UTILITY OF THE ACS NSQIP SURGICAL RISK CALCULATOR TO ACCURATELY PREDICT POSTOPERATIVE OUTCOMES AFTER COLON RESECTION

by

Rihab Al Lawati

M.D., Sultan Qaboos University, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Surgery)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

July 2016

Abstract

BACKGROUND:

Predicting potential complications from surgery is a crucial step to aid decision to operate. The American College of Surgeons (ACS) initiated the National Surgical Quality Improvement Program (NSQIP) which collects and analyses patients' outcomes from surgery. ACS NSQIP developed a Surgical Risk Calculator (RC) to predict risks of postoperative complications. Aim of this study was to assess RC accuracy for predicting complications in patients undergoing colon resection.

METHODS:

Validation study with secondary use of administrative data conducted in a tertiary care center. Patients who received colorectal procedures in our Enhanced Recovery After Surgery (ERAS) program from November 2013 to December 2015 were enrolled. RC predictions were calculated and compared with observed NSQIP outcomes within 30 days follow-up. Observed versus predicted outcomes were compared. RC accuracy was assessed by graphical examination of the model calibration for outcomes that exceeded 50 events. Predicted versus observed length of stay (days mean±SD) was compared.

RESULTS:

A total of 368 patients were enrolled. RC predicted versus observed outcomes (n) were: serious complication 40.3 vs. 51; any complication 60.5 vs. 70; surgical site infection (SSI) 31.8 vs. 51; pneumonia 5.8 vs. 15; cardiac complication 16.5 vs. 9; urinary tract infection 9.8 vs. 11; venous thromboembolism 4.8 vs. 4; acute renal failure 16.6 vs. 5; return to operating room 14.6 vs. 6; death 4.2 vs. 2; Discharge to facility 20.2 vs. 12. Good calibration was observed for any complication and serious complications. SSI was underestimated but RC adjustment by surgeon improved SSI prediction. Length of stay was inaccurately predicted: 4.4±1.3 predicted versus 8.6±12.1 days observed ($p < 0.01$, Wilcoxon Rank Sum Test).

CONCLUSION:

Application of RC in our population closely predicts serious and any complication but less accurately predicts SSI unless adjusted by surgeon and inaccurately predicts length of hospital stay. All outcomes including the above require analysis of greater number of events to permit final conclusions on RC use.

Preface

- Chapter two – Validation study with secondary use of administrative data was conducted as collaboration between Vancouver General Hospital Department of Surgery the Vancouver Coastal Health Department of Clinical Quality and Safety for NSQIP. The supervising committee consisted of Dr. Garth Warnock (as the Principal Investigator) and Drs. Kelly Mayson and Penny Brasher as Co-investigators.

    - Under the supervision of the research committee, I was responsible of formulating the research objectives, study design, obtaining the ethical approval, data collection, data analysis and writing the manuscript.
    - Ms. Tracey Hong (Department of Clinical Quality and Safety) provided the study participants under ERAS protocol who underwent colorectal procedures.
    - Dr. Penny Brasher (Center for Clinical Epidemiology and Evaluation) provided assistance in the project design and data analysis.
    - Mr. Markus Zurberg (Department of Clinical Quality and Safety) provided an insight about the current situation and how NSQIP assists in quality assurance.
    - Ethical approval was obtained from the UBC Clinical Ethics Board (UBC CREB No. H16-00821).

Table of Contents

List of Tables

List of Figures

List of Abbreviations

VGH	Vancouver General Hospital
UBC	University of British Columbia
ACS NSQIP	American College of Surgeons - National Surgical Quality Improvement Program.
NSQIP	National Surgical Quality Improvement Program.
ERAS	Enhanced Recovery After Surgery.
M&M	Mortality and Morbidity
MRN	Medical Record Number
RC	Risk Calculator
UTI	Urinary Tract Infection
SSI	Surgical Site Infection
PE	Pulmonary Embolism
DVT	Deep Vein Thrombosis
VTE	Venous Thromboembolism
COPD	Chronic Obstructive Pulmonary Disease
DM	Diabetes Mellitus
HTN	Hypertension
CHF	Congestive Heart Failure
OR	Operation Room
CPT	Current Procedural Terminology
ASA	American Society of Anesthesiologists
Cm	Centimeters
Kg	Kilograms

Acknowledgments

“We always aim for the best science” The inspiring sentence I frequently heard from Dr. Warnock and which immensely encouraged me to always aim for the best quality outcomes throughout my project period.

No words can express my deep and genuine gratitude to my supervisor, Dr. Warnock, for his sincere dedication, keen interest and overwhelming support throughout my master’s journey.
His constructive advice, guidance and kind co-operation enabled me to complete and accomplish this task.

I am also thankful to my committee members Dr. Kelly Mayson for her input and guidance, Dr. Penny Brasher for her opinions in the statistical analysis plan and content revision, Ms. Tracey Hong for her constant cooperation during the data collection phase, Mr. Markus Zurberg for his valuable opinions during the period of finalizing the study design.

My special Thanks go to Dr. Alice Mui (MSc program coordinator) who always was there to facilitate any obstacles appeared through the way and Dr. Morad Hameed for accepting to be part of the defense committee.

Finally, I would be honored to express my sincere appreciation and gratitude to Dr. Hani Al Qadhi for his ongoing support, keen encouragement and for making this opportunity of pursuing my ambitions possible.

Dedication

I dedicate this thesis to:

My perfect match, Sadiq for his continuous support and encouragement, without him this would not have been possible.

My loving parents, who raised me to the person I am today and who taught me that no matter how hard the journey might be, ambitions must be chased and goals must be reached.

My wonderful three sisters, Noor, Bedoor and Lubna, with whom I shared my day-to-day details and difficulties until this work was finally formulated. Despite the distance their presence will always be of a great value to me.

My only best friend, Safiya, without hearing her advises and weekly complaints my journey would not have been as joyful as it was.

My son, Ali, Who I cannot wait to see him grow up into a successful gentleman with enough confidence, knowledge and wisdom to choose the right path in life.

Chapter 1 Introduction And Literature Survey

1.1 Background

Complications after surgery increase mortality substantially (1). Colorectal surgery itself carries a significant mortality risk, with reported rates of 1–6% for elective surgery and up to 22% in the emergency setting (2-5).Understanding the risks and possible complications of any surgical intervention before proceeding with it is an important issue. The process of this decision is shared between physicians and patients. In order to reach a final decision and provide the informed consent that is required before any intervention, the patient has to have a thorough understanding of the potential risks of surgery (6,7).

Providing accurate information to patients about potential complications is essential. Historically, health institutions adopted what is called a minimum standard program in which every staff is required to review and analyze at regular intervals their clinical experience and discuss it in regular meetings (8,9). With some modifications over the years, this experience nowadays is called morbidity and mortality meeting or morbidity and mortality conference.

