Published in final edited form as:

J Gastrointest Surg. 2014 May ; 18(5): 986–994. doi:10.1007/s11605-013-2448-2.

Development of an improved risk calculator for complications in proctectomy

Scott K. Sherman, MD¹, Jennifer E. Hrabec, MD¹, Mary E. Charlton, PhD², John W. Cromwell, MD¹, and John C. Byrn, MD¹

¹Department of Surgery, University of Iowa Carver College of Medicine, Iowa City, IA

²Department of Epidemiology, University of Iowa College of Public Health, Iowa City, IA

Abstract

Rectal surgery is associated with high complication rates, but tools to prospectively define surgical risk are lacking. Improved preoperative risk assessment could better inform patients and refine decision making by surgeons. Our objective was to develop a validated model for proctectomy risk prediction. We reviewed non-emergent ACS-NSQIP proctectomy data from 2005-2011 (n=13,385). Logistic regression identified variables available prior to surgery showing independent association with 30-day morbidity in 2010-2011 (n=5,570). The resulting risk model's discrimination and calibration were tested against the NSQIP-supplied morbidity model, and performance was validated against independent 2005-2009 data. Overall morbidity for proctectomy in 2010-2011 was 40.2%; significantly higher than the 23.0% rate predicted by the NSQIP-provided general and vascular surgery risk model. Frequent complications included bleeding (16.3%), superficial infection (9.2%), and sepsis (7.4%). Our novel model incorporating 17 preoperative variables provided better discrimination and calibration (p<0.05) than the NSQIP model, and was validated against 2005-2009 data. A web-based calculator makes this new model available for prospective risk assessment. We conclude that the NSQIP-supplied risk model underestimates proctectomy morbidity and that this new, validated risk model and risk-prediction tool (<http://myweb.uiowa.edu/sksherman>) may allow clinicians to counsel patients with accurate risk estimates using data available in the preoperative setting.

Introduction

Accurate estimates of surgical risk are essential for decision making by surgeons and patients. Surgeons often base risk estimates on a combination of clinical judgment, the surgical literature, and personal experience[1]. This method has variable accuracy, may not account for all risks, and is subject to bias[2-4]. More accurate, data-driven estimates of surgical risk could better inform patient and surgeon expectations and contribute to improved comparisons between surgeons and hospitals[5]. Despite a need for valid risk-estimation, available tools for proctectomy are limited by their origin in single-institution

Corresponding Author: John C. Byrn, MD Clinical Assistant Professor, Department of Surgery University of Iowa Carver College of Medicine 200 Hawkins Drive, 4577 JCP Iowa City, IA 52242 Phone: 319-384-7359 Fax: 319-356-4609 john-byrn@uiowa.edu.

No financial conflicts of interest to disclose

experiences, lack of validation, inclusion of other types of procedures, or restriction to particular diagnoses[6-8].

The American College of Surgeons National Surgical Quality Improvement Project (ACS-NSQIP) collects information on preoperative risk factors and 30-day outcomes abstracted from the medical records of patients undergoing procedures at participating institutions[9]. While not a random sample, these data represent a national cohort of patients in both community and academic settings, and present opportunities for generation of robust and generalizable conclusions regarding rectal surgery risk.

Despite the wealth of data in NSQIP and range of NSQIP-derived risk-prediction tools[8, 10, 11], proctectomy presents particular challenges to accurate risk estimation. With case series reporting complication rates of 30 to over 50%[12-14], proctectomy carries a risk of morbidity substantially higher than the 11% rate reported for general surgery overall[10]. As such, models developed based on general surgery data may be overly influenced by lower-risk surgeries and underestimate risk in proctectomy or other high-risk subgroups[8, 15]. During a recent analysis of the effect of body-mass index (BMI) on proctectomy outcomes, we observed that the NSQIP-supplied morbidity probabilities seemed lower than our expectations[16]. These NSQIP morbidity probabilities are based on the NSQIP general and vascular surgery risk model[17]. Although this is not NSQIP's most sophisticated model for estimating proctectomy risk[15], it is the one supplied with NSQIP data and thus the one most readily available. We therefore set out to test the hypothesis that the NSQIP morbidity risk-model supplied in the Participant-Use-Data-File would underestimate morbidity in patients undergoing proctectomy, and to develop a new, more accurate, and accessible risk prediction tool.

Material and Methods

Patients

Data were obtained from ACS-NSQIP Participant-Use-Data-Files for 2005-2011 (n=13,385) [9]. Included were NSQIP “proctectomy basket” major proctectomy and rectal surgical CPT codes performed by a general surgery (**including colorectal surgery**) primary team (Table 1). These formed 16 primary procedure categories, as described[16]. Due to the unique risk profile of rectal prolapse surgery[18], prolapse procedures were excluded, except for proctopexy with sigmoid resection which was included when diagnoses other than rectal prolapse were designated. Low anterior resection is not supplied with the “proctectomy basket”. Patients requiring chronic ventilator use or undergoing emergency procedures were excluded. This study was Institutional Review Board-exempt.

