



Published in final edited form as:

*J Am Coll Surg.* 2023 June 01; 236(6): 1093–1103. doi:10.1097/XCS.0000000000000658.

## Quantifying the Prognostic Value of Preoperative Surgeon Intuition: Comparing Surgeon Intuition and Clinical Risk Prediction as Derived from the American College of Surgeons NSQIP Risk Calculator

Jayson S Marwaha, MD MBI<sup>1,2</sup>, Brendin R Beaulieu-Jones, MD MBA<sup>1,2</sup>, Margaret Berrigan, MD<sup>1</sup>, William Yuan, PhD<sup>2</sup>, Stephen R Odom, MD, FACS<sup>1</sup>, Charles H Cook, MD, FACS<sup>1</sup>, Benjamin B Scott, MD<sup>1</sup>, Alok Gupta, MD, FACS<sup>1</sup>, Charles S Parsons, MD, FACS<sup>1</sup>, Anupamaa J Seshadri, MD<sup>1</sup>, Gabriel A Brat, MD, FACS, MPH<sup>1,2</sup>

<sup>1</sup>Department of Surgery, Beth Israel Deaconess Medical Center, Boston, MA

<sup>2</sup>Department of Biomedical Informatics, Harvard Medical School, Boston, MA

### Abstract

**Background:** Surgical risk prediction models traditionally use patient attributes and measures of physiology to generate predictions about postoperative outcomes. However, the surgeon's assessment of the patient may be a valuable predictor, given the surgeon's ability to detect and incorporate factors that existing models cannot capture. We compare the predictive utility of surgeon intuition and a risk calculator derived from the American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP).

**Study Design:** From 10/1/2021 to 9/1/2022, surgeons were surveyed immediately before performing surgery to assess their perception of a patient's risk of developing any postoperative complication. Clinical data were abstracted from ACS NSQIP. Both sources of data were independently used to build models to predict the likelihood of a patient experiencing any 30-day postoperative complication as defined by ACS NSQIP.

**Results:** Preoperative surgeon assessment was obtained for 216 patients. NSQIP data were available for 9182 patients who underwent general surgery (1/1/17 to 9/1/22). A binomial regression model trained on clinical data alone had an AUC of 0.83 (95% CI: 0.80-0.85) in predicting any complication. A model trained on only preoperative surgeon intuition had an AUC of 0.70 (95% CI: 0.63-0.78). A model trained on surgeon intuition and a subset of clinical predictors had an AUC of 0.83 (95% CI: 0.77-0.89).

**Conclusions:** Preoperative surgeon intuition alone is an independent predictor of patient outcomes; however, a risk calculator derived from ACS NSQIP is a more robust predictor of post-operative complication. Combining intuition and clinical data did not strengthen prediction.

---

**Corresponding Author:** Gabriel A. Brat, MD, MPH, Department of Surgery, Beth Israel Deaconess Medical Center, Department of Biomedical Informatics, Harvard Medical School, 110 Francis Street, Suite 2G, Boston, MA, 02215, gbrat@bidmc.harvard.edu. Drs Marwaha and Beaulieu-Jones contributed equally to this work.

Disclosure Information: Nothing to disclose.

Presented at the 103rd Annual meeting of the New England Surgical Society, Boston, MA, September 2022.

## Precis

We compared the predictive utility of preoperative surgeon intuition and surgical risk calculators and found that, while preoperative surgeon intuition alone is an independent predictor of patient outcomes, traditional risk calculators are more robust predictors of postoperative complication.

---

## Introduction

Physicians frequently integrate nuanced and potentially conflicting data from a number of sources to make complex clinical decisions. There is significant inter-physician variation in decision-making: in certain specialties, patients seeking a second opinion only receive the same diagnosis from both physicians 12% of the time.<sup>1</sup> That physicians evaluating the same patient with the same physiology can arrive at different conclusions suggests that a substantial portion of the decision-making process occurs beyond the objective evaluation of clinical attributes and physiology. In addition to this objective input, which includes demographics, medical history and current physiologic data, a physician's training, past experiences, clinical gestalt and even their "gut" feeling about a patient – which can collectively be considered the physician's intuition – all likely play an important role in clinical assessment.

Across many specialties, including surgery, provider intuition is hypothesized to play a significant role in clinical decision-making.<sup>2,3</sup> There is considerable debate regarding the utility of physician intuition in stratifying risk and guiding appropriate care. In the book *Noise: A Flaw in Human Judgment*, Nobel laureate Daniel Kahneman argues that human decision-making is often irrational and clouded by extraneous details.<sup>4</sup> Literature in surgery and other specialties has also exposed how several common cognitive biases and counterproductive heuristics affect clinical decision-making.<sup>2,3,5</sup> Despite these recognized shortcomings of intuition, there is also acknowledgement that clinicians are capable of capturing and synthesizing important information about a patient for which current statistical models cannot account, such as a nuanced characterization of patients' functional status and social situation, indistinct signs of distress or change in clinical status, and/or subtle implications based on overall appearance.

Much of this debate is theoretical; the true prognostic value of physician intuition remains largely unexplored. Within the realm of clinical prediction models, intuition is rarely used as a source of information. In fact, physician intuition and clinical prediction models are often viewed as adversarial. Many published prediction models benchmark their performance against the predictive capacity of a human physician, and consider their model to have real-world utility if it exceeds human performance.<sup>6</sup> Few prior studies have managed to merge these two sources of data to construct a model that exceeds the predictive power of either alone.

