < 0.001$ ) and 0.701 (CI = 0.695-0.707,  $p < 0.001$ ).

**Conclusion:** This study demonstrates that although NSQIP-generated morbidities used to create AUROCs are accurate for patients in an overall surgical model, predictive models for morbidity are marginal for laparoscopic and open abdominal colectomies. NSQIP risk models tend to emphasize comorbidities rather than intraoperative details or technical aspects of colonic resections.

## Introduction

In 1994, the Veterans Health Administration (VHA) established the National Surgical Quality Improvement Program (NSQIP) for monitoring and improving the quality of surgical care across all VHA medical centers where major surgery is performed. The impact of NSQIP on quality of care was substantial, with a 47% decrease in the 30-day

postoperative mortality and a 43% reduction in postoperative complications.<sup>1</sup> The implementation of NSQIP in VHA hospitals demonstrated that systematic collection, analysis, and feedback of risk-adjusted surgical data could lead to improved outcomes.<sup>1</sup> NSQIP collects 250 preoperative, intraoperative, and 30-day postoperative variables to quantify 30-day risk-adjusted surgical outcomes.

This prospective, peer-controlled, and validated database includes 95% of the data points.<sup>2</sup> The data represent a sample of institutional operative cases. Data are collected by specially trained nurse coordinators and validated by standard methods to ensure reliable comparison between institutions. This approach is gaining wide acceptance and is rapidly becoming a standard for measuring and improving quality of care for general, vascular, and colon and rectal surgery practices in many health care institutions in the US. Initiatives have been undertaken to broaden the implementation of NSQIP in additional surgical subspecialties, including gynecology, orthopedics, and neurosurgery; the “multispecialty” hospital membership includes these surgical subspecialties and more. In the colon and rectal surgical realm, NSQIP has had an impact on decreasing surgical site infections (SSI) and has been used to study the impact of a laparoscopic or open approach on the frequency of SSI.<sup>3,4</sup> Fleming and colleagues<sup>5</sup> recently used NSQIP data to demonstrate that a laparoscopic approach for restorative proctocolectomy was associated with a statistically significant reduction in both minor and major postoperative complications compared with the traditional open approach.

Currently, risk models for morbidity and mortality are adjusted each year,

**Jeffrey A Neale, MD, FACS, FASCRS**, is a Colorectal Surgeon in the Department of Surgery in the Division of Colon and Rectal Surgery at the Lee Memorial Health System in Fort Meyer, FL. E-mail: [Jeffneale@hotmail.com](mailto:Jeffneale@hotmail.com). **Craig Reickert, MD, FACS, FASCRS**, is a Colorectal Surgeon in the Division of Colon and Rectal Surgery at the Henry Ford Hospital in Detroit, MI. E-mail: [creicke1@hfhs.org](mailto:creicke1@hfhs.org). **Andrew Swartz** is a Medical Student/Research Assistant in the Department of Surgery at the Henry Ford Hospital in Detroit, MI. E-mail: [answartz@med.wayne.edu](mailto:answartz@med.wayne.edu). **Subhash Reddy, MBBS**, is a General Surgery Resident in the Department of General Surgery at the Cleveland Clinic Main Campus in Cleveland, OH. E-mail: [sreddydr@gmail.com](mailto:sreddydr@gmail.com). **Maher A Abbas, MD, FACS, FASCRS**, is the Chair of the Digestive Disease Institute of Cleveland Clinic Abu Dhabi in the United Arab Emirates and an Associate Professor of Surgery at the Cleveland Clinic Lerner College of Medicine of Case Western Reserve University in OH. E-mail: [abbasm@clevelandclinicabudhabi.ae](mailto:abbasm@clevelandclinicabudhabi.ae). **Ilan Rubinfeld, MD, MBA, FACS**, is the Director of the Surgical Intensive Care Units, Trauma Surgeon, and Associate Professor in the Department of Surgery in the Division of Trauma and Acute Care Surgery at the Henry Ford Hospital and the Center for Health System Research at the Henry Ford Health System in Detroit, MI. E-mail: [irubmd@gmail.com](mailto:irubmd@gmail.com).