Variables

The primary outcome, morbidity, was defined as death, reoperation, or any of the 21 NSQIP-recorded cardiac, neurologic, pulmonary, infectious, or bleeding complications within 30 days after surgery[17]. For 2005-2009, total perioperative blood transfusion is not available, and intraoperative transfusion of 3 or more units of blood was coded as a bleeding complication. Variables with low numbers of affected patients were collapsed into

composite variables where applicable. Variables with fewer than 5 cases were censored. BMI categories were defined according to the World Health Organization[19]. Age greater than 90, recorded in NSQIP as “90+”, was considered to be 90. Continuous variables were compared with Welch's t-test; categorical variables were compared by Chi-Squared or Fisher's Exact tests. Multiple comparisons were P-value-adjusted using the Benjamini-Hochberg false discovery rate (FDR) correction[20].

Model development

Univariate logistic regression identified risk factors associated with complications at $p < 0.2$ in 2010-2011 ($n = 5,570$), which were considered for inclusion in a multivariable model. Manual forward and reverse stepwise logistic regression was used to develop a minimum model with $p < 0.2$ for entry and $p > 0.05$ for exit. Laboratory values available for $< 90\%$ of cases were excluded.

Model Validation

Model predictions were compared against NSQIP morbidity probabilities supplied with the Participant-Use-Data-File (“MORBPROB”). The models' predictive abilities were assessed for discrimination by the c-statistic (area under the receiver operating characteristic curve), with DeLong's test to compare curves[21, 22]. For calibration, deviance from the actual complication rate (proportion of predicted minus actual complications to actual complications) was compared by Chi-Squared test. Brier Skill Scores were calculated to provide an additional measure of model performance[23, 24]. Results were validated against NSQIP 2005-2009 data ($n = 7,815$). To estimate the effect of underreported bleeding complications in the older dataset, corrected bleeding complication data were simulated by multivariate imputation by chained equations (MICE)[25]. Results of 5 separately-imputed datasets were analyzed individually and pooled to estimate true deviance and c-statistics in 2005-09. Statistical analyses were performed in R v.2.15.2 (Vienna, Austria).

Results

Patient characteristics

Included were 5,570 patients in 2010-11 (Table 2). The median patient age was 56.0, 45.1% percent were female, and 76.2% were white. The most common procedure was abdominoperineal resection (APR, 24.4%), followed by total proctocolectomy with ileostomy (9.3%), and proctectomy (9.2%). The most common preoperative risk factor was use of antihypertensive medication (37.1%), followed by smoking (17.6%), recent radiation or chemotherapy (15.3%), and recent steroid or immunosuppression treatment (12.7%). The most frequent postoperative diagnosis was cancer (56.1%).

The 2005-09 validation set included 7,815 patients and was similar to the 2010-11 group in age, BMI, gender distribution, three most common procedures, and rates of diabetes, smoking, and immunosuppressant use. Patients in 2005-09 were significantly less likely to be Asian and more likely to be Native American, were less likely to undergo laparoscopic procedures, and more frequently had dyspnea at rest or with exertion, a history of

cardiovascular problems (including stroke, MI, angina, and claudication), and a history of recent radiation or chemotherapy ($p<0.01$ for all).

Proctectomy complications were frequent in 2010-11, with an overall morbidity rate of 40.2% (Table 3). The most common complications were intra/postoperative bleeding requiring transfusion of 5 or more units of blood, superficial wound infection, and sepsis. Thirty-day mortality was 1.2%. In 2011, the only year for which data were available, readmission within 30 days occurred in 16.5% of patients.

The proctectomy complication rate in 2005-09 of 32.4% was significantly lower than in 2010-11, due to a much lower (5.3 vs. 16.3%, $p<0.0001$) rate of bleeding complications. The difference in bleeding complications exists because of a change in the definition and procedure for coding this complication that began in 2010 to correct previous underreporting. Superficial infections occurred more frequently in 2005-09. Rates of other complications did not differ significantly between the time periods.

Development of a Proctectomy Morbidity Risk Model

The NSQIP morbidity model's performance was evaluated in the entire 2010-11 population of 5,570 proctectomy patients, where it predicted an overall morbidity of 23.0%, significantly less than the actual 40.2% morbidity rate ($p<0.0001$). To provide more accurate risk estimates, more than 70 individual and composite preoperative variables were tested for association with complications. Fifteen variables showed significant independent contribution to morbidity risk on multivariable analysis and would be available prior to operation. These were included in the Iowa Proctectomy Risk Model (Table 4). Also included were poor preoperative functional status and the presence of preoperative sepsis, which due to low event numbers did not reach significance, but which improved model performance and held a plausible connection to complication risk. Intraoperative factors, such as wound classification and whether additional procedures were performed, were strongly associated with complications, but were not included in this model. Laboratory values unavailable for large numbers of patients were not included. A preoperative INR was available for only 54% of patients, but that an INR had been checked in 30 days prior to surgery in fact correlated significantly with complications ($p<0.005$), and was included.

The most influential factor in the model was the primary procedure performed. Among procedures with more than 200 cases, complication rates varied between 29.8% for laparoscopic proctectomy with j-pouch and coloanal anastomosis to 50.1% for total proctocolectomy with ileostomy. Well-recognized operative risk factors such as advanced age, obesity, steroid use, dyspnea, and poor functional status all correlated with increased complication rates. Interestingly, a history of cardiovascular problems was not independently associated with complications, and a history of recent radiation treatment correlated with decreased risk of morbidity. Despite differences in male compared to female pelvic anatomy, gender was not associated with complications on uni- or multivariate analysis.