In part, this is likely because individual physician gestalt is inherently dynamic and difficult to measure directly. Some success has been had in estimating human intuition using surrogate markers found in the electronic health record. One study found that the volume of diagnostic imaging tests ordered by a physician corresponded with their assessment of the

patient's prognosis: physicians who exhibited data-seeking behavior, manifested by ordering imaging tests, usually thought the patient was at risk of poor outcomes.<sup>7</sup> Another study found that the time of day during which a physician chooses to order a laboratory test is more strongly associated with patient outcome than the actual value of the test result; for instance, a white blood cell count (WBC) ordered by a physician in the middle of the night, representing deviation from more routine testing, is more predictive of a poor outcome than an abnormal WBC value or a WBC ordered during the day.<sup>8</sup>

Recognizing the potential value of this data, attempts have been made to incorporate intuition into surgical risk calculators; however, the results are mixed. The ACS NSQIP surgical risk calculator allows surgeons to adjust a patient's risk of postoperative complications based on their intuition, but incorporates the provider's intuition into the risk calculation imprecisely. Further, the efficacy of this method of adjustment has never been prospectively studied.<sup>9</sup> Another study used a combination of physician intuition and patient physiology (vital signs, laboratory values, and other objective data elements) to predict the likelihood of inpatient admission from a pediatric emergency department, finding that models trained on a combination of intuition and physiology outperformed all other models.<sup>10</sup> Lastly, one prospective study among surgical patients compared the surgeon's risk assessment with the output of a validated risk calculator, and secondarily assessed whether a surgeon's risk assessment improved after being provided the output from the standardized risk algorithm. The authors found that the risk calculator was more accurate in predicting risk for complications compared to physician assessment, and while many physicians modified their risk assessment after interacting with the standardized algorithm, the change in the accuracy of the physician's assessment did not reach clinical significance.<sup>11</sup>

In this study, we sought to develop and validate prediction models to compare the prognostic value of surgeon intuition and clinical attributes traditionally used in predicting the likelihood of any postoperative complication among surgical patients at our institution. We also sought to combine surgeon intuition and clinical factors to better understand the prognostic utility of surgeon intuition in relation to patient-specific clinical data. We hypothesized that incorporating surgeon intuition into surgical risk models would strengthen the overall risk prediction.

## Methods

### Study Design

In this study, we prospectively surveyed general surgeons between 10/1/21 and 9/1/22 at one academic medical center (Beth Israel Deaconess Medical Center [BIDMC]) immediately before performing surgery to assess their intuition and/or assessment of their patient's risk for any postoperative complication, as defined by ACS NSQIP. Concurrently, clinical data, as used in the ACS NSQIP surgical risk calculator, were abstracted from an institutional registry for all eligible patients who had surgery between 1/1/17 to 9/1/22. Binomial regression models to predict postoperative complication were independently trained using (a) the surgeon's preoperative intuition and (b) patient-specific clinical data, currently used in the ACS NSQIP surgical risk calculator. This study was approved by the institutional review board of BIDMC (2021P000484).

### Data Collection - Clinical Data

Clinical data were retrospectively abstracted from our institutional ACS NSQIP database<sup>12</sup> which includes all patients undergoing general surgery. The database includes the following baseline variables as defined by ACS NSQIP: age (continuous), sex (dichotomous), functional status (ordinal), emergency case (dichotomous), American Society of Anesthesiologists (ASA) classification (ordinal), body mass index (BMI) (continuous), wound class (ordinal), steroid use (dichotomous), ascites within 30 days prior to surgery (dichotomous), systemic sepsis within 48 hours of surgery (dichotomous), dependent on mechanical ventilation (dichotomous), disseminated cancer (dichotomous), diabetes (nominal), hypertension requiring medication (dichotomous), congestive heart failure (CHF) within 30 days prior to surgery (dichotomous), dyspnea (ordinal), current smoker within one year (dichotomous), severe chronic obstructive pulmonary disease (COPD) (dichotomous), dialysis dependence (dichotomous), and acute renal failure (dichotomous). Functional status is classified as independent, partially dependent or totally dependent. Diabetes is classified as either insulin-dependent or non-insulin-dependent. CHF refers to newly diagnosed CHF within 30 days of surgery or chronic CHF with signs and symptoms of CHF within the 30 days prior to surgery. Severe COPD is defined as one of the following: functional disability secondary to COPD, prior hospitalization secondary to COPD, chronic bronchodilator therapy or forced expiratory volume (FEV1) of <75% predicted.

Surgical operations were classified using procedural codes. Operations were further classified as elective or non-elective. Elective cases included non-urgent, non-emergent cases, scheduled from the outpatient setting. Non-elective cases included urgent and emergent cases, with priority designation ranging from immediate to within 24 hours of booking.

The ACS NSQIP registry records 30-day postoperative morbidity and mortality, including the occurrence of any of the following complications: superficial surgical site infection (SSI), deep incisional SSI, organ space SSI, wound disruption, pneumonia, unplanned intubation, pulmonary embolism, mechanical ventilation requirement for greater than 48 hours after surgery, progressive renal insufficiency, acute renal failure, urinary tract infection (UTI), cerebrovascular accident, cardiac arrest requiring cardiopulmonary resuscitation (CPR), myocardial infarction, blood transfusion requirement within 72 hours after surgery, vein thrombosis requiring therapy, *Clostridium difficile* infection, sepsis or septic shock, death within 30 days of surgery, unplanned readmissions within 30 days of surgery, and unplanned reoperation within 30 days of surgery. Trained surgical clinical reviewers complete all chart review and data abstraction for the ACS NSQIP program.

Details regarding each of the data variables can be found in the ACS NSQIP data dictionary.<sup>13</sup> ACS NSQIP data was available and complete for all patients.