and institutional outcomes are based on the acuity-adjusted observed-to-expected ratios. These models' operative results are highly predictive when applied to the general population of NSQIP. For any of these models, accuracy can be judged by 2 components: its ability to separate diseased from nondiseased (discrimination) and its ability to correctly estimate the risk (calibration). The area under the receiver operating characteristic curve (AUROC) is one of the most common means of measuring discrimination. The AUROC and its associated c-statistic are functions of the sensitivity and specificity for each value of the measure or model. Because specificity and sensitivity can be manipulated on the basis of threshold choice, the c-statistic allows one to balance the view of the predictive model across the various metrics. The c-statistic value can range from 0.5 (no predictive ability) to 1 (perfect discrimination). The AUROC and its c-statistic are optimized semiannually to ensure accurate risk adjustment for reliable interinstitutional comparison. These models tend to favor demographic and comorbidity data, as these are common to all procedures. Another tool that one could use to evaluate goodness of fit in logistic regression is the Hosmer-Lemeshow test. However, this test cannot be used for large datasets such as ours because "[a]s with any statistical test, the power increases with sample size; this can be undesirable for goodness of fit tests because in very large data sets, small departures from the proposed model will be considered significant."<sup>6</sup> Given NSQIP's need to gather a dataset common to all procedures, there are no specific colon and rectal data points collected. Despite proven broader surgical and specific colon and rectal predictive benefits, current NSQIP risk models are slightly better at predicting mortality than morbidity. In a review of semiannual reports, both mortality and morbidity are accurately predicted, with c-statistics on the AUROC curve of 0.94 (range = 0.85-0.87). We hypothesized that these models tend to emphasize comorbidity data rather than intraoperative details and technical aspects of surgery, and therefore are not solely reliable in predicting the outcome of patients undergoing colectomy.

**Table 1. Current Procedural Terminology codes**

Procedure	Codes	
Open colectomy	44139	44140
	44141	44143
	44144	44145
	44146	44147
	44150	44151
	44152	
Laparoscopic colectomy	44204	44205
	44206	44207
	44208	44210
	44211	44212
	44213	44215

### Materials and Methods

NSQIP participant use files were obtained under a data use agreement of the American College of Surgeons, and the study was approved by the Henry Ford Health institutional review board. We evaluated the most recent 4 years available at the time of analysis, January 1, 2005 to December 31, 2008. Patients were selected using Current Procedural Terminology (CPT) coding for major colectomy and labeled as either open or laparoscopic. For open colectomy and laparoscopic colectomy, the CPT codes are listed in Table 1. The noncolectomy group was defined as patients undergoing procedures other than those listed under open and laparoscopic colectomy. Postoperative morbidity was defined as the occurrence of 1 or more of the following events: SSI (superficial, deep, or organ space), wound disruption, pneumonia, unplanned intubation, pul-

monary embolism, mechanical ventilation longer than 48 hours, renal insufficiency, acute renal failure, urinary tract infection, stroke or cerebrovascular accident, coma lasting longer than 24 hours, peripheral nerve injury, cardiac arrest requiring cardiopulmonary resuscitation, myocardial infarction, bleeding transfusions, graft/prosthesis/flap failure, deep vein thrombosis or thrombophlebitis, sepsis, and septic shock. It should be noted that each of these points, even though not directly applicable to colectomy surgery, are part of the standard NSQIP adverse events that all NSQIP surgical clinical reviewers look for. NSQIP-generated predicted morbidities were then used to create AUROCs for the various populations: all of NSQIP, noncolon-related surgeries, all colectomies, laparoscopic colectomies, and open colectomies. AUROC (a curve generated by the modeling process, the c-statistic gives you an objective understanding if that curve is a good one) is defined as the probability that predicting the outcome is better than that of chance.<sup>7</sup> The c-statistic can range from 0.5 (no predictive ability) to 1 (perfect discrimination). AUROCs were judged by the c-statistic: < 0.70 (no clinical utility), 0.70 to 0.79 (marginal clinical utility), 0.80 to 0.89 (adequate clinical utility), and greater than 0.90 (excellent clinical utility).<sup>8</sup> All analyses were verified using segmentation and subset methods. Data were analyzed using statistical analysis software (SPSS version 19, IBM SBSS, New York, NY), and  $p < 0.05$  was considered significant.