Model Performance and Validation

To assess the Iowa Model's performance compared to the NSQIP model, discrimination and calibration of both models were calculated. In 2010-11, the NSQIP model provided reasonable discrimination between patients who did and did not have complications, with a c-statistic of 0.643. The Iowa Model showed significantly better discrimination ($p<0.01$), with a c-statistic of 0.660 (Table 5).

We assessed calibration by comparing the proportion of total predicted minus actual complications to actual complications, and report the results as the percent deviance of predicted to actual complications (Table 5). The Iowa Model showed perfect calibration when employed on 2010-2011 data, with its predicted 40.2% complication rate equal to the actual complication rate (Figure 1A). The NSQIP model performed significantly worse, predicting a 23.0% complication rate with a percent deviance of -42.9% ($p<0.001$).

To further assess model performance, Brier Skill Scores (BSS) were calculated for the NSQIP and Iowa Models. The BSS is defined as $BSS = 1 - (\text{Brier Score} / \text{Null Brier Score})$, and reflects the Brier Score normalized to a null model where the observed average morbidity serves as the prediction for all cases. Higher BSS values show better predictive performance, while negative BSS values indicate predictive skill worse than guessing the average for all cases. In 2010-2011, the NSQIP model had a Brier Skill Score of -0.073 while the Iowa model performed much better, at +0.075.

To validate its performance on an independent dataset, the Iowa Model was applied to 2005-09 NSQIP proctectomy data. Discrimination of the Iowa and NSQIP models was not significantly different in the validation set (c-statistic 0.608 vs. 0.606, $p=0.6$). The Iowa Model still showed significantly better calibration, with +25.9% deviance in predicted-to-actual complications, versus -33.9% for the NSQIP model ($p<0.001$). Due to their large percent deviances, both models had negative Brier Skill Scores, with -0.016 for the Iowa Model and -0.024 for the NSQIP model.

To more rigorously assess the impact of bleeding underreporting in the 2005-09 dataset, bleeding complications were simulated by multivariate imputation by chained equations. Bleeding complication data were removed from 2005-09, and the data combined with 2010-11. Using all variables not dependent on bleeding complications as predictors, bleeding complications were imputed for 2005-09 in 5 independent simulations. In these simulations, the mean overall morbidity rate for 2005-09 was 41.4%, significantly higher than the underreported rate of 32.4%. This led to improved discrimination in both models, with mean c-statistics of 0.636 for the Iowa Model and 0.623 for the NSQIP model, with the Iowa Model outperforming the NSQIP model in all simulations. The mean deviance of the Iowa Model was -1.4%, significantly better than the -48.3% for the NSQIP model (Figure 1B). Brier Skill Scores for the Iowa and NSQIP models were +0.052 and -0.118, respectively. These simulations demonstrate 41.4% as a reasonable estimate of the actual morbidity rate in 2005-09, which is close to the Iowa Model's predicted rate of 40.8%. Thus, underreporting of bleeding complications in 2005-09 disadvantaged the Iowa Model, but it nevertheless performed significantly better than the NSQIP model using unadjusted data. When the data were corrected for bleeding underreporting, performance of both models

improved, with the Iowa Model showing greater improvement and significantly better performance. We therefore conclude that the Iowa Model shows superior discrimination and calibration based on these validation studies using data independent from those used to develop the model.

Online Risk Calculator

Using the validated Iowa Model, an open-source online proctectomy risk calculator, available at <http://myweb.uiowa.edu/sksherman>, was created. After entering the anticipated procedure, the patient's age, BMI, and the other variables included in the Iowa Model, the calculator returns the probability of experiencing a complication.

Discussion

For procedures with high complication rates, the true risk of morbidity may not be well-estimated by standard models. In a previous analysis of the effect of BMI on morbidity in proctectomy, we observed that morbidity probabilities assigned in the NSQIP Participant-Use-Data-File were lower than expected for these seemingly high-risk patients[16]. The present study demonstrates that for proctectomy, the NSQIP Participant-Use-Data-File morbidity probabilities drastically underestimate the true morbidity rate, and that the Iowa Model provides significantly better risk estimates. This model is valid based on independent 2005-09 data.

Discrimination and calibration define performance for risk models. Discrimination refers to the model's ability to correctly distinguish patients at higher risk of complications from those at lower risk and calibration describes how closely the model's assigned risk matches actual risk. A method to assess model calibration, the Hosmer-Lemeshow Goodness-of-Fit Test has been applied to other NSQIP risk-models[8, 10], but its limitations in large datasets are well-established[26, 27]. The Brier Score has been proposed as a superior measure of surgical risk calculator performance[11], but interpreting Brier Scores from different datasets, or for models addressing different outcomes is not straightforward. Brier Scores depend strongly on the underlying rate of the event being predicted, which can cause confusion because Brier Scores for predictions of more likely events may seem worse than predictions for less-likely outcomes, even if the models are equally informative. The Brier Skill Score mitigates this somewhat by incorporating the event probability in the null model score[24], and the Iowa Model returned a greater Brier Skill Score than the NSQIP model in all cases. Still, despite development of methods for hypothesis testing in two models applied to the same dataset, assessing models' relative performance in different datasets by Brier Score remains problematic[28-31]. We therefore assessed calibration using percent deviance. This metric allows comparisons between different models with a Chi-Squared statistic, and offers an intuitive presentation of a model's predicted average morbidity compared to that actually observed.