### Data Collection - Surgeon Intuition

To assess surgeons' preoperative assessment of each patient's surgical risk, we administered a prospective single-question survey immediately prior to surgery in which we asked

participating surgeons to rate the patient's risk of 30-day morbidity and/or mortality compared to the average risk across all patients undergoing the procedure. Surgeons were asked to select one of the following responses: "lower than average risk," "average risk," or "higher than average risk". The single-question survey was designed to model the surgeon risk adjustment which is currently available in the ACS NSQIP surgical risk calculator and enables surgeons to adjust the estimated risk based on factors that are not already entered into the risk calculator. Participating surgeons were not provided data on the ACS NSQIP risk calculator outputs for their patients. Surgeons were instructed to continue their existing practice of assessing risk pre-operatively. Assessment of risk via the ACS NSQIP calculator may have been completed for a subset of patients, though this completion of this step was not recorded as part of the current study protocol.

Study data were collected and managed using REDCap (Research Electronic Data Capture) tools hosted at our institution.<sup>14,15</sup> REDCap is a secure, web-based software platform designed to support data capture for research studies, providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for data integration and interoperability with external sources. Participating surgeons were sent a short messaging service (SMS) message with the survey link immediately before each surgery. The survey instrument is shown in Supplemental Digital Content 1.

### Statistical Analysis

Based on medical record number (MRN), surgeon survey responses were matched with the patient's clinical and postoperative outcome data to create a complete dataset of surgeon intuition, baseline patient clinical data and postoperative 30-day outcomes. Postoperative 30-day outcomes were aggregated to create a single composite variable to indicate whether the patient experienced any 30-day postoperative complications as defined by the ACS NSQIP.

Using our institutional ACS NSQIP database, we performed multivariate lasso regression analysis to identify clinical features in the existing ACS NSQIP surgical risk calculator that are associated with increased risk of any postoperative complication in our study cohort.

Independently, the preoperative surgeon intuition data and the ACS NSQIP clinical data were used to build binomial logistic regression models to predict the likelihood of experiencing any 30-day postoperative complication. For the model trained using clinical data alone, logistic regression was performed with the following independent variables, as currently used by the ACS NSQIP surgical risk calculator: age, sex, BMI, principal procedural code, wound class, functional status, emergency case, ASA classification, steroid use, ascites within 30 days, systemic sepsis within 48 hours, ventilator dependence, disseminated cancer, diabetes, hypertension, CHF within 30 days, current smoker within one year, severe COPD, dialysis dependence and acute renal failure. For the model trained using surgeon intuition, the outcome remained the composite measure of any 30-day complication and the only exposure was the surgeon's assessment of the patient's postoperative risk, classified as "lower than average," "average risk," or "higher than average".

A third model was trained using the surgeon's preoperative risk assessment and a subset of clinical variables identified based on the feature importance of the lasso regression analysis. Given the smaller number of patients with both data elements, the model could not be effectively trained on all of the clinical inputs included in the ACS NSQIP surgical risk calculator. Ultimately, the following variables were included: age (continuous), sex (dichotomous), functional health status (ordinal), ASA classification (ordinal), CHF within 30 days prior to surgery (dichotomous), current smoker within one year (dichotomous) and history of severe COPD (dichotomous).

In addition to the above analysis incorporating all data, we performed two subgroup analyses. These were designed to assess differences in surgeon intuition among more and less experienced surgeons, as well as in the context of elective and non-elective surgeries. To assess provider experience, we stratified the responses by provider role: (a) surgical attendings and (b) fellows and chief residents. Predictive models were also completed independently for elective and non-elective operations.

For model development, data were randomly divided into training and test sets with 80:20 split. Model performance was measured as the area under the receiver operating characteristic curve (AUC) using 5-fold cross-validation. Statistical analysis was performed using R version 3.6.3.

A study framework is shown in Supplemental Digital Content 2. We developed, validated, and reported each model in accordance with the Equator Network Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines.<sup>16</sup>

## Results

Preoperative surgeon intuition was obtained for 216 patients, representing an 18.8% response rate among all surgeries and a 34.9% response rate among surgeries in which an SMS text with the survey link was sent to the surgeon. Retrospective clinical data from the ACS NSQIP database were available for 9182 patients who underwent surgery at our institution. In total, 216 patients had baseline clinical data, surgeon intuition data and 30-day outcome data (Supplemental Digital Content 3).

Patient demographics and baseline clinical characteristics are presented in Table 1.

Among all patients with abstracted ACS NSQIP data, the most common operations were laparoscopic cholecystectomy (N=585), laparoscopic appendectomy (N=579), laparoscopic partial colectomy with anastomosis (N=451), and laparoscopic partial colectomy with low pelvic anastomosis (N=340). Among patients with preoperative intuition data, the most common surgical operations were laparoscopic cholecystectomy (N=45), laparoscopic appendectomy (N=13), parathyroidectomy (N=12) and thyroidectomy (N=8) (Table 2).

With regard to surgical outcomes, 19.8% of all patients and 24.1% of patients with intuition data had a postoperative complication (Table 2). Among all patients, the most common complications were organ space SSI (N=400, 4.4%), ventilator dependence for >48 hrs

(N=226, 2.5%) and superficial incisional SSI (N=181, 2.0%). Approximately 8% of all patients (N=701) and 3.2% of patients with intuition data were readmitted within 30 days. Surgical re-operation occurred in 3.6% of all patients (N=326) and 1.4% of patients with intuition data. Mortality within 30-days of surgery was 1.4% for all patients and 3.2% for patients with intuition data.

Exactly half of survey responses (50%, N=108) were from attending surgeons, 21.3% (N=46) were from clinical fellows in surgery, and the remaining 28.7% (N=62) were post-graduate year 5 resident physicians in general surgery (Table 3). Among attending surgeons, the median (IQR) years of attending experience was 8.5 years (5.3-11.0). The majority of respondents were associated with the acute care surgery service (72.7%). Nearly 60% of cases were non-elective.