**Table 2. Findings by group**

Characteristic	All NSQIP procedures (N = 635,265) <sup>a</sup>	Noncolectomies (n = 589,620) <sup>a</sup>	All colectomies (n = 45,645) <sup>a</sup>	Laparoscopic (n = 12,455) <sup>a</sup>	Open (n = 33,190) <sup>a</sup>
Mean age, years	55.1	54.5	62.1	60.0	62.9
Male, %	42.6	42.2	48.1	48.4	47.9
Emergency procedure, %	12.9	12.4	18.6	3.6	24.2
Mean relative value unit	15.6	14.8	25.6	28.1	24.7
ASA 1, %	10.4	11.0	3.3	5.0	2.7
ASA 2, %	45.6	45.7	44.4	59.6	38.7
ASA 3, %	36.7	36.3	41.5	32.3	45.0
ASA 4, %	6.6	6.4	9.8	2.9	12.4
ASA 5, %	0.3	0.3	0.9	0.1	1.1
Average predicted morbidity	0.12	0.11	0.24	0.17	0.26
Actual morbidity occurrence	0.14	0.13	0.14	0.18	0.35

<sup>a</sup> Significant at  $p < 0.001$ .

ASA = American Society of Anesthesiologists class; NSQIP = National Surgical Quality Improvement Program.

Colectomy emergency status	N	AUROC <sup>a</sup>	Confidence interval	p value	Corresponding figure
<b>Emergency</b>					
Morbidity	4408	0.73	0.72-0.74	< 0.001	7
Mortality	1295	0.86	0.85-0.87	< 0.001	8
<b>Elective</b>					
Morbidity	9376	0.64	0.64-0.65	< 0.001	9
Mortality	777	0.88	0.87-0.89	< 0.001	10

<sup>a</sup> AUROC value > 0.8 (accurate predictive model); value ≤ 0.7 (marginal predictive model).  
AUROC = area under the receiver operator characteristic curve.

Group	N	AUROC <sup>a</sup>	Confidence interval	p value	Corresponding figure
All patients	635,265	0.817	0.815-0.819	< 0.001	5
Noncolectomy	589,620	0.816	0.814-0.818	< 0.001	6
Open colectomy	33,190	0.701	0.694-0.707	< 0.001	3
All colectomies	45,645	0.702	0.697-0.708	< 0.001	4
Laparoscopic colectomy	12,455	0.627	0.619-0.647	< 0.001	2

<sup>a</sup> AUROC Value > 0.8 is the accurate predictive model; value ≤ 0.7 is a marginal predictive model.  
AUROC = area under the receiver operator characteristic curve.