1.1.1 Mortality and morbidity conference

The morbidity and mortality conference has a long history in the academic pathway as many medical centers use it as a teaching tool to provide trainees real examples of medical errors or problematic decision-making situations. Although this type of meeting has not always resulted in lessons to prevent future error, it has been increasingly used as a part of the measures that are taken to enhance patient safety and quality of care (10).

Currently, this type of meeting is considered a key educational tool which is essential for training programs accreditation and a critical contribution to quality assurance (11). From these meetings, critical information can be gathered to inform knowledge about certain cases and complications.

As part of the educational process, trainees must learn to report errors and quality issues, to participate in the wider picture of providing a better environment for patient safety. Trainees should also be engaged in projects to improve systems of care, decrease health care disparities and improve patient outcomes (12). If this process is conducted with attention to best practices such as non-punitive review, debriefing, and follow up on systems improvements it can support building strong safety cultures in medicine (13).

1.1.2 UBC Mortality and Morbidity Tool

To be contemporary with the recommendations and apply the best quality measures, UBC has established Mortality and Morbidity (M&M) Tool in 2013. Many modifications were introduced to the tool over the previous years to improve its quality. Appendix A demonstrates the website page of the tool and the categories that must be completed in order to report a case of mortality or morbidity, for example, patients’ last name, MRN, type of complication and Clavien Dindo grade. Recently the tool was approved by British Columbia Freedom of Information and Protection of Privacy Act and The Vancouver Costal Health Authority. Since August 2015, the tool has been used for residents in the UBC General Surgery Program to report complications for presentations and discussions at a weekly M&M conference. Furthermore, a database was established and a summary provided for feedback to surgeon attending faculty on the frequency of complications on their teams. Surgical team members are requested to provide feedback to front line care providers on their team-encountered patients’ mortality and morbidity. This experience was prepared as an abstract submitted for presentation in one of the future NSQIP meetings (personal communication; Mr. Markus Zurberg, Dec 2015).

Despite potential advantages of M&M reporting tool such as the one we have developed, literature is showing a very low reporting rate for custom reporting tools in other centers (9) which made it necessary to find other ways to improve patient safety by creating a variety of pre-procedural risk assessment tools.

1.1.3 Risk assessment tools/models

Many institutions created tools for risk assessment and risk prediction to help in the process of decision making. Some of the currently available models in the literature include Cleveland Clinic Foundation colorectal cancer model and the Physiological and Operative Severity Score for enUmeration of Mortality and morbidity (POSSUM). These models either solely assess risk of mortality or are difficult and complex to assess at patient's bedside (14). Other models which tried to simplify the previously mentioned ones include the Colorectal preOperative Surgical Score (CrOSS), which may be easier to apply but still encounters the same problem of assessing the risk of mortality only without addressing the other important aspects of postoperative morbidity (15). In order to solve this issue, ACS NSQIP developed a model called the Surgical Risk Calculator (RC) which addresses postoperative morbidity and mortality.

1.1.4 National Surgical Quality Improvement Program (NSQIP)

The American College of Surgeons (ACS) initiated a program called National Surgical Quality Improvement Program (NSQIP) that collects high-quality, standardized clinical data on preoperative risk factors and postoperative complications from patients who have surgery in more than 500 hospitals in North America and selected international sites (16,17). Clinical reviewers in these centers are extensively trained to collect data in different methods like chart review, surgeon and patient interview to ensure high quality of the collected data (18). This program applied high-quality data that they collected to develop a tool to predict surgical risk in the form of software that predicts postoperative mortality and morbidity and from here it gains its importance, as it is not addressing mortality only like the previous models but it also includes morbidity assessment (17).

1.1.5 ACS NSQIP Surgical Risk Calculator

The ACS NSQIP Surgical Risk Calculator is generated from 1.4 million patients' information gathered between 2009 till 2012 from all the NSQIP-participating institutions (19,29). Those institutions ranged from rural community hospitals to large academic and university-affiliated centers representing a wide variety of surgeries in variable clinical settings (29).

The initial RC first released in 2013 (29) was an online software requesting information input of demographics, functional status, comorbidities and American Society of Anesthesiologists (ASA) class. Figure 1, Appendix B shows the risk calculator software interface and a screen shot of the risk factor entry screen. This data is entered by the surgeon or anesthesiologist who is assessing the patient in the preoperative period. All the risk factors can be entered using a drop down menu of each category which makes the process of entering the data easier and quicker. The information required to generate the risk estimates includes procedure name, patient age, sex, comorbidities (DM, HTN, cardiac events, etc.), ASA class, wound class and others. Figure 2 (appendix B) is the risk generated report screen which demonstrates the way that risk estimates are presented in each category starting from serious complication and ending with discharge to nursing or rehabilitation facility. Clear definitions are provided by NSQIP explaining what each category of outcomes indicates and the outcome definition will appear in a pop up dialogue if clicked on the question mark sign beside each outcome. For example, "any complication" defined as all the NSQIP recorded morbidity and they are: Superficial incisional SSI, deep incisional SSI, organ space SSI, pneumonia, unplanned intubation, PE, DVT, ventilator > 48 hours, acute renal failure, UTI, cardiac arrest, myocardial infarction, return to the operating room, systemic sepsis. While Serious complication included all the outcomes mentioned in any complication except superficial incisional SSI and ventilator > 48 hours. The report also provides an explanation on how to interpret the results by showing a sample at the bottom of the reporting page. The sample shows that the bolded black line represents average patient risk and the concerned patient risk is demonstrated in three methods: graphically, percentage of estimated risk and chance of outcome (below average, average, above average). The report also provides an estimation of the length of hospital stay postoperatively. Finally, the calculator gives the surgeon or physician assessing the patient the option of adjusting the risks because the calculator doesn't capture all risk factors. The risks adjustment has three levels, level 0 with no adjustment, level one is when the risk is somewhat higher than estimated and level two is when the risk is significantly higher than estimated.

So far, the literature presents scattered papers studying the accuracy of ACS NSQIP Surgical Risk Calculator in estimating postoperative complications in certain types of surgeries. Bilimoria et al concluded in their study that ACS NSQIP Surgical Risk Calculator level of prediction was reasonable and this was demonstrated by c-statistics results which ranged between 0.806 to 0.944 for most of the outcomes (20). Some studies showed that complications might not be accurately estimated (21). Other studies showed that complications were effectively estimated in patients with average risk factors but less so in predicting complications in patients with lots of risk factors (22). These studies were performed in surgical specialties of gynecology, orthopedics and surgical oncology. One study compared the data collected using a traditional M&M tool with data collected using the ACS NSQIP techniques concluding that the M&M tool considerably underreported for both in-hospital and post-discharge complications and deaths compared with ACS NSQIP techniques (9). No study in the literature was identified that addresses utility of the

RC in the context of general surgery cases rather than subspecialized fields like surgical oncology, gynecology and orthopedics surgery.

Recently, ACS NSQIP Surgical Risk Calculator was updated. New prediction equations were based on larger and more recent samples of surgical patients. The updated calculator tends to assign, for highest risk patients, higher predicted risk of mortality than the old calculator. Otherwise, the prediction for other outcomes is almost similar to the version used in this study (personal communication with Tracey Hong, May 2016).