Nearly half of surgeon respondents (45.4%, N=98) indicated that their patient's preoperative risk of any postoperative complication was "average risk", with 40.3% (N=87) responding "higher than average" and the remaining 14.4% (N=31) responding "lower than average" (Table 4).

We performed multivariate lasso regression analysis to identify the clinical attributes in the baseline ACS NSQIP surgical risk predictor associated with increased risk of any postoperative complication in our study cohort (N=9182 patients) (Table 5). After shrinkage, the following variables were most strongly associated with risk of any complication: ASA classification (coefficient=1.032); emergency operation (coefficient=0.562); disseminated cancer (coefficient=0.502); functional health status (coefficient=0.430); CHF (coefficient=0.397); immunosuppressive therapy (coefficient=0.321); and active smoker (coefficient=0.245).

A series of logistic regression models were trained using different covariates to predict the risk of any 30-day postoperative complication; the results of internal validation are shown in Table 6. The model trained on preoperative clinical data alone was associated with an AUC of 0.828 (95% CI: 0.802-0.853), as well as a sensitivity and specificity of 0.970 and 0.258, respectively. Use of preoperative intuition data alone resulted in an AUC of 0.700 (95% CI: 0.624-0.778); the model exhibited improved specificity (0.711) but substantially decreased sensitivity (0.695). Combining surgeon intuition with a subset of important patient attributes and clinical characteristics resulted in an AUC of 0.825 (CI: 0.765-0.885), with specificity and sensitivity of 0.392 and 0.926, respectively. A comparison of baseline data for the development and validation cohorts is presented in Supplemental Digital Content 4.

The complete output of the three models are presented in Supplemental Digital Content 5, 6 and 7. Regarding the baseline clinical model, the following features were significantly associated with postoperative complication: totally dependent functional health status (coefficient 1.57, p=0.02); ASA class 3 (coefficient 0.08, p=0.013); ASA class 4 (coefficient 1.28, p<0.001); emergency case (coefficient 0.574, p=0.006); dyspnea with moderate exertion (coefficient 2.81, p=0.006), ascites (coefficient 1.57, p<0.001), history of severe COPD (coefficient 0.039, p=0.036), disseminated cancer (coefficient 0.039, p=0.04),

preoperative dialysis dependence (coefficient 1.36,  $p=0.001$ ), and CHF (coefficient 0.061,  $p\text{-value}=0.047$ ).

As shown in Supplemental Digital Content 7, preoperative surgeon intuition (coefficient 1.24,  $p<0.001$ ) remains a significant predictor of postoperative complications, even when combined with patient demographics and important clinical attributes. It should be noted that this model was evaluated among the 216 patients with both data elements and thus, many clinical elements did not reach significance, which contrasts the baseline model that is based on 9182 data points.

The outcomes of the subgroup analysis by provider role and surgical urgency (i.e., elective versus non-elective) is presented in Tables 7 and 8. With regard to provider experience, the intuition-only model derived from attendings' responses achieved an AUC of 0.718 (95% CI: 0.637-0.800), which substantially outperformed the model derived from trainees' responses (AUC 0.564, 95% CI: 0.408-0.720). No differences were observed for the combined intuition and clinical model based on respondent experience (attendings: AUC = 0.860, 95% CI 0.792-0.930 and fellow and resident: AUC = 0.868, 95% CI: 0.771-0.965). In terms of surgical urgency, intuition-only models derived from elective (N=89) and non-elective (N=129) cases exhibited an AUC of 0.708 (95% CI: 0.490-0.926) and 0.652 (0.564-0.739), respectively. A similar trend was observed for combined intuition and clinical models (elective cases: AUC = 0.885, 95% CI 0.802-0.967 and non-elective cases: AUC = 0.815, 95% CI: 0.742-0.888).

## Discussion

In this study, we assessed preoperative surgeon intuition via a single-question survey in order to determine the predictive accuracy of surgeon intuition, compared to ACS NSQIP risk prediction. Principally, we found that preoperative surgeon intuition is an independent predictor of 30-day postoperative complications among surgical patients; however, the predictive performance was inferior to traditional clinical models, as derived from the ACS NSQIP risk calculator. Combining clinical data and preoperative intuition did not markedly improve prediction; however, the study size may not have been adequate to identify this outcome.

These findings underscore the independent prognostic value of physician intuition, indicating that a surgeon's preoperative assessment of a patient's postoperative risk is an alternate and acceptable mechanism of assessing surgical risk. However, surgeon intuition alone was inferior to the predictive value of a model derived from the ACS NSQIP risk calculator. Additionally, combining clinical data with surgeon intuition did not markedly improve the performance of the risk calculator. This may be secondary to inherent limitations in our study, or it could suggest that there is a large degree of redundancy between surgeon intuition and objective clinical attributes, such as patient demographics, functional status and clinical status. Nonetheless, compared to the arbitrary adjustment available within the NSQIP risk calculator, the current analysis provides a framework for more precise integration of surgeon intuition into surgical risk prediction. The predictive value of a surgeon's intuition is likely impacted by surgeon experience and



the operation type, given the varied occurrence of post-operative complications between surgical operations. Sub-group analysis among attending surgeons and surgical trainees revealed that the preoperative risk assessment of attendings was substantially more robust in predicting post-operative complications than the risk assessment of trainees. Additional subgroup analysis among elective and non-elective cases identified minor differences in point estimates. Higher performance was observed for elective cases although this was not significantly different, as the wide confidence intervals overlap. This observation is somewhat unexpected, as we initially hypothesized that surgeon intuition may play a larger role among urgent and emergent cases, where structured clinical data may not capture the patient physiology and surgical risk. Alternatively, for elective cases, surgeons may have access to more robust non-structured data, derived from the outpatient work-up and previous patient encounters. Larger samples, as discussed below, may further reveal clinical domains in which surgeon intuition has greater benefit—overall and relative to other traditional risk metrics. Future research is also needed to investigate divergence between surgeon prediction and patient outcomes in order to understand the potential value and/or misdirection of surgeon intuition.