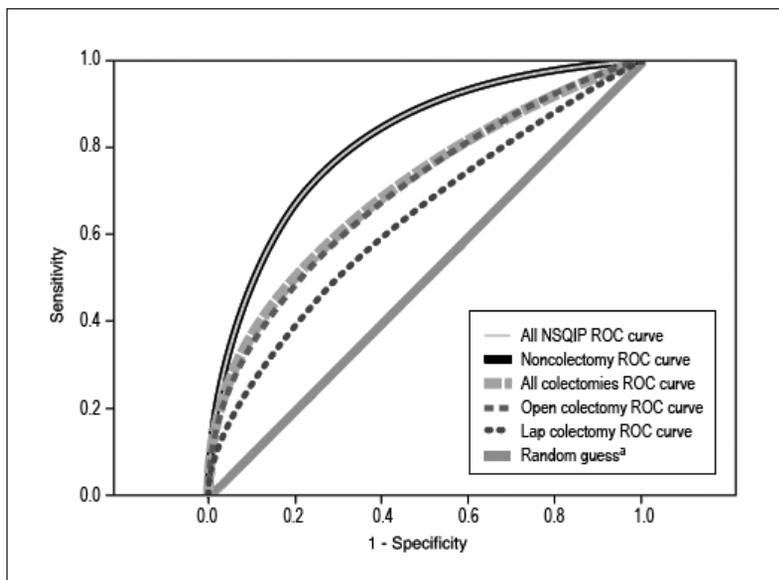


Figure 1. Comparison of receiver operator characteristic (ROC) curves.

<sup>a</sup> The random guess line, also known as the line of no discrimination, represents the strategy of randomly guessing. That is, the result is no more accurate than a random guess, for example, heads or tails. The closer the sample curve approaches the random guess line, the less accurate the test. The closer the sample line approaches the left and upward direction (toward y = 1), the more accurate the test.<sup>1,2</sup> Diagonal segments are produced by ties.

NSQIP = National Surgical Quality Improvement Program.

- Harrell FE Jr. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York, NY: Springer-Verlag New York, Inc; 2001.
- Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007 Feb 20;115(7):928-35. DOI: <http://dx.doi.org/10.1161/CIRCULATIONAHA.106.672402>.

## Results

The general NSQIP population from January 1, 2005 to December 31, 2008 included 635,265 patients, of whom 45,645 underwent colonic resections (Table 2). Of the colonic resections, 12,455 (27.2%) were laparoscopic and 33,190 (72.8%) were open procedures. The mean age of all patients undergoing colectomy—"colectomy" group—was 62.1 years, and 48.1% were male. The patients undergoing procedures unrelated to the colon—"noncolectomy group"—were younger (mean = 54.5 years) and approximately the same proportion of male sex as in the other group. Emergent colectomies comprised 18.6% of all colectomies; 3.6% of laparoscopic colectomies were emergent, and 24.2% of open procedures were also emergent. Compared with other NSQIP-captured noncolorectal abdominal procedures, a higher proportion of colectomies were performed as emergency procedures, and most often employed the open approach. The AUROC for emergent morbidity, mortality, and elective morbidity and mortality were 0.73, 0.86, 0.64, and 0.88, respectively (Table 3). The mean relative value unit was 25.6 for all colectomies, 28.1 for laparoscopic colectomies, and 24.7 for open colectomies. As displayed in Table 2, the American Society of Anesthesiologists (ASA) status for colectomies was significantly higher than noncolectomy cases, especially for open procedures. As expected, the predicted morbidity of the colectomy group was much higher than that of the noncolectomy groups (24% for all colectomies vs 17% for laparoscopic colectomies, and 26% for open procedures; all univariate data significant at  $p < 0.001$ ). The occurrence of actual morbidity for all of the NSQIP, all of NSQIP noncolectomy procedures, all colectomies, laparoscopic colectomies, and open procedures was 14.2%, 13.0%, 14.2%, 17.9%, and 34.9%, respectively.

The detail of each AUROC curve is aggregated and summarized in Table 4. The AUROC for the general NSQIP model was 0.817, which was accurate in predicting morbidity in the entire patient population; the confidence interval (CI) was appropriate, and the p value was of statistical significance. The AUROC for the combined laparoscopic and open