A persistent challenge in risk modeling is accounting for risk in subgroups where the risk differs substantially from the average or when less variation exists between affected and unaffected patients. For example, Cohen *et al.* reported that c-statistics for discriminating morbidity decreased from above 0.80 to 0.68 when their model considered colorectal

surgery instead of all of general surgery[8]. Our subgroup of only proctectomy, which represents approximately 8% of all colorectal procedures in NSQIP, should be expected to complicate accurate discrimination. In fact, the maximum-achievable c-statistic varies based on the magnitude and variance of risk in a population. It can be demonstrated that while a c-statistic of 0.99 is theoretically achievable for an outcome with a narrowly-distributed risk averaging close to zero (as might be observed when studying mortality), for an outcome with a risk of around 50% and a large variance (such as is observed for morbidity in proctectomy), the maximum-possible value for the c-statistic is in the range of 0.62-0.72[32]. As such, a c-statistic of 0.66 represents outstanding discrimination for a group of procedures with a high preoperative probability of complications, even if it is lower than would be expected when measuring more rare outcomes, such as mortality. Similarly, a difference in c-statistic of 0.02 between the two models is not large in absolute terms, but for a c-statistic with a possible range of 0.5-0.72, represents an improvement of at least 9% and is statistically and potentially clinically significant.

One explanation of the strength of our model is its reliance on the procedure performed as a strong predictor of risk. Although early NSQIP models focused solely on patient-characteristics to define risk, it was recognized that procedure-based risk often carries more information[15]. Later models moved increasingly towards the procedure as a risk predictor, first grouping CPT codes together[10], and most recently considering CPT codes individually[11]. The NSQIP model assessed here relies on large CPT groups, and the difference between morbidity in all colorectal surgery (reported at 23%, which is very close to the rate predicted by the NSQIP model in this analysis[8]) compared to only proctectomy, where complication rates approximating 40% have been described[12-14, 33], likely contributes to the poor calibration of the NSQIP model as applied to this subgroup. By using smaller CPT groups, the Iowa Model achieves improved calibration, with the predicted risk close to the true average morbidity (Figure 1). Furthermore, we conclude that the NSQIP model, while displaying reasonable discrimination, is poorly calibrated for predicting risk in proctectomy.

Another factor affecting the NSQIP model's performance in this analysis is its failure to include bleeding as a complication. Omission of bleeding from previous and existing NSQIP morbidity models is important as bleeding represents the most frequent complication in proctectomy. The NSQIP model therefore systematically underestimates morbidity. Any inclusion or exclusion of factors as "complications" is at some level arbitrary, but since transfusion of 5 or more units of blood is rarely planned, and as evidence documenting the negative effects of transfusion accumulates[34, 35], it seems reasonable that our model, which is designed to inform understanding of preoperative risk, should count this occurrence as morbidity.

Validation against independent data is essential, as risk models usually perform well when tested against the same data used to develop them. In applying the Iowa Model to 2005-2009 data, the validation dataset was limited by its underreported complication rate of 32.4%, but as it remained the best data available, it was used. Considering that underreporting of bleeding complications by approximately 10% occurred in 2005-09, the actual complication rate is likely to be higher than the reported 32.4% and closer to the Iowa Model's predictions

while further from the NSQIP model's. Our analysis using a more realistic imputed bleeding rate for 2005-09 supports this conclusion. Thus, despite the spuriously low complication rate in 2005-09 data, the Iowa Model provided better overall risk estimation than the NSQIP model and is valid.

Other risk calculators for proctectomy exist. NSQIP has offered an online calculator since 2008, but it was available on a subscription basis only and was not readily accessible. The recently-released NSQIP Universal Risk Calculator is freely available, and represents a powerful tool providing risk estimates for a huge range of procedures and patient characteristics[11]. Yet despite its strong composite performance, its validity for overall morbidity in high-risk procedures, and subsets of the total is unknown. Its performance in one subset, colectomy, was reported by Bilimoria *et al.* In these colectomy cases, the Universal Risk Calculator's performance was compared to a colectomy-specific model. Although the authors characterize the differences as small, calculating Brier Skill Scores for the two models' predictions demonstrates that the colectomy-specific model outperformed the Universal Model by 10% for mortality, 12% for morbidity, and 40% for surgical site infection[11]. These results support that risk models developed for specific subsets of patients and outcomes may outperform omnibus models if one has the time to develop them.

The Iowa Model **is one such specially-developed model**, and offers several advantages for proctectomy. The variables in the Iowa Model were manually chosen specifically for **these procedures**. The large volume of procedures in the NSQIP Universal Calculator necessitates a standardized and automated approach to model development, which while efficient, may not always select the most powerful predictors for all procedures[15]. Secondly, although the most recent NSQIP model appropriates all available NSQIP data, it was not tested on an independent dataset[11]. Particularly for procedures with lower n, such as proctectomy, this increases the risk of overfitting and could decrease predictive power. Thirdly, the Iowa Model is highly accessible. The online proctectomy risk calculator allows clinicians to provide validated risk estimates to patients, and the open-source code for the calculator allows it to be tested by other investigators against additional risk-assessment tools.