Compared to prior attempts at measuring intuition,<sup>7,8</sup> which have generally relied on measuring downstream actions, such as physician orders, as proxies of clinical judgment, our study directly and independently assessed a surgeon's preoperative assessment of the patient's overall risk for postoperative complication. Measuring physician intuition in this fashion may help meaningfully incorporate intuition into future clinical decision support tools.<sup>11</sup> This approach does not bias the physician's assessment, and permits the surgeon to selectively prioritize specific data attributes.

Our approach to risk assessment complements prior research by our group, which assessed the value of clinician-initiated data, defined as data elements created through specific actions or insights of the clinician.<sup>17</sup> The authors found that clinician-initiated data is a filtered representation of patient physiology and may capture most of the inherent value of clinical models. Specifically, a machine learning model trained on clinician-initiated administrative data alone to predict mortality, readmission and length of stay achieved comparable performance to models trained on complete electronic medical record data. Based on these findings, it is unsurprising that in our current study, the model with intuition and clinical data did not substantially outperform either model alone. While most of the inputs to the ACS NSQIP risk calculator are not clinician-initiated, the decision to operate, except for a small subset of urgent and emergent operations, is a strong indication of a surgeon's risk assessment. Surgical decision-making is informed by structured clinical data and thus, among most patients who receive surgical intervention, the output of the ACS NSQIP risk calculator reflects surgeon intuition. Interestingly, within the ACS NSQIP-derived risk calculator, we found that ASA class—a subjective yet directed measure of a patient's preoperative illness severity—was the strongest predictor of experiencing a postoperative complication. Future research is needed to capture aspects of surgeon intuition, which are not reflected in structured EMR data or clinical decision-making, and delineate the role and value of intuition compared to established clinical attributes such as ASA class, functional status, comorbidities and demographics.

A previous study comparing surgeon intuition with a standardized risk algorithm found that the risk calculator, derived from structured, patient-specific clinical elements, was more accurate in predicting risk for postoperative complication.<sup>11</sup> This is consistent with our findings, as we observed slightly superior predictive performance for a risk model using clinical attributes alone compared to a second model using surgeon intuition. While assessment of intuition was assessed uniquely between the two studies, and the clinical inputs included in the model varied across the studies, the consistency in results suggests that preoperative intuition alone is not optimal for predicting risk of postoperative complication. Still, the role of surgeon intuition may vary depending on the specific clinical context and patient population in which it is employed as a predictor of risk. Further research is needed to measure surgeon intuition and assess its role in risk prediction.

In addition, it is notable that more than 85% of surgeon respondents indicated that their patient had “average” or “higher than average” risk for 30-day postoperative complication relative to all patients undergoing the same surgical operation. The study cohort is derived from a tertiary, academic medical center, which may explain this observed skew toward higher perceived risk. Additionally, the majority of surgeons who responded to the survey were acute care surgeons, and most operations were not electively scheduled. It is unsurprising that their preoperative assessment of most patients was “average” or “higher than average” as it is well-established that non-elective operations generally involve patients who are more acutely ill and more likely to experience postoperative complications, relative to the general population.<sup>18-20</sup> Relative to the entire ACS NSQIP cohort, we observed higher mortality within 30 days (3.2% vs 1.4%) and a higher incidence of postoperative adverse events (24.1% vs 19.8%) among patients with corresponding intuition data—suggesting that this subgroup was higher risk than the entire ACS NSQIP cohort.

This study has several limitations. First, we modeled preoperative intuition as an ordinal variable and restricted possible survey responses to three categories (higher than average, average, and lower than average) for simplicity and ease of data collection. However, there are undoubtedly alternative mechanisms to characterize surgeon intuition and statistically incorporate this data into regression analysis and predictive modeling. Second, while we sought to directly compare the prognostic value of surgeon intuition and a structured risk calculator such as the ACS NSQIP calculator, our study design combined retrospectively collected clinical data and adverse events and prospectively collected data on surgeon intuition. Predictive performance degradation when incorporating retrospective and prospective data elements has been well-documented, including with regard to predicting postoperative outcomes.<sup>21</sup> Thus, despite our stated objective, this analysis likely represents an indirect comparison of the two approaches to patient risk assessment. Third, the study may be limited by sample size. Our overall response rate was low, which likely reflects the demands of patient care priorities, limiting a surgeon’s capacity to complete a non-critical survey immediately prior to surgery. Our data include responses from providers of varying clinical experience, as well as a combination of elective and urgent/emergent operations, with varying expectations for the occurrence of post-operative adverse events. The value of a surgeon’s intuition likely depends on the acuity of the case, the frequency of postoperative complications and the experience of the provider. Our study was likely not sufficiently powered, as evidenced by the wide confidence intervals, to effectively perform subset

analysis in order to investigate the role of surgeon intuition in distinct contexts. While we performed a stratified analysis, future prospective research, with larger samples, is needed to assess the relative value of more and less experienced surgeons. We did not perform a power analysis prior to commencing the study, a further limitation, as this study was principally an exploratory analysis, which is a limitation of the current study. Future research is needed to quantify the independent effect size of surgeon intuition and clinical data. Finally, while modeled after the ACS NSQIP risk calculator, it is important to acknowledge that the clinical variables used in this study may not fully capture the patient's entire physiology, baseline health, and acuity of presentation; other objective measurements of risk may exist that we did not incorporate in our model, given our aim to reproduce the ACS NSQIP risk calculator.