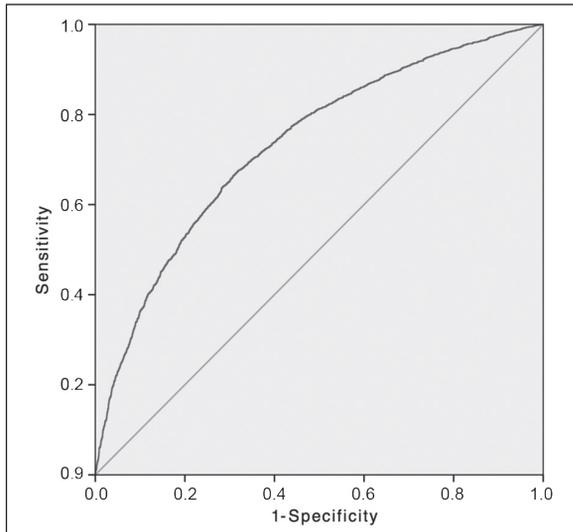


Figure 2. Morbidity of emergency colectomy: receiver operator characteristic curve.

Solid diagonal line represents the random guess line.

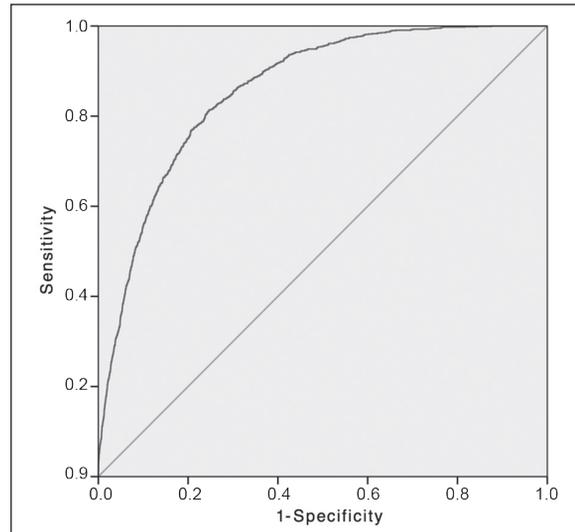


Figure 3. Mortality of emergency colectomy: receiver operator characteristic curve.

Solid diagonal line represents the random guess line.

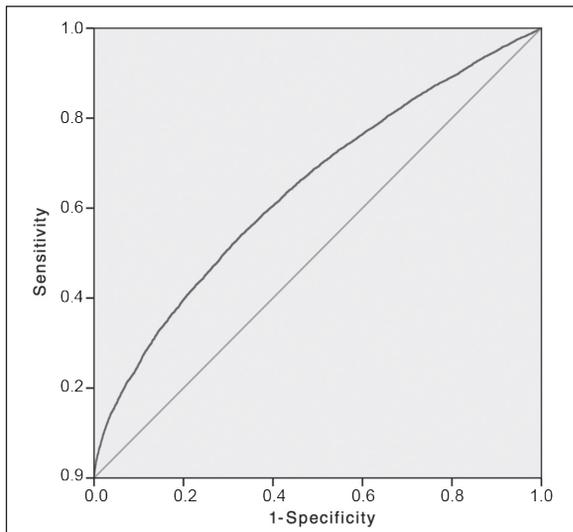


Figure 4. Morbidity of elective colectomy: receiver operator characteristic curve.

Solid diagonal line represents the random guess line.

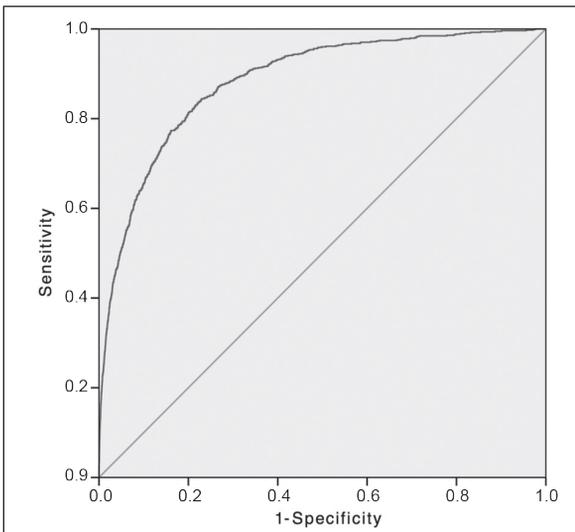


Figure 5. Mortality of elective colectomy: receiver operator characteristic curve.