Hence, the idea of this project came to study the accuracy of ACS NSQIP Surgical Risk Calculator in the setting of elective colorectal surgery procedures which are enrolled in ERAS program. ERAS database combined with standardized NSQIP reporting allows for reliable definitions to test the RC tool.

1.1.6 Enhanced Recovery After Surgery (ERAS)

The project aims to use the Enhanced Recovery After Surgery (ERAS) database as a source for the sample population that will be studied. ERAS program (also called fast track perioperative care) is an evidence-based collection of protocols that patients undergoing elective surgeries are recommended to follow (23,24).
Literature shows that undergoing colorectal surgery involving bowel resection carries s 15% to 20% rate of complications (25,26). In an effort to reduce post colorectal surgery complications and decrease the length of hospital stay, Kehlet et al. (27) was the first to describe in detail the fast track or the enhanced recovery after surgery protocols. This was achieved mainly by harnessing the physiological principles to improve patient outcomes by reducing the profound stress response induced by surgery, there by reducing postoperative complications, minimizing hospital stay, and ultimately reducing health costs without compromising patients' safety (28,30). Moreover, the aim is to provide pain and stress-free pathway to full recovery.

Combining all the above factors, the main objective of this study is to examine the evidence that supports and validates the accuracy of the level of risk prediction generated by the RC in our patient population and how accurately it may help in predicting postoperative complications. The ultimate aim is to observe if pre-surgical application of the RC will highlight ways to predict and reduce postoperative complications further in patients entering ERAS programs (30).

1.2 Overall project hypothesis:

We hypothesize that the RC predicts postoperative complications that can be detected through routine NSQIP screening.

1.3 Overall project objective:

The overall objective of this thesis is to provide a more robust ability to predict and prevent postoperative complications in a population of patients who are undergoing scheduled elective colon resections using a standardized procedure care protocol.

Chapter 2 Utility Of The ACS NSQIP Surgical Risk Calculator To Accurately Predict Postoperative Outcomes After Colon Resection: A Validation Study.

2.1 Methods

2.1.1 Study design

This project is a validation study with secondary use of administrative data. This study compared observed postoperative outcomes for patients undergoing elective colorectal procedures in VGH to the predicted outcomes generated by the RC. This comparison was conducted to assess validity of the surgical risk prediction model provided by NSQIP in our patient population. Data for a cohort of patients who underwent elective colorectal surgery under ERAS protocol at VGH during the period from November 2013 to December 2015, and who were selected for the VGH NSQIP sample was extracted from the ERAS database.

2.1.2 Patient selection (inclusion, exclusion criteria)

All adult (≥18 years old) patients of all ages who underwent colorectal procedure under Enhanced Recovery After Surgery program in Vancouver General Hospital were included.
Any patients presented and operated in an emergency setting were excluded from the study.

2.1.3 Data collection and definitions

Consecutively treated patients enrolled in ERAS who underwent elective colorectal procedures between the periods from November 2013 to December 2015 were identified through reviewing the lists of OR slates for patients who underwent colorectal procedure under ERAS during the above mentioned period. Information on patients' demographics, functional status, smoking, medical background (HTN, DM, COPD, previous cardiac history, ventilation dependence, cancer, acute renal failure, dialysis, ascites and sepsis) and procedural details (procedure name, date, CPT code and description) was abstracted from the ERAS database and entered into Excel spreadsheets as summarized in table 2-1.

Admission, discharge dates and discharge destination were also obtained to calculate the length of hospital stay. Duration of hospital length of stay was defined as the total days postoperatively in hospital from the surgery date until discharge. NSQIP outcomes detected through the 30 days follow up included any complication, serious complication, pneumonia, cardiac complications, surgical site infection, urinary tract infection, acute renal failure, ileus, deep vein thrombosis, pulmonary embolism, unplanned intubation, ventilation more than 48 hours, death and discharge destination as shown in table 2-1.

All categories were defined by NSQIP standard definitions. The category "any complication" included an aggregate of all the NSQIP-recorded morbidities including: Superficial incisional SSI, deep incisional SSI, organ space SSI, pneumonia, unplanned intubation, PE, DVT, ventilator > 48 hours, acute renal failure, UTI, cardiac arrest, myocardial infarction, return to the operating room, systemic sepsis. "Serious complication" was defined as all the outcomes mentioned in any complication except superficial incisional SSI and ventilator > 48 hours.

Table 2-1 Data collection form variables

Preoperative variables (ERAS database)	30 days follow up variables (NSQIP database)
Age	Pneumonia
Sex	Cardiac complications
Date of admission	SSI
Date of discharge	UTI
Date of surgery	Ileus
Procedure name	DVT
CPT code and description	PE
Height in (cm) & weight (kg)	Acute renal failure
DM	Transfusion
HTN	Sepsis
Smoking	Return to OR
Dyspnea	Death
Ventilation dependence	Unplanned intubation
Disseminated cancer	Ventilation >48 hours
Functional status	Discharge destination
COPD	
Previous cardiac history/ CHF	
Acute renal failure/ dialysis	
Steroid use	
Ascites	
Systematic sepsis	
ASA class	
Wound class	

If study data was missing from the ERAS or NSQIP database the information was obtained by chart reviews of preoperative assessment reports in the Vancouver General Hospital Patient Care Information System (PCIS) files. Since all of the patients were managed according to the standard of care at VGH, the project was approved to be of minimal risk to patient confidentiality by the University of British Columbia Clinical Research Ethics Board which approved the study protocol to be conducted with a waived consent (UBC CREB No. H16-00821).

2.1.4 Data analysis

Descriptive statistics were used to describe the demographic, preoperative and operative characteristics of the study population. Discrete variables were summarized by frequencies and percentages. Continuous variables were summarized by mean (standard deviation).

A table of the number (percent) of patients with a NSQIP-recorded complication was constructed. Postoperative outcomes were broken down by NSQIP category, i.e., serious complication, any complication, pneumonia, cardiac complication, SSI, UTI, VTE, acute renal failure, return to OR, death and discharge to nursing or rehab facility.

The NSQIP predicted and the observed lengths of hospital stay postoperatively were compared visually using histograms. A table showing the mean, standard deviation, range of hospital stay was constructed and the Wilcoxon rank sum test was used to compare the two groups.

Evaluation of the accuracy of the ACS NSQIP Surgical Risk Calculator:

To assess the model performance, calibration measures were assessed for the outcome categories which had at least 50 events.
Calibration of a prediction model generally studies the agreement between observed outcome frequencies and predicted probabilities (31).

In our study the calibration of the model was assessed graphically by comparing predicted and observed risks for the categories serious complication, any complication and SSI. The predicted risk for each outcome was cut into approximately equally-sized "bins".

A bin, is away of sorting data by evenly distributing the data set in carefully chosen categories, as demonstrated in figure 2-1 (x-axis).