## Conclusions

In this study, we effectively measured preoperative surgeon intuition and found that surgeon intuition alone is an independent predictor of postoperative outcomes among surgical patients, with a predictive accuracy that is slightly less than established clinical prediction models. Models trained on both surgeon intuition and clinical data did not outperform a model derived from clinical data alone, although a larger sample may be needed to delineate significant differences. Nonetheless, this analysis suggests that surgeon intuition can be assessed and that it acceptably predicts postoperative outcomes. Additional research is needed to determine the value of surgeon intuition among certain providers and under specific clinical circumstances, and how best to incorporate intuition data into clinical risk prediction and postsurgical care of patients.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgement

The authors thank Dr. Kevin Schuster, MD, MPH, FACS, and the New England Surgical Society Publication Committee for their support and contributions to this manuscript.

Support: Drs Marwaha and Beaulieu-Jones are supported by National Library of Medicine/NIH grant [T15LM007092].

## References

1. Gao J, Xiao C, Glass LM, Sun J. Dr. Agent: Clinical predictive model via mimicked second opinions. *JAMIA*. 2020;27(7):1084–1091. doi:10.1093/JAMIA/OCAA074 [PubMed: 32548622]
2. Patel SH, Itri JN. The Role of Intuitive Cognition in Radiologic Decision Making. *Journal of the American College of Radiology : JACR*. 2022;19(5):669–676. doi:10.1016/J.JACR.2022.02.027 [PubMed: 35346618]
3. Hughes TM, Dossett LA, Hawley ST, Telem DA. Recognizing Heuristics and Bias in Clinical Decision-making. *Ann Surg*. 2020;271(5):813–814. doi:10.1097/SLA.0000000000003699 [PubMed: 31850992]
4. Kahneman D, Sibony O, Sunstein C. *Noise: A Flaw in Human Judgment*. Little, Brown and Company; 2021.

5. Olenski AR, Zimerman A, Coussens S, Jena AB. Behavioral Heuristics in Coronary-Artery Bypass Graft Surgery. *The New England journal of medicine*. 2020;382(8):778–779. doi:10.1056/NEJMC1911289 [PubMed: 32074429]
6. Marwaha JS, Beaulieu-Jones B, Yuan W, Brat GA. Comment on: Truth and truthiness: evidence, experience and clinical judgement in surgery. *The British journal of surgery*. 2021;108(12):e417. doi:10.1093/BJS/ZNAB319 [PubMed: 34529769]
7. Ghassemi MM, Al-Hanai T, Raffa JD, et al. How is the Doctor Feeling? ICU Provider Sentiment is Associated with Diagnostic Imaging Utilization. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual International Conference*. 2018;2018:4058–4064. doi:10.1109/EMBC.2018.8513325
8. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ (Clinical research ed)*. 2018;361. doi:10.1136/BMJ.K1479
9. American College of Surgeons. ACS Surgical Risk Calculator.
10. Barak-Corren Y, Agarwal I, Michelson KA, et al. Prediction of patient disposition: comparison of computer and human approaches and a proposed synthesis. *JAMIA*. 2021;28(8):1736–1745. doi:10.1093/JAMIA/OCAB076 [PubMed: 34010406]
11. Brennan M, Puri S, Ozrazgat-Baslanti T, et al. Comparing clinical judgment with the MySurgeryRisk algorithm for preoperative risk assessment: A pilot usability study. *Surgery*. 2019;165(5):1035–1045. doi:10.1016/J.SURG.2019.01.002 [PubMed: 30792011]
12. American College of Surgeons. ACS National Surgical Quality Improvement Program. <https://www.facs.org/quality-programs/data-and-registries/acs-nsqip/>.
13. American College of Surgeons. ACS NSQIP Participant Use Data File.
14. Harris PA, Taylor R, Thielke R, et al. A metadata-driven methodology and workflow process for providing translational research informatics support, *J Biomed Inform*. *J Biomed Inform*. Published online 2009.
15. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform*. Published online 2019. doi:10.1016/j.jbi.2019.103208
16. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ (Clinical research ed)*. 2015;350. doi:10.1136/BMJ.G7594
17. Beaulieu-Jones BK, Yuan W, Brat GA, et al. Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians? *NPJ Digit Med*. 2021;4(1). doi:10.1038/S41746-021-00426-3
18. Mullen MG, Michaels AD, Mehaffey HJ, et al. Risk Associated With Complications and Mortality After Urgent Surgery vs Elective and Emergency Surgery: Implications for Defining “Quality” and Reporting Outcomes for Urgent Surgery. *JAMA surgery*. 2017;152(8):768–774. doi:10.1001/JAMASURG.2017.0918 [PubMed: 28492821]
19. Ingraham AM, Cohen ME, Bilimoria KY, et al. Comparison of 30-day outcomes after emergency general surgery procedures: potential for targeted improvement. *Surgery*. 2010;148(2):217–238. doi:10.1016/J.SURG.2010.05.009 [PubMed: 20633727]
20. Kassahun WT, Babel J, Mehdorn M. Assessing differences in surgical outcomes following emergency abdominal exploration for complications of elective surgery and high-risk primary emergencies. *Scientific reports*. 2022;12(1). doi:10.1038/S41598-022-05326-4
21. Ren Y, Loftus TJ, Datta S, et al. Performance of a Machine Learning Algorithm Using Electronic Health Record Data to Predict Postoperative Complications and Report on a Mobile Platform. *JAMA Netw Open*. 2022;5(5). doi:10.1001/JAMANETWORKOPEN.2022.11973