Solid diagonal line represents the random guess line.

colectomy group was 0.703 and therefore marginal in predicting morbidity for the entire colectomy group. An appropriate CI was also obtained, and the p value demonstrated statistical significance. The AUROCs for the individual laparoscopic and open colectomy groups were 0.633 and 0.701, respectively. The NSQIP-generated AUROCs for these patient populations were marginal at predicting morbidity, which was supported by

adequate sample size, CI, and p values (Figure 1). Figures 2 to 5 show AUROCs for morbidity and mortality for elective and emergency colectomies.

### Discussion

Our review demonstrates that the NSQIP-generated morbidities used to create AUROCs are accurate for patients in an overall surgical model. However, NSQIP-generated morbidities used to

create AUROCs to predict morbidity in patients undergoing open colectomy demonstrated marginal accuracy at best and even less reliability for laparoscopic colectomy. The NSQIP risk models tend to emphasize comorbidities rather than intraoperative details or technical aspects of colonic resections. It is our opinion that certain factors may affect the surgical morbidity, including the case volume experience of the surgeon, the surgeon's training

**The NSQIP risk models tend to emphasize comorbidities rather than intraoperative details or technical aspects of colonic resections.**

(eg, specialized training in colorectal surgery or minimally invasive surgical fellowship), institutional support for colorectal oncology therapy, conversion from laparoscopic to open procedure, and institutional investment in laparoscopic equipment and dedicated surgical teams in the operating room. For instance, Bates and colleagues<sup>9,10</sup> compared operative mortality rates of board-certified colorectal surgeons vs other institutional general surgeons, finding that overall mortality rates for colorectal operations were 1.4% for colorectal surgeons and 7.3% for other general surgeons. Specific patient factors such as type of prior abdominal surgery, adhesions, severity of disease process (in cases such as diverticulitis and inflammatory bowel disease), quality of bowel preparation, intraoperative decision making, intraoperative technique choices, or unexpected findings that change the planned strategy are not tracked or monitored by NSQIP. For example, the study conducted by Van't Sant et al<sup>11</sup> showed that anastomotic leakage developed in 7.8% of patients treated with mechanical bowel preparation and in 5.7% of patients not treated with mechanical bowel preparation ( $p = 0.79$ ). Anastomotic leakage and intraabdominal abscess adverse events, which are of far greater concern to surgeons, are collected under the broad category of the organ/deep space infection variable in NSQIP. Therefore, it is plausible that NSQIP could be enhanced to better evaluate the technical and process-related variables that might affect morbidity rates in colon and rectal surgery, which current NSQIP data have not routinely considered. Investigators for the Michigan Colectomy Collaborative currently are focusing independent efforts on colectomy procedures, using a broader NSQIP approach to produce uniformity

across the data and accurate comparison of different institutions.<sup>12</sup>

Future research efforts are needed to further understand and quantify the impact of various intraoperative factors on postoperative outcome to improve the value of the NSQIP program as it pertains to colorectal surgical procedures. This is of paramount importance considering that colorectal surgical procedures contribute to a substantial percentage of postoperative complications among all general surgical procedures.<sup>13</sup>

### Conclusions

This study demonstrated the limitations of NSQIP as a risk-adjusted program used to monitor postoperative outcomes in patients undergoing colorectal resections. When evaluating practice improvement opportunities on the basis of expected outcomes for colon and rectal surgery in NSQIP reports, an organization and its physicians must balance the significant power of statistical measurements with a need for granularity about specific patient factors that may also influence outcome. Further research is needed to delineate the impact of various intraoperative technical and process-related factors that can affect outcome. ❖

### Disclosure Statement

*The author(s) have no conflicts of interest to disclose.*

### Acknowledgment

*Kathleen Loudon, ELS, of Loudon Health Communications provided editorial assistance.*