The Iowa Model is subject to the limitations of the NSQIP data on which it is based, and is not informative for patient groups that were excluded, such as emergency cases. Some complications relevant to proctectomy, such as urinary retention, or those which may become apparent only after 30 days, such as sexual dysfunction, effects on fertility, and long-term pouch outcomes, are not captured by NSQIP data, and cannot be addressed by this model. Several included CPT codes had low case numbers, and risk estimates for these procedures are likely less accurate than those for better-represented procedures. Moreover, the high-risk nature of these procedures frustrates efforts at definitive prediction, as both patients with favorable and unfavorable risk characteristics are expected to experience complications at relatively high rates.

The Iowa Model sacrifices predictive power for utility in the decision to limit predictors to those available before surgery. After including informative factors such as operative time, wound classification, and whether additional procedures were performed, an expanded

version of the model achieved a significantly improved c-statistic of 0.685 ($p < 0.001$ vs. base Iowa Model). However, because these factors are unavailable prior to the operation, their usefulness is questionable even if they improve the model's on-paper performance. Notably, the NSQIP Universal Calculator requires ASA and wound classifications, which are not always known in advance[11].

The Iowa Model confers benefits to both patients and surgeons. Armed with improved preoperative knowledge, patients may make more informed decisions and align their expectations with the reality of the often-complicated postoperative proctectomy course. For surgeons, the action suggested by a high preoperative risk is less clear. The correlative nature of these data and their inability to account for selection biases limits their recommendations for alternative treatments. That total proctocolectomy with ileal-pouch-anal anastomosis (IPAA) had a lower rate of complications than total proctocolectomy with ileostomy (37 vs. 50%) is not reason to offer high-risk patients the more complex procedure, because selection of healthier patients for IPAA likely accounts for the lower morbidity. Instead, we propose that after finding a patient at above-average risk, the surgeon may focus additional attention on altering modifiable risk factors, consider delaying surgery, or if surgery is more immediately necessary, select the least complex procedure indicated for the patient's illness. The risk-calculator could also help identify high-risk patients most appropriate for inclusion in studies of perioperative interventions to reduce morbidity.

Conclusion

The morbidity probabilities derived from the NSQIP risk model and supplied with NSQIP data significantly underestimate morbidity in proctectomy. While newer NSQIP risk models may show improved performance, their utility in high-risk subsets is unknown. Our new, validated proctectomy risk model and web-calculator (<http://myweb.uiowa.edu/sksherman>) provide significantly improved estimates of 30-day morbidity, and promise to assist clinicians and patients in making informed surgical decisions based on accurate risk-estimates.

Acknowledgments

Supported by NIH 5T32#CA148062-03 (SKS and JEH). Thanks to Mary E. Belding-Schmitt RN, BSN for NSQIP and database support.

References

1. Markus PM, Martell J, Leister I, Horstmann O, Brinker J, Becker H. Predicting postoperative morbidity by clinical assessment. *Br J Surg.* Jan; 2005 92(1):101–6. [PubMed: 15635697]
2. Ferguson MK, Stromberg JD, Celauro AD. Estimating lung resection risk: a pilot study of trainee and practicing surgeons. *Ann Thorac Surg.* Apr; 2010 89(4):1037–42. discussion 42-3. [PubMed: 20338304]
3. Volk ML, Roney M, Merion RM. Systematic bias in surgeons' predictions of the donor-specific risk of liver transplant graft failure. *Liver Transpl.* Jun 19.2013
4. Karliczek A, Harlaar NJ, Zeebregts CJ, Wiggers T, Baas PC, van Dam GM. Surgeons lack predictive accuracy for anastomotic leakage in gastrointestinal surgery. *Int J Colorectal Dis.* May; 2009 24(5):569–76. [PubMed: 19221768]