**Table 1.**

## Patient Demographics and Baseline Clinical Attributes

| Variable                           | NSQIP data<br>(n = 9182) | Intuition data<br>(n = 216) |
|------------------------------------|--------------------------|-----------------------------|
| Age, y, median (IQR)               | 60.6 (37.9-83.3)         | 65.0 (40.7-89.3)            |
| Sex, f, n (%)                      | 5031 (54.8)              | 120 (55.6)                  |
| Race, White, n (%)                 | 6596 (71.8)              | 138 (63.8)                  |
| Hispanic ethnicity, n (%)          | 598 (6.5)                | 21 (9.7)                    |
| Diabetes mellitus, n (%)           |                          |                             |
| Insulin-dependent                  | 468 (5.1)                | 11 (5.1)                    |
| Non-insulin dependent              | 704 (7.7)                | 35 (16.2)                   |
| Current smoker, n (%)              | 1003 (10.9)              | 21 (9.7)                    |
| Dyspnea <sup>*</sup> , n (%)       | 147 (1.6)                | 6 (2.8)                     |
| Functional health status, n (%)    |                          |                             |
| Independent                        | 8976 (97.8)              | 201 (93.1)                  |
| Partially dependent                | 130 (1.3)                | 13 (6.0)                    |
| Totally dependent                  | 30 (0.3)                 | 2 (0.9)                     |
| History of severe COPD, n (%)      | 393 (4.3)                | 16 (7.4)                    |
| Ascites, n (%)                     | 76 (0.8)                 | 2 (0.9)                     |
| Heart failure, n (%)               | 142 (1.5)                | 23 (10.6)                   |
| Hypertension, on medication, n (%) | 4096 (44.6)              | 125 (57.9)                  |
| Acute renal failure, n (%)         | 62 (0.7)                 | 6 (2.8)                     |
| Preoperative dialysis, n (%)       | 106 (1.2)                | 13 (6.0)                    |
| Disseminated cancer, n (%)         | 369 (4.0)                | 10 (4.6)                    |
| Open wound, n (%)                  | 94 (1.0)                 | 10 (4.6)                    |
| Immunosuppressive therapy, n (%)   | 647 (7.0)                | 15 (6.9)                    |
| Malnourishment, n (%)              | 164 (1.8)                | 7 (3.2)                     |

\* At rest or with moderate exertion

IQR, interquartile range.

**Table 2 –**  
Operative Procedure and Postoperative Complication

| Variable                                                       | NSQIP data<br>(n = 9182) | Intuition data<br>(n = 216) |
|----------------------------------------------------------------|--------------------------|-----------------------------|
| Operative procedure                                            |                          |                             |
| Laparoscopic cholecystectomy                                   | 585 (6.4)                | 45 (20.8)                   |
| Laparoscopic appendectomy                                      | 579 (6.3)                | 13 (6.0)                    |
| Laparoscopic partial colectomy, with anastomosis               | 451 (4.9)                | ---                         |
| Laparoscopic partial colectomy, with low pelvic anastomosis    | 340 (3.7)                | ---                         |
| Thyroidectomy, total or complete                               | 318 (3.5)                | 8 (3.7)                     |
| Repair initial inguinal hernia                                 | 291 (3.2)                | ---                         |
| Parathyroidectomy or exploration of parathyroid                | 291 (3.2)                | 12 (5.6)                    |
| Mastectomy, partial                                            | 277 (3.0)                | 3 (1.4)                     |
| Laparoscopic partial colectomy, with removal of terminal ileum | 234 (2.5)                | 1 (0.5)                     |
| Total thyroid lobectomy, unilateral                            | 231 (2.5)                | 2 (0.9)                     |
| Enterectomy, resection of small intestine                      | 224 (2.4)                | 9 (4.2)                     |
| Repair initial incisional or ventral hernia                    | 206 (2.2)                | 3 (1.4)                     |
| Pylorus-sparing, Whipple-type procedure                        | 185 (2.0)                | 1 (0.5)                     |
| Repair umbilical hernia, reducible                             | 180 (2.0)                | ---                         |
| Colectomy, partial; with anastomosis                           | 160 (1.7)                | 6 (2.8)                     |
| Closure of enterostomy, large or small intestine               | 142 (1.5)                | ---                         |
| Other                                                          | 4488 (48.9)              | 113 (52.3)                  |
| Postoperative complication                                     |                          |                             |
| Superficial incisional SSI                                     | 181 (2.0)                | 1 (0.5)                     |
| Deep incisional SSI                                            | 19 (0.2)                 | 1 (0.5)                     |
| Organ space SSI                                                | 400 (4.4)                | 10 (4.6)                    |
| Wound disruption                                               | 46 (0.5)                 | 12 (5.6)                    |
| Pneumonia                                                      | 112 (1.2)                | 6 (2.8)                     |
| Unplanned intubation                                           | 81 (0.9)                 | 4 (1.9)                     |
| Pulmonary embolism                                             | 30 (0.3)                 | 0 (0)                       |
| Ventilator >48 h                                               | 226 (2.5)                | 17 (7.9)                    |
| Renal insufficiency                                            | 115 (1.3)                | 11 (5.1)                    |
| Acute renal failure                                            | 49 (0.5)                 | 2 (0.9)                     |
| Urinary tract infection                                        | 105 (1.1)                | 4 (1.9)                     |
| Cerebrovascular accident                                       | 12 (0.1)                 | 0 (0)                       |
| Cardiac arrest with CPR                                        | 39 (0.4)                 | 0 (0)                       |
| Myocardial infarction                                          | 42 (0.5)                 | 0 (0)                       |
| Deep venous thrombosis *                                       | 80 (0.9)                 | 4 (1.9)                     |
| <i>Clostridium difficile</i> infection                         | 60 (0.7)                 | 3 (1.4)                     |
| Mortality within 30 d                                          | 125 (1.4)                | 7 (3.2)                     |

| Variable                        | NSQIP data<br>(n = 9182) | Intuition data<br>(n = 216) |
|---------------------------------|--------------------------|-----------------------------|
| Readmission within 30 d         | 701 (7.6)                | 7 (3.2)                     |
| Reoperation within 30 d         | 326 (3.6)                | 3 (1.4)                     |
| Any postoperative adverse event | 1815 (19.8)              | 52 (24.1)                   |