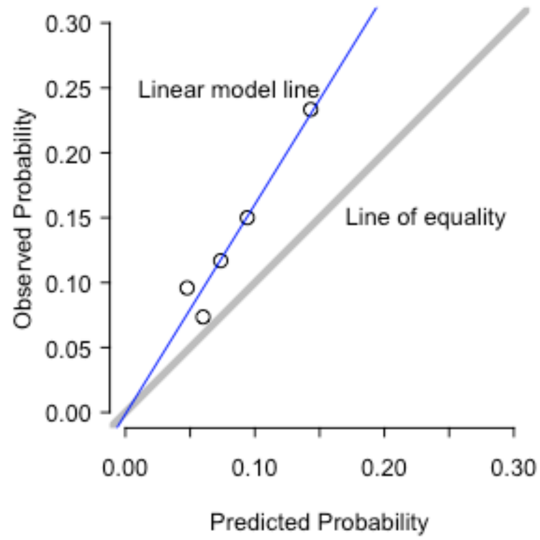


Figure 2-1 Calibration graph example: The thick line indicates perfect calibration; the thin line shows the relationship between the predicted and observed probabilities. Circles indicate observed events per quintile of predicted probabilities

After that, the mean predicted risk for each bin was calculated (x-axis) and plotted versus the observed risk (y-axis) (Figure 2-1) and a linear model line was fit to the data (thin line). Finally the line of equality was superimposed on the graph (Thick line).

The line of equality is used as a reference in comparing two sets of data expected to be identical. If the prediction model is perfectly accurate, the fitted linear model line should follow the line of equality. In the example provided in figure 2-1 the linear model line is shifted to the left of the line of equality, which indicates that the observed events are higher than predicted. For outcomes that did not have a sufficient number of events to assess calibration, we provided a table of observed versus predicted events.

For length of postoperative hospital stay, a p value <0.05 was considered statistically significant. All data was collected using Microsoft Excel (2011) and data analysis was carried out using R program (version 3.2.3).

2.2 Results

2.2.1 Baseline characteristics of the study population

Figure 2-2 shows the selection of patients to be included. A total of 416 patients were enrolled in ERAS program to undergo a colorectal procedure during the period from Nov 1st, 2013 to December 31st, 2015. Study population colorectal procedures included left and right hemicolectomy, sigmoid and segmental colon resection, rectosigmoid and rectal excision. Diagnoses of colon resection were malignancy, benign polyp, diverticular disease and inflammatory bowel disease. Out of these, thirty-seven were excluded because they were not sampled by NSQIP and another eleven patients were excluded because of missing data.

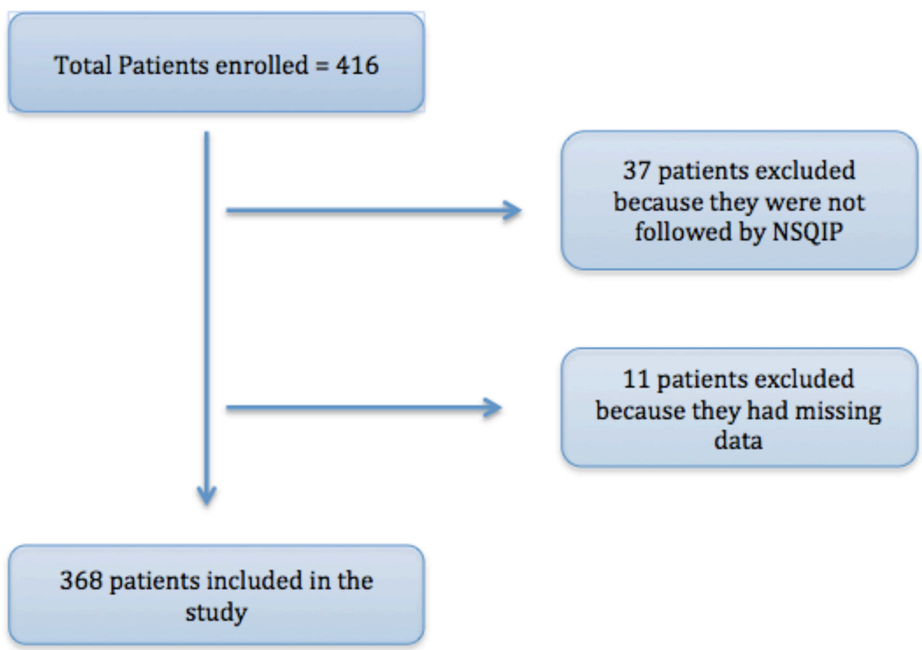


Figure 2-2 Flow diagram of patients through the study

Finally, a total of 368 patients were included in the study. Table 2-2 shows the demographics, procedure measures and operative characteristics of the population. The age of the study population ranged between 24 and 100 years and the mean was 69 years with a standard deviation of 13.4 and 54.1 % of them were males. The mean population BMI was 26.7 (SD= 5.3) and the preoperative measures showed that 45.4% of the sample were hypertensive and 15.8 % were with a previous cardiac history.57.3 % were classified as ASA class 2 and 71.2 % of the total procedures performed were minimally invasive surgery.

Table 2-2 Characteristics of the study population and procedures (n=368)

Demographics	Mean±SD or Number	%
Age (years)	69±13.4	—
Males – number (%)	199	54.1
Body Mass Index (kg/m^2)	26.7±5.3	—
- Height (cm)	163.5±18.8	—
- Weight (kg)	74.1±18.5	—
Preoperative measures	**Number**	**%**
Functional status		
- Independent	365	99.2
- Partially Dependent	3	0.8
Diabetes*	37	10.1
Hypertension	167	45.4
Smoker**	29	7.9
Dyspnea on moderate exertion	18	4.9
Ventilator dependent	5	1.6
Disseminated Cancer	17	4.6
COPD†	14	3.8
Previous cardiac history	58	15.8
Dialysis	3	0.8
Recent steroids use	15	4.1
Sepsis††	5	1.4
Operative characteristics	**Number**	**%**
ASA Class ∞		
- Class 1	15	4.1
- Class 2	211	57.3
- Class 3	128	34.8
- Class 4	14	3.8
Procedure		
- Laparoscopic technique	262	71.2
- Open technique	105	28.5
Wound Class		
- Clean-Contaminated	350	95.1
- Contaminated	10	2.7
- Dirty-Infected	8	2.2

Data are shown as number (%) or mean (standard deviation). •Diabetes category including Insulin and non-insulin dependent; ••smoking recorded within a year from the surgery; †COPD, Chronic Obstructive Pulmonary Disease; ††Sepsis includes its different stages; ∞ ASA, American Society of Anesthesiologists.

2.2.2 Frequency of complications

The occurrence of a complication was recorded by NSQIP is shown in table 2-3. There were a total of 70 patients with recorded NSQIP morbidity, representing 19 % of the total sample population. Fifty-one of the total complications met the criteria of serious complication representing 13.9% of the total population of the study. Surgical site infection was the highest represented complication against all other recorded outcomes reaching up to 13.8 % of the total complications. Postoperative blood transfusion and ileus scored 10.3 % and 8.7 % respectively but these two complications are not predicted by the tool that we are evaluating in this study so they were not further included in the analysis.