5. Merkow RP, Bilimoria KY, Ko CY. Surgical quality measurement: an evolving science. *JAMA Surg.* Jul 1; 2013 148(7):586–7. [PubMed: 23864242]
6. Ragg JL, Watters DA, Guest GD. Preoperative risk stratification for mortality and major morbidity in major colorectal surgery. *Dis Colon Rectum.* Jul; 2009 52(7):1296–303. [PubMed: 19571708]
7. Fazio VW, Tekkis PP, Remzi F, Lavery IC. Assessment of operative risk in colorectal cancer surgery: the Cleveland Clinic Foundation colorectal cancer model. *Dis Colon Rectum.* Dec; 2004 47(12):2015–24. [PubMed: 15657649]
8. Cohen ME, Bilimoria KY, Ko CY, Hall BL. Development of an American College of Surgeons National Surgery Quality Improvement Program: morbidity and mortality risk calculator for colorectal surgery. *J Am Coll Surg.* Jun; 2009 208(6):1009–16. [PubMed: 19476884]
9. ACS-NSQIP Info Book; Chicago: 2013. Available at: <http://site.acsnsqip.org/wp-content/uploads/2012/11/NSQIP-info-book-10.12.pdf>. [1/3/13]
10. Raval MV, Cohen ME, Ingraham AM, Dimick JB, Osborne NH, Hamilton BH, Ko CY, Hall BL. Improving American College of Surgeons National Surgical Quality Improvement Program risk adjustment: incorporation of a novel procedure risk score. *J Am Coll Surg.* Dec; 2010 211(6):715–23. [PubMed: 20846884]
11. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko C, Cohen ME. Development and Evaluation of the Universal ACS NSQIP Surgical Risk Calculator: A Decision Aide and Informed Consent Tool for Patients and Surgeons. *J Am Coll Surg.* 2013 10.1016/j.jamcollsurg.2013.07.385.
12. Balentine CJ, Robinson CN, Marshall CR, Wilks J, Buitrago W, Haderxhanaj K, Sansgiry S, Petersen NJ, Bansal V, Albo D, Berger DH. Waist circumference predicts increased complications in rectal cancer surgery. *J Gastrointest Surg.* Nov; 2010 14(11):1669–79. [PubMed: 20835770]
13. Ballian N, Yamane B, Levenson G, Harms B, Heise CP, Foley EF, Kennedy GD. Body mass index does not affect postoperative morbidity and oncologic outcomes of total mesorectal excision for rectal adenocarcinoma. *Ann Surg Oncol.* Jun; 2010 17(6):1606–13. [PubMed: 20077020]
14. Canedo JA, Pinto RA, McLemore EC, Rosen L, Wexner SD. Restorative proctectomy with ileal pouch-anal anastomosis in obese patients. *Dis Colon Rectum.* Jul; 2010 53(7):1030–4. [PubMed: 20551755]
15. Cohen ME, Ko CY, Bilimoria KY, Zhou L, Huffman K, Wang X, Liu Y, Kraemer K, Meng X, Merkow R, Chow W, Matel B, Richards K, Hart AJ, Dimick JB, Hall BL. Optimizing ACS NSQIP Modeling for Evaluation of Surgical Quality and Risk: Patient Risk Adjustment, Procedure Mix Adjustment, Shrinkage Adjustment, and Surgical Focus. *J Am Coll Surg.* Aug; 2013 217(2):336–46. e1. [PubMed: 23628227]
16. Hrabe JE, Sherman SK, Charlton ME, Cromwell JW, Byrn JC. The Effect of BMI on Outcomes in Proctectomy. *Dis Colon Rectum.* 2013 In Press.
17. User Guide for the 2011 Participant Use Data File. American College of Surgeons National Surgical Quality Improvement Program; Chicago, IL: Oct. 2012
18. Fang SH, Cromwell JW, Wilkins KB, Eisenstat TE, Notaro JR, Alva S, Bustami R, Chinn BT. Is the abdominal repair of rectal prolapse safer than perineal repair in the highest risk patients? An NSQIP analysis. *Dis Colon Rectum.* Nov; 2012 55(11):1167–72. [PubMed: 23044678]
19. Report of a WHO consultation. Vol. 894. World Health Organ Tech Rep Ser; 2000. Obesity: preventing and managing the global epidemic.; p. i-xii.p. 1-253.
20. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med.* Jul; 1990 9(7):811–8. [PubMed: 2218183]
21. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011; 12:77. [PubMed: 21414208]
22. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* Apr; 1982 143(1):29–36. [PubMed: 7063747]
23. Brier GW. Verification of forecasts expressed in terms of probability. *Monthly Weather Review.* 1950; 78:1–3. 2/10/1950.
24. Wilks, DS. *Statistical Methods in the Atmospheric Sciences.* Academic Press; 1995.

25. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. Dec; 2011 45(3):1–67.
26. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Crit Care Med*. Sep; 2007 35(9):2052–6. [PubMed: 17568333]
27. Paul P, Pennell ML, Lemeshow S. Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets. *Stat Med*. Jan 15; 2013 32(1):67–80. [PubMed: 22833304]
28. Redelmeier DA, Bloch DA, Hickam DH. Assessing predictive accuracy: how to compare Brier scores. *J Clin Epidemiol*. 1991; 44(11):1141–6. [PubMed: 1941009]
29. Ferro CAT. Comparing Probabilistic Forecasting Systems with the Brier Score. *Weather and Forecasting*. Oct; 2007 22(5):1076–88. 2007.
30. Bradley AA, Schwartz SS, Hashino T. Sampling Uncertainty and Confidence Intervals for the Brier Score and Brier Skill Score. *Weather and Forecasting*. Oct; 2008 23(5):992–1006. 2008.
31. Rufibach K. Use of Brier score to assess binary predictions. *J Clin Epidemiol*. Aug; 2010 63(8): 938–9. author reply 9. [PubMed: 20189763]
32. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. Feb 20; 2007 115(7):928–35. [PubMed: 17309939]
33. Clark W, Siegel EM, Chen YA, Zhao X, Parsons CM, Hernandez JM, Weber J, Thareja S, Choi J, Shibata D. Quantitative measures of visceral adiposity and body mass index in predicting rectal cancer outcomes after neoadjuvant chemoradiation. *J Am Coll Surg*. Jun; 2013 216(6):1070–81. [PubMed: 23523147]
34. Amato A, Pescatori M. Perioperative blood transfusions for the recurrence of colorectal cancer. *Cochrane Database Syst Rev*. 2006; (1):CD005033. [PubMed: 16437512]
35. Ball CG, Pitt HA, Kilbane ME, Dixon E, Sutherland FR, Lillemoe KD. Peri-operative blood transfusion and operative time are quality indicators for pancreaticoduodenectomy. *HPB (Oxford)*. Sep; 2010 12(7):465–71. [PubMed: 20815855]

Figure 1.