Data presented as n (%)

\* Deep venous thrombosis requiring therapy, as defined by ACS NSQIP

SSI, surgical site infection

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3 –**

## Respondent Characteristics

| Characteristic             | n (%)<br>(n = 216) |
|----------------------------|--------------------|
| Training level             |                    |
| Attending                  | 108 (50.0)         |
| Fellow                     | 46 (21.3)          |
| PGY-5                      | 62 (28.7)          |
| Service                    |                    |
| Acute care surgery         | 157 (72.7)         |
| Surgical oncology          | 40 (18.5)          |
| Minimally invasive surgery | 11 (5.1)           |
| Vascular surgery           | 8 (3.7)            |
| Surgical urgency           |                    |
| Elective                   | 87 (40.3)          |
| Non-elective               | 129 (59.7)         |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 4.**

## Distribution of Preoperative Surgeon Risk Assessment

| Response                 | n (%)      |
|--------------------------|------------|
| Preoperative risk        |            |
| Higher than average risk | 87 (40.3%) |
| Average risk             | 98 (45.4%) |
| Lower than average risk  | 31 (14.4%) |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 5.**

Multivariate Lasso Regression Analysis to Identify Clinical Features Associated with Increased Risk for Any Postoperative Complication in Our Institutional Patient Population

| Clinical feature*                      | Beta coefficient |
|----------------------------------------|------------------|
| ASA classification (ordinal)           | 1.032            |
| Emergency operation                    | 0.562            |
| Disseminated cancer                    | 0.502            |
| Functional health status (ordinal)     | 0.430            |
| Congestive heart failure               | 0.397            |
| Preoperative immunosuppressive therapy | 0.321            |
| Active smoker                          | 0.245            |
| History of severe COPD                 | 0.242            |
| Acute renal failure                    | 0.062            |
| Preoperative malnourishment            | 0.006            |
| Age (continuous)                       | 0.003            |
| Hypertension                           | Excluded         |
| Sex                                    | Excluded         |
| BMI (continuous)                       | Excluded         |
| Diabetes (ordinal)                     | Excluded         |
| Dialysis dependence                    | Excluded         |

\* Features are dichotomous unless otherwise specified

ASA, American Society of Anesthesiologists

**Table 6.**

Internal Validation Performance for a Baseline Model to Predict the Probability of Any Postoperative Complication, Independently Using Surgeon Intuition, Baseline Clinical Attributes, and a Combination of Surgeon Intuition and Baseline Clinical Attributes

| Model                                           | AUC (95% CI)        | Sensitivity | Specificity |
|-------------------------------------------------|---------------------|-------------|-------------|
| Preoperative clinical data                      | 0.828 (0.802-0.853) | 0.970       | 0.258       |
| Preoperative intuition data alone               | 0.700 (0.624-0.778) | 0.695       | 0.711       |
| Preoperative intuition data and clinical data * | 0.825 (0.765-0.885) | 0.926       | 0.392       |

\* Subset of clinical data: age (continuous), sex (dichotomous), functional health status (ordinal), American Society of Anesthesiologists classification (ordinal), congestive heart failure within 30 days of operation (dichotomous), current smoker within 1 year (dichotomous) and history of severe COPD (dichotomous).

AUC, area under the receiver operating characteristic curve

**Table 7.**

Stratified Analysis by Role of Respondent: Attending Surgeon vs Trainee

| Model                                           | AUC (95% CI)        | Sensitivity | Specificity |
|-------------------------------------------------|---------------------|-------------|-------------|
| Preoperative intuition data alone               |                     |             |             |
| Surgical attendings (n = 108)                   | 0.718 (0.637-0.800) | 0.569       | 0.861       |
| Surgical fellows and chief residents (n = 108)  | 0.564 (0.408-0.720) | N/A         | N/A         |
| Preoperative intuition data and clinical data * |                     |             |             |
| Surgical attendings (n = 108)                   | 0.860 (0.792-0.930) | 0.889       | 0.543       |
| Surgical fellows and chief residents (n = 108)  | 0.868 (0.771-0.965) | 0.956       | 0.375       |

\* Subset of clinical data: age (continuous), sex (dichotomous), functional health status (ordinal), American Society of Anesthesiologists classification (ordinal), congestive heart failure within 30 days of operation (dichotomous), current smoker within 1 year (dichotomous) and history of severe COPD (dichotomous).

AUC, area under the receiver operating characteristic curve; N/A, not applicable

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 8.**

Stratified Analysis by Case Urgency: Elective vs Non-Elective Cases

| Model                                           | AUC (95% CI)        | Sensitivity | Specificity |
|-------------------------------------------------|---------------------|-------------|-------------|
| Preoperative intuition data alone               |                     |             |             |
| Elective case (n = 87)                          | 0.708 (0.490-0.926) | N/A         | N/A         |
| Urgent or emergent case (n = 129)               | 0.652 (0.564-0.739) | 0.576       | 0.727       |
| Preoperative intuition data and clinical data * |                     |             |             |
| Elective case (n = 87)                          | 0.885 (0.802-0.967) | 0.987       | 0.125       |
| Urgent or emergent case (n = 129)               | 0.815 (0.742-0.888) | 0.821       | 0.512       |

\* Subset of clinical data: age (continuous), sex (dichotomous), functional health status (ordinal), American Society of Anesthesiologists classification (ordinal), congestive heart failure within 30 days of operation (dichotomous), current smoker within 1 year (dichotomous) and history of severe COPD (dichotomous).

AUC, area under the receiver operating characteristic curve; N/A, not applicable