Table 2-3 Frequency of complications recorded in NSQIP database

NSQIP-recorded complication	Number	Percent %
Serious complication	51	13.9
Any complication	70	19
Pneumonia	15	4.1
Cardiac complication	9	2.4
SSI	51	13.9
UTI	11	3
DVT	3	0.8
Ileus	32	8.7
Acute renal failure	5	1.4
PE	1	0.3
Transfusion	38	10.3
Sepsis	20	5.4
Return to OR	6	1.6
Death	2	0.5
Intubation	10	2.7
Ventilation >48 hours	11	3

2.2.3 Length of postoperative hospital stay

Length of postoperative hospital stay is shown in table 2-4. The mean of the predicted length of stay was 4.4±1.3 days (range 2.5-10.5, median 4). Corresponding stay was 8.6±12.1 days (range 1-171, median 6). The difference between predicted and observed lengths was significant ($p < 0.01$, Wilcoxon rank sum test).

Table 2-4 Length of postoperative hospital stay comparison

Length of postoperative hospital stay (days)			
	Predicted	**Observed**	**P value** *
Mean (days)	4.4	8.6	< 0.01
SD	1.3	12.1	
Range of hospital stay (days)	2.5 - 10.5	1 - 171	

* Wilcoxon rank sum test

Figure 2-3 is shows the distribution of the predicted length of postoperative hospital stay versus the observed length of postoperative hospital stay. It clearly illustrates a wider pattern of distribution. Multiple outliers are shown in the graph of observed length of stay.

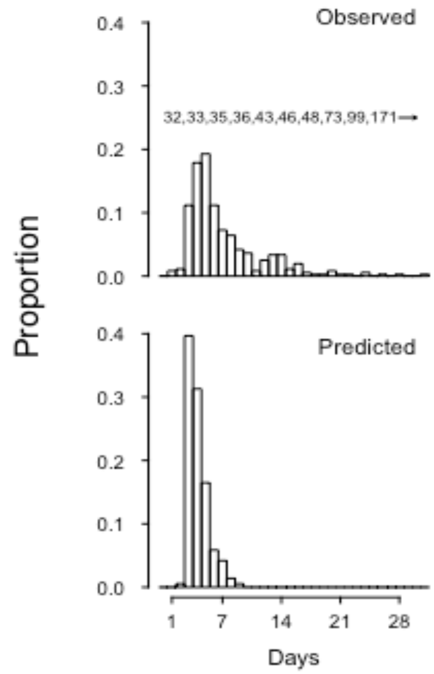


Figure 2-3 Distribution of predicted vs. observed length of hospital stay
(The numerical values in the observed graph are the outliers detected in that group)

2.2.4 Evaluation of the accuracy of ACS NSQIP Surgical Risk Calculator

In this part of the analysis the calibration of the risk prediction model was assessed for the following outcomes:

1. Surgical site infection:

Surgical site infection was the highest detected complication as shown previously in table 2-3 and the pattern of its predicted distribution is shown in figure 2-4. The calibration of the model for this outcome was assessed without increasing the level of surgeon adjusted risk in the RC in figure 2-5 and showed that the observed outcomes are higher than the predicted outcomes, which means that this model is underestimating the risk of surgical site infection.

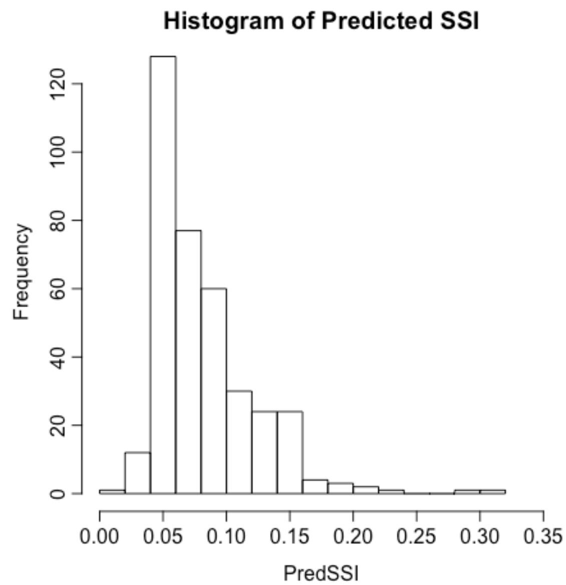


Figure 2-4 Distribution of the predicted SSI

To further assess the model, the same process was repeated for but with introduction of the RC surgeon adjusted risk according to the RC report screen. As shown in figure 2-6 and figure 2-7 respectively, predicted outcomes for SSI were much more accurate and closer to the observed outcomes in level 1 adjustment but in level 2 the predicted outcomes exceeded the observed outcomes. From the assessment of the model calibration taking in consideration its three levels (without adjustment, adjustment level 1, adjustment level 2), prediction with risk adjustment level one provided the closest prediction of postoperative surgical site infection in our patient population.