Density plots showing the distribution of predicted morbidity probabilities (x-axis) in 2010-11 (A) and 2005-09 (B) datasets. The Iowa Model (solid line) is better calibrated than the NSQIP model (dashed line), as indicated by predicted morbidities centered more closely around the actual morbidity rates of 40.2% (A) and 41.5% (B) (dotted line). Morbidity in 2005-09 based on imputed bleeding complications.

Table 1

Included CPT codes, procedure groups and descriptions, sample sizes, and percentages from 2010-11 and 2005-09 datasets.

CPT Code	Procedure Type	2010-11		2005-09	
		n	%	n	%
45110	Abdominoperineal Resection	1359	24.4	2151	27.5
44155, 44156, 45121	Total Proctocolectomy with Ileostomy	519	9.3	874	11.2
45111, 45123	Proctectomy	510	9.2	903	11.6
44211	Laparoscopic Total Proctocolectomy with Ileal-Pouch-Anal-Anastomosis	497	8.9	603	7.7
45395	Laparoscopic Abdominoperineal Resection	448	8.0	309	4.0
44158	Total Proctocolectomy with Ileal-Pouch-Anal-Anastomosis	357	6.4	555	7.1
45112, 45114, 45120	Proctectomy with Anastomosis	346	6.2	431	5.5
45119	Proctectomy with J-Pouch and Coloanal Anastomosis	332	6.0	553	7.1
45113	Proctectomy with Ileal-Pouch-Anal-Anastomosis	328	5.9	510	6.5
45397	Laparoscopic Proctectomy with J-Pouch and Coloanal Anastomosis	299	5.4	245	3.1
44212	Laparoscopic Total Proctocolectomy with Ileostomy	237	4.3	248	3.2
44157	Total Proctocolectomy with Ileal-Anal Anastomosis	145	2.6	176	2.3
45126	Pelvic Exenteration	80	1.4	131	1.7
45550	Proctopexy with Sigmoid Resection	42	0.8	58	0.7
45402	Laparoscopic Proctopexy with Sigmoid Resection	39	0.7	24	0.3
45160	Transsacral Proctotomy	31	0.6	44	0.6

Table 2

Patient characteristics compared between 2005-09 and 2010-11 time periods.

Preoperative Characteristic	2010-2011 n=5570	2005-2009 n=7815	P Value
Age, median years (IQR)	56.0 (45.0-67.0)	56.0 (44.0-68.0)	0.45
Sex (Female %)	45.1	43.9	0.26
Race (White %)	76.2	77.7	0.078
Asian	2.9	1.3	<0.001
Black	7.3	6.3	0.060
Hispanic	4.0	4.0	0.89
Native American	0.9	1.7	<0.001
Unknown/Other	8.6	9.1	0.47
Laparoscopic procedure (%)	27.2	18.3	<0.001
BMI median (IQR)	26.4 (22.9-30.5)	26.4 (23.0-30.5)	0.55
Diabetes (%)	11.7	11.2	0.55
Smoking (%)	17.6	17.9	0.69
Steroid treatment in past 30 days (%)	12.7	13.3	0.47
Dyspnea at rest or with exertion (%)	5.6	7.4	<0.001
Poor functional status (%)	2.9	3.2	0.38
Cardiovascular problems (%)	5.2	7.9	<0.001
Congestive heart failure (%)	0.3	0.4	0.33
Hypertension medication (%)	37.1	37.5	0.64
Disseminated cancer (%)	4.8	5.4	0.18
Large weight loss (%)	6.0	6.9	0.076
Impaired clotting (%)	2.8	3.1	0.55
Recent chemo- or radio-therapy (%)	15.3	21.3	<0.001
Hematocrit, median % (IQR)	38.4 (34.9-41.3)	38.5 (35.0-41.7)	0.065
Leukocyte count, median 1000/dL (IQR)	6.5 (5.1-8.4)	6.6 (5.2-8.4)	0.065
Post-op Diagnosis (%)			
Benign disease	15.6	14.4	0.21
Cancer	56.1	57.0	0.46
Inflammatory bowel disease	28.3	28.6	0.77

P values represent comparison between two time periods and are false discovery rate-adjusted.

IQR: Interquartile Range; BMI: Body Mass Index

Table 3

Actual rates of morbidity, mortality, and selected individual complications.