### References

1. Khuri SF, Henderson WG, Daley J, et al; Principal Site Investigators of the Patient Safety in Surgery Study. The Patient Safety in Surgery Study: background, study design, and patient populations. *J Am Coll Surg* 2007 Jun;204(6):1089-102. DOI: <http://dx.doi.org/10.1016/j.jamcollsurg.2007.03.028>.
2. Mutch MG. Laparoscopic restorative proctocolectomy: does the national surgical quality improvement program tell the whole story? *Dis Colon Rectum* 2011 Feb;54(2):142-3. DOI: <http://dx.doi.org/10.1007/DCR.0b013e3181f42429>.
3. Wick EC, Vogel JD, Church JM, Remzi F, Fazio VW. Surgical site infections in a "high outlier" institution: are colorectal surgeons to blame? *Dis Colon Rectum* 2009 Mar;52(3):374-9. DOI: <http://dx.doi.org/10.1007/DCR.0b013e31819a5e45>.
4. Kiran RP, El-Gazzaz GH, Vogel JD, Remzi FH. Laparoscopic approach significantly reduces surgical site infections after colorectal surgery: data from national surgical quality improvement program. *J Am Coll Surg* 2010 Aug;211(2):232-8. DOI: <http://dx.doi.org/10.1016/j.jamcollsurg.2010.03.028>.
5. Fleming FJ, Francone TD, Kim MJ, Gunzler D, Messing S, Monson JR. A laparoscopic approach does reduce short-term complications in patients undergoing ileal pouch-anal anastomosis. *Dis Colon Rectum* 2011 Feb;54(2):176-82. DOI: <http://dx.doi.org/10.1007/DCR.0b013e3181f42423>.
6. Paul P, Pennell ML, Lemeshow S. Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets. *Stat Med* 2013 Jan 15;32(1):67-80. DOI: <http://dx.doi.org/10.1002/sim.5525>.
7. Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc; 2000.
8. Ford MK, Beattie WS, Wijeyesundera DN. Systematic review: prediction of perioperative cardiac complications and mortality by the revised cardiac risk index. *Ann Intern Med* 2010 Jan 5;152(1):26-35. DOI: <http://dx.doi.org/10.7326/0003-4819-152-1-201001050-00007>.
9. Bates EW, Berki SE, Homan RK, Lindenauer SM. The challenge of benchmarking: surgical volume and operative mortality in Veterans Administration Medical Centers. *Best Pract Benchmarking Healthc* 1996 Jan-Feb;1(1):34-42.
10. Longo WE, Virgo KS, Johnson FE, et al. Risk factors for morbidity and mortality after colectomy for colon cancer. *Dis Colon Rectum* 2000 Jan;43(1):83-91. DOI: <http://dx.doi.org/10.1007/BF02237249>.
11. Van't Sant HP, Sliker JC, Hop WC, et al. The influence of mechanical bowel preparation in elective colorectal surgery for diverticulitis. *Tech Coloproctol* 2012 Aug;16(4):309-14. DOI: <http://dx.doi.org/10.1007/s10151-012-0852-3>.
12. Englesbe MJ, Brooks L, Kubus J, et al. A state-wide assessment of surgical site infection following colectomy: the role of oral antibiotics. *Ann Surg* 2010 Sep;252(3):519. DOI: <http://dx.doi.org/10.1097/SLA.0b013e3181f244f8>.
13. Schilling PL, Dimick JB, Birkmeyer JD. Prioritizing quality improvement in general surgery. *J Am Coll Surg* 2008 Nov;207(5):698-704. DOI: <http://dx.doi.org/10.1016/j.jamcollsurg.2008.06.138>.

## Not By Any Other Means

Surgery cures diseases that cannot be cured by any other means,  
not by themselves, not by nature, not by medicine.

—Henri de Mondeville, 1260-1320, French surgeon