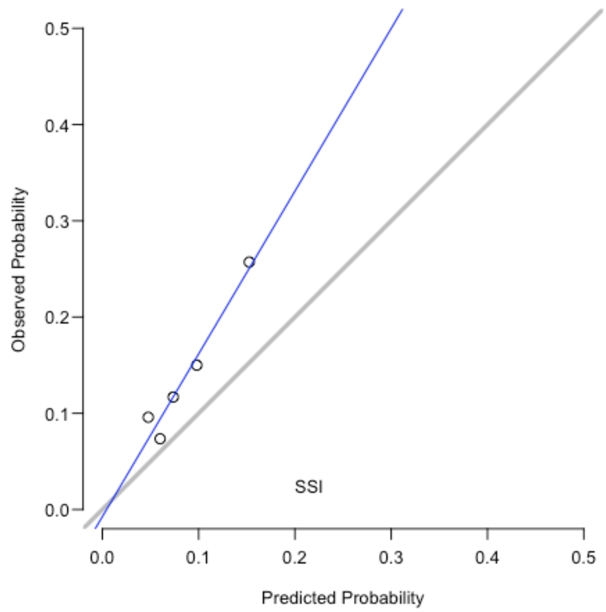


Figure 2-5 Comparison of predicted vs. observed SSI without risk adjustment

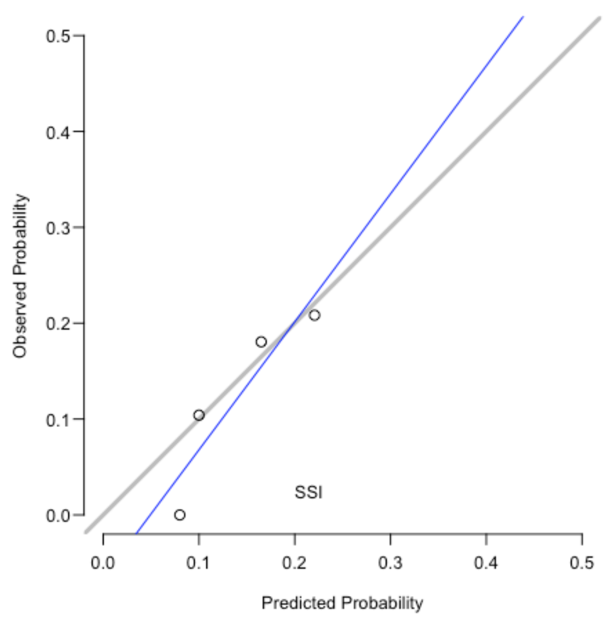


Figure 2-6 Comparison of predicted vs. observed SSI with risk adjustment level 1

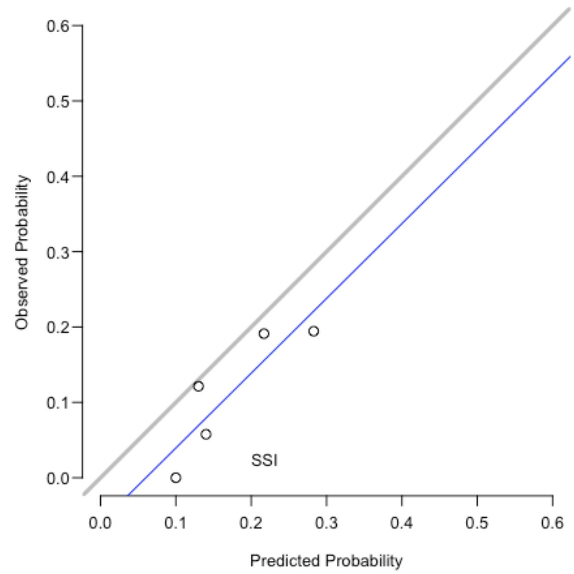


Figure 2-7 Comparison of predicted vs. observed SSI with risk adjustment level 2

2. Any complication:

This category of outcomes included the sum of all the predicted morbidity outcomes by NSQIP and figure 2-8 is showing the pattern of its distribution.

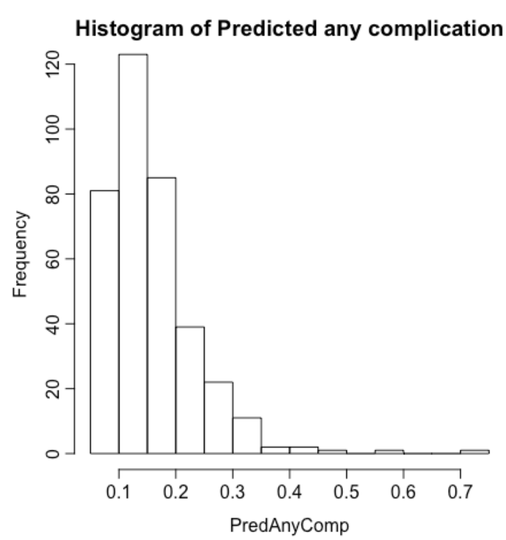


Figure 2-8 Distribution of predicted any complication

Figure 2-9 and 2-10 are the graphs of the model calibration for outcome category "any complication" without risk adjustment and with risk adjustment level 1 respectively. For this outcome, the model showed a similar level of prediction without risk adjustment, but a lower agreement of prediction at higher adjustment.

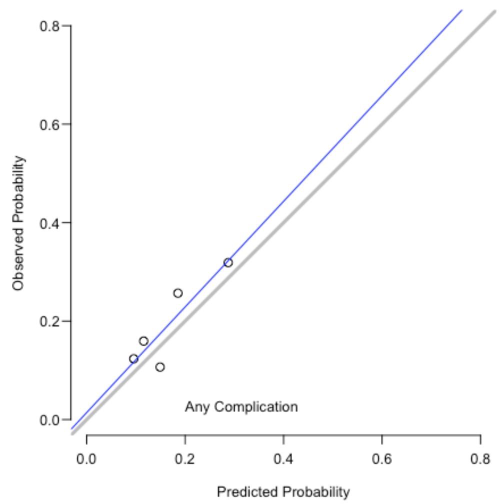


Figure 2-9 Comparison of predicted vs. observed any complication without risk adjustment

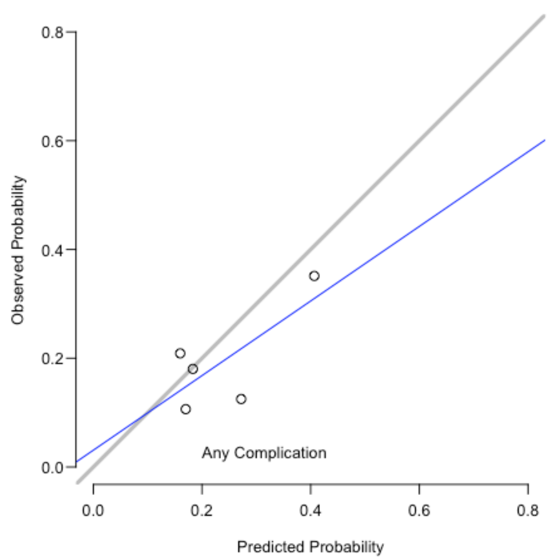


Figure 2-10 Comparison of predicted vs. observed any complication with risk adjustment level 1

3. Serious complication:

This category of outcomes shows the distribution in figure 2-11. The calibration graph showed that observed risks were slightly higher than predicted (figure 2-12) but using the level 1 adjustment produced larger discrepancies (figure 2-13).

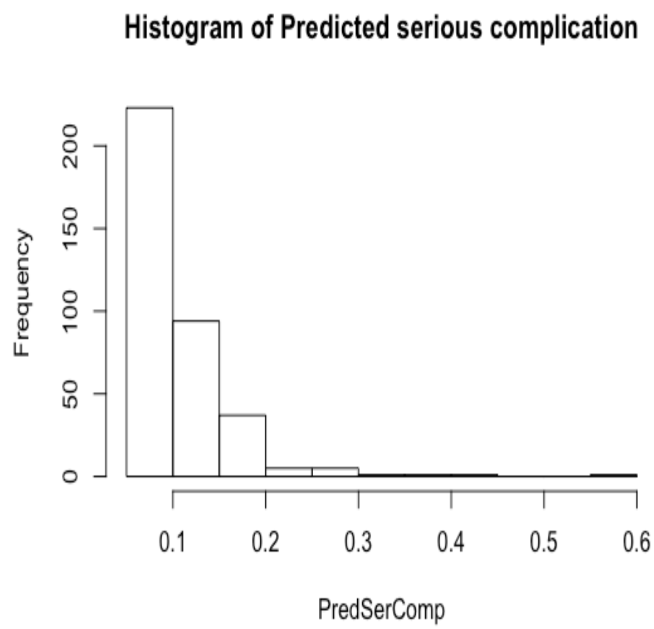


Figure 2-11 Distribution of predicted serious complication

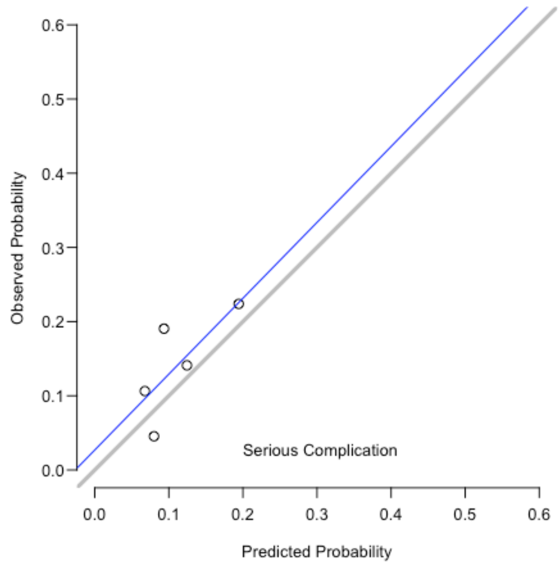


Figure 2-12 Comparison of predicted vs. observed serious complication without risk adjustment