Complication	2010-2011 n=5570	2005-2009 n=7815	P Value
Morbidity (%)	40.2	32.4	<0.001
Bleeding (%)	16.3	5.3	<0.001
Superficial wound infection (%)	9.2	11.0	<0.01
Deep wound infection (%)	2.5	2.9	0.20
Organ space infection (%)	7.0	6.3	0.20
Dehiscence (%)	1.8	2.3	0.11
Sepsis (%)	7.4	8.1	0.20
Septic shock (%)	1.4	1.8	0.12
Deep vein thrombosis (%)	1.7	1.8	0.74
Pulmonary embolism (%)	0.5	0.6	0.74
Stroke (%)	0.3	0.2	0.32
Need for dialysis (%)	0.5	0.6	0.34
MI or cardiac arrest (%)	0.9	0.7	0.20
Mortality (%)	1.2	0.8	0.11
Readmission within 30 days (%)	16.5	N/A	

Readmission data available for 2011 only, MI: Myocardial Infarction

P values are false discovery rate-adjusted

Table 4

The Iowa Proctectomy Risk Model. Shown are included predictive variables with associated odds-ratios (OR) for morbidity, 95% confidence intervals (CI), and P values.

Variable	OR	95% CI	P Value
Primary procedure (APR reference)	1.00	-	<0.001 ^a
Lap APR	0.74	0.58-0.93	<0.01
Lap proctectomy with j-pouch, coloanal anastomosis, pos. stoma	0.59	0.44-0.78	<0.001
Lap proctopexy with sigmoid resection	0.18	0.06-0.42	<0.001
Lap TPC ileostomy	0.81	0.60-1.10	0.18
Lap TPC IPAA pos. ileostomy	0.68	0.53-0.87	<0.01
Pelvic exenteration	2.53	1.51-4.37	<0.001
Proctectomy	0.51	0.40-0.64	<0.001
Proctectomy with j pouch, coloanal anastomosis, pos. stoma	0.73	0.56-0.95	0.018
Proctectomy IPAA pos. ileostomy	0.77	0.58-1.02	0.068
Proctectomy with anastomosis	0.60	0.46-0.78	<0.001
Proctopexy with sigmoid resection	1.08	0.56-2.07	0.81
TPC IAA pos. ileostomy	0.61	0.42-0.88	<0.01
TPC ileostomy	0.91	0.73-1.14	0.41
TPC IPAA pos. ileostomy	0.76	0.58-0.99	0.041
Transsacral proctotomy	0.23	0.08-0.56	<0.01
Age	1.006	1.002-1.010	<0.01
Smoking	1.30	1.12-1.52	<0.001
Dyspnea at rest or with exertion	1.57	1.23-2.02	<0.001
Poor functional status	1.36	0.95-1.95	0.091
History of stroke	2.01	1.13-3.66	0.02
Disseminated cancer	1.69	1.29-2.21	<0.001
Preoperative open wound	1.79	1.25-2.57	<0.01
Preoperative steroid/immunosuppressant use	1.25	1.04-1.52	0.02
Preoperative weight loss >10%	1.28	1.01-1.64	0.042
Transfusion of >4U PRBC in 72 hours prior to surgery	2.51	1.37-4.88	<0.01
Radiation therapy within 90 days	0.83	0.69-0.98	0.031
Preoperative sepsis	1.40	0.93-2.13	0.11
INR checked within 30 days	1.19	1.06-1.33	<0.01
Preoperative leukocyte count	1.02	1.002-1.04	0.033
Preoperative hematocrit	0.95	0.93-0.96	<0.001
BMI Category (Normal; BMI 18.5-24.9 reference)	1.00	-	<0.001 ^a
Underweight; BMI <18.5	0.90	0.66-1.21	0.47
Overweight; BMI 25-29.9	1.04	0.90-1.20	0.60
Class I Obesity; BMI 30-34.9	1.38	1.16-1.64	<0.001

Variable	OR	95% CI	P Value
Class II Obesity; BMI 35-39.9	1.97	1.55-2.51	<0.001
Class III Obesity; BMI 40 or greater	1.42	1.03-1.95	0.031

APR: Abdominoperineal resection; Lap: Laparoscopic; Pos.: Possible; TPC: Total Proctocolectomy; IPAA: Ileal-Pouch-Anal-Anastomosis; IAA: Ileal-Anal-Anastomosis; U PRBC: Units Packed Red Blood Cells

^aThese p-values represent overall significance of Primary Procedure and BMI Category variables.

Table 5

Comparison of model performance. Discrimination and calibration analyses show significantly better performance by the Iowa Model compared to the NSQIP Model.

Dataset		Model		P Value
		Iowa Model	NSQIP Model	
Training Set 2010-2011 n=5570	Predicted Morbidity (%)	40.2	23.0	<0.001
	Actual Morbidity (%)	40.2	40.2	
	C-statistic	0.660	0.643	<0.01
	Percent Deviance	0.0%	-42.9%	<0.001
Validation Set 2005-2009 n=7815	Predicted Morbidity (%)	40.8	21.4	<0.001
	Actual Morbidity (%)	32.4	32.4	
	C-statistic	0.608	0.606	0.60
	Percent Deviance	+25.9%	-33.9%	<0.001
Imputed Validation Set 2005-2009 n=7815	Predicted Morbidity (%)	40.8	21.4	<0.001
	Actual Morbidity (%)	41.5	41.5	
	C-statistic	0.637	0.621	<0.01
	Percent Deviance	-1.7%	-48.4%	<0.001

Results are shown for the training set (2010-11), unadjusted validation set (2005-09), and for validation dataset with imputed bleeding complications.