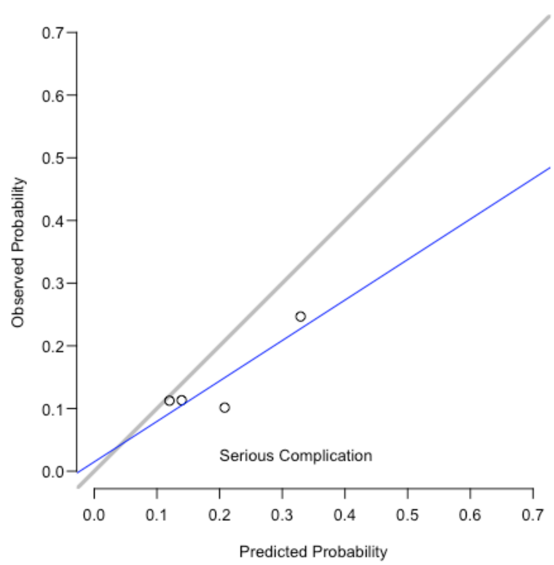


Figure 2-13 Comparison of predicted vs. observed serious complication with risk adjustment level 1

2.2.5 Less frequent outcomes

The rest of the outcome categories were not assessed graphically because the number of the detected events per category was very small. Pneumonia and UTI were the only two categories that showed a larger number of observed complications than the predicted ones. The outcome category discharge to nursing or rehab facility showed less number of observed events than predicted and this was the same for all other complications.

Table 2-5 Total predicted vs. total observed outcomes

Outcome category	Total predicted	Total observed
Pneumonia	5.8	15
Cardiac complications	16.5	9
UTI	9.8	11
VTE	4.8	4
Acute renal failure	16.6	5
OR return	14.6	6
Death	4.2	2
Discharge to facility	20.2	12

2.3 Discussion

Validation studies are crucial to evaluate any prediction model performance. Two qualities are assessed, including calibration (compare observed and predicted event rates for a group of patients) and, discrimination (quantify the model ability to distinguish between patients who do or do not experience the event of interest) or both (36). Testing the generalizability of these models before recommending them for clinical use is essential (32). The more diverse the setting the model is tested and found accurate, the more it will generalize and become widely applicable (33, 34).

In any surgical field, surgeons are frequently asked to provide prognostic assessment

for operative and postoperative risks (39). To do so, surgeons usually depend on the literature that is based on results of aggregate data or they depend on their personal experience and they often worry that their assessment will prove incorrect (33,35).
The American College of Surgeons provided a tool that may guide surgeons to accurately predict postoperative risk and make the process of assessment consistent in different surgical environments worldwide. The ACS NSQIP Surgical Risk Calculator is a model that was built on a multi-institutional, well-collected data of a large sample size. However, this tool is still underutilized in the clinical field. One of the possible reasons for this is that this tool remains incompletely tested to prove its validity and gain trust of surgeons and stakeholders to incorporate it in the preoperative assessment protocols.

2.3.1 Main findings

Our study applied this tool on a sample of patients from a single tertiary center and compared the predictions to our actual experience. Starting with the length of postoperative hospital stay, the observed stay showed multiple outliers. Even after we tried to exclude the outliers, the observed distribution was still different than the predicted one. The RC did not show an accurate prediction, as the range of the prediction was very small despite having multiple patients with many risk factors. This may indicate that the tool incompletely adjusts the risk on the individual level. This observation agrees with other studies in the literature which had similar findings (22). Our study population was ERAS which has designed to reduce postoperative length of hospital stay, yet we still did not see accuracy.

To test the model performance our study assessed the calibration graphically. Three outcome categories of any complication, serious complication and SSI were chosen to be assessed by this method and the reason for choosing them was that they contained a reasonable number of events. In order to rigorously assess calibration at least 100 events are necessary (31).

Despite the fact that our sample population is on standardized procedure care protocols within ERAS under high compliance with perioperative protocols to reduce SSI, we still found that SSI scored the highest. The model prediction for this outcome showed an overall underestimation of the postoperative SSI. In order to improve the model prediction, we generated the calibration graphs for the predicted outcomes with risk adjustment in its two levels. This step was done to examine if simple risk adjustment will modulate the model performance. This particular step adds to the concepts in previous literature. Our analysis for this outcome showed by adjusting the risk to level one, the calibration graph demonstrated a better agreement between the predicted and the observed outcomes. Hence, we can suggest that to improve the tool performance for this outcome in our patient population, the first level risk adjustment can be employed.

The same steps of testing the agreement between the predicted and observed outcomes were performed for the outcome categories any complication and serious complication. For these two variables, the model showed fairly good calibration.

Our study did not perform the same steps of assessment and analysis for the rest of the outcome categories (Pneumonia, Cardiac complications, UTI, VTE, acute renal failure, OR return, death) because the number of events detected was too small to draw conclusions from it. A comparison between the sums of observed outcomes and the sums of the detected outcomes were performed and it showed that VTE predicted and observed events were similar, while Pneumonia and UTI showed higher observed events and all other outcome categories showed a lower number of observed events than predicted by the model. However, these findings must be interpreted cautiously given the small number of events.

We observed a total of 19% of overall complications in our sample population. This percent is still considered high and greater efforts are required to address this problem in order to find new methods for improvement. It is possible the prospective application of the RC in advance of surgery for this population could allow better preoperative preparation and enhance partnerships between patients, surgeons, anesthesiologists and nurses, but this possibility remains unproven. As ERAS standardized protocols improve in this population, the overall complication rate is trending down in slow steps and this was noticed in comparison to a recent study done in the same center with similar characteristics of the sample population which showed an overall complication rate of 22%. SSI remained the outcome of concern with a rate of 14% in both studies (30, personal communication with Dr. Garth Warnock, March 2016).

Accurate assessment of treatment risks is an important aid for good decision-making and a recent study published in the literature found that providing surgeons with objective data from a well-validated risk calculator resulted in improved and less varied judgments of operative risks that more closely approximate the risk calculator values (38). Due to this fact, our study tried to focus on one of the tools that are currently available and proof its validation or suggest a simple "fix" to improve the model and make the tool better utilized clinically.

The current study aims to determine if the RC is accurate in a colon resection population. This is unique in the literature and essential step to incorporating this tool in preoperative clinics. The calculator can be completed by surgeon or anesthesiologists in the preoperative assessment period where they can have the chance to discuss the results with patients to make a shared decision about the best quality of care provided focusing on individual patient risk factors and needs.

2.3.2 Limitations

There were several limitations in this study that were recognized. First, the reliance on historical data which exposed the study to the issue of missing data and followed by patients exclusion due to this fact. Secondly, the sample size and we were unable to rigorously assess calibration, thus our findings should be interpreted cautiously. Third, our sample population was based on patients treated in a single tertiary referral center. Therefore, the results are unlikely to be generalizable to other centers. Despite the limitations, our study has made unique contribution to the literature by assessing the calibration of the RC prediction model, including prediction among the three levels of risk adjustment. This allowed us to test if the model can be improved by simply raising the level of risk adjustment instead of just stating that this model is not appropriately predicting postoperative outcomes.

2.4 Conclusion

Prediction models for treatment risks are of crucial benefits for surgeons to predict their decision to operate. They can also allow patients to understand and comprehend the measures of the possible risks associated with their treatment in order to reach a final decision and sign the informed consent.

For a risk prediction model implementation in a clinical system, it must have a rigorous proof of its validity and generalizability. Our study focused on the ACS NSQIP Surgical Risk Calculator, tool generated from a very well collected data and vast sample size. Calibration of the model was mainly examined in three categories of NSQIP outcomes for any complication, serious complication and SSI categories which showed: any complication and serious complication predictions were fairly accurate but SSI prediction was lower than the actual. By adjusting the level of risk, model predictions improved notably. Length of postoperative hospital stay was examined too but the model showed inaccurate predictions for this variable.
Operationally, results from the study to date would mean using the values predicted by the RC for all outcomes including those less frequent, because we have no evidence to support not using them.

We conclude that this tool is a potentially useful tool in this population of colon resection patients but we recommend further analysis to be conducted to test all the ACS NSQIP Surgical Risk Calculator predicted outcomes after obtaining larger number of events to properly validate the model and reach a solid conclusion about its level of accuracy as planned for the ongoing part of this study.

2.5 Future plan

All that have been done so far is considered as phase I of the project that we proposed and it is the initial step to move forward towards the ultimate goal that we have set and trying to achieve. Further data will be collected for validating the tool in the study population giving our center continues to collect NSQIP data on all colon resection patients who enter ERAS program. Also continuing the project will give us the privilege of increasing the sample size, which will increase the overall number of events and the number of events per category of complication. The next step will be moving to phase II where we will be collecting the data prospectively to avoid the problem of missing data when depending on hospital records alone. This will allow us to assess the model calibration for more outcome categories and perhaps recommend it as a useful preoperative tool placed on the chart for patients, surgeons, nurses and anesthesiologist to incorporate into the preadmission clinic protocols or define a certain recommendation to prepare patients for a safe surgical experience.

Bibliography

1. Sandblom G, Videhult P, Crona Guterstam Y, Svenner A, Sadr Azodi O. Mortality after a cholecystectomy: a population-based study. HPB. 2015 Mar;17(3):239–43.

2. Population-based audit of colorectal cancer management in two. 2005 Nov 10;:1–6.

3. Operative mortality in colorectal cancer: prospective national study. 2003 Nov 18;:1–6.

4. Slim K, Panis Y, Alves A, Kwiatkowski F, Mathieu P, Mantion G. Predicting Postoperative Mortality in Patients Undergoing Colorectal Surgery. World J Surg. 2005 Dec 13;30(1):100–6.

5. Ragg JL, Watters DA, Guest GD. Preoperative Risk Stratification for Mortality and Major Morbidity in Major Colorectal Surgery. Diseases of the Colon & Rectum. 2009 Jul;52(7):1296–303.

6. Knops AM, Legemate DA, Goossens A, Bossuyt PMM, Ubbink DT. Decision Aids for Patients Facing a Surgical Treatment Decision. Annals of Surgery. 2013 May;257(5):860–6.

7. Schenker Y, Fernandez A, Sudore R, Schillinger D. Interventions to Improve Patient Comprehension in Informed Consent for Medical and Surgical Procedures: A Systematic Review. Medical Decision Making. 2011 Feb 4;31(1):151–73.

8. Brand RA. Biographical Sketch: Ernest Amory Codman, MD (1869–1940). Clinical Orthopaedics and Related Research®. 2012 Dec 18;471(6):1775–7.

9. Hutter MM, Rowell KS, Devaney LA, Sokal SM, Warshaw AL, Abbott WM, et al. Identification of Surgical Complications and Deaths: An Assessment of the Traditional Surgical Morbidity and Mortality Conference Compared with the American College of Surgeons-National Surgical Quality Improvement Program. Journal of the American College of Surgeons. Elsevier; 2006 Nov;203(5):618–24.

10. Ackerman AD. Morbidity and Mortality Conference. Pediatric Critical Care Medicine. 2016 Jan;17(1):94–5.

11. Aaronson E, Wittels K, Nadel E, Schuur J. Morbidity and Mortality Conference in Emergency Medicine Residencies and the Culture of Safety. WestJEM. 2015 Nov 12;16(6):810–7.

12. Weiss KB, Bagian JP, Wagner R. CLER Pathways to Excellence: Expectations for an Optimal Clinical Learning Environment (Executive Summary). Journal of Graduate Medical Education. 2014 Sep;6(3):610–1.

13. McVeigh TP, Waters PS, Murphy R, O'Donoghue GT, McLaughlin R, Kerin MJ. Increasing Reporting of Adverse Events to Improve the Educational Value of the Morbidity and Mortality Conference. Journal of the American College of Surgeons. American College of Surgeons; 2013 Jan 1;216(1):50–6.

14. Kong CH, Guest GD, Stupart DA, Faragher IG, Chan STF, Watters DA. Recalibration and Validation of a Preoperative Risk Prediction Model for Mortality in Major Colorectal Surgery. Diseases of the Colon & Rectum. 2013 Jul;56(7):844–9.

15. Kong CH, Guest GD, Stupart DA, Faragher IG, Chan STF, Watters DA. Colorectal preOperative Surgical Score (CrOSS) for mortality in major colorectal surgery. ANZ J Surg. 2015 Mar 30;85(6):403–7.

16. Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY. Does Surgical Quality Improve in the American College of Surgeons National Surgical Quality Improvement Program. Transactions of the Meeting of the American Surgical Association. 2009;127:7–20.

17. Ingraham AM, Richards KE, Hall BL, Ko CY. Quality Improvement in Surgery: the American College of Surgeons National Surgical Quality Improvement Program Approach. Advances in Surgery. Elsevier Inc; 2010 Sep 1;44(1):251–67.

18. Henderson WG, Daley J. Design and statistical methodology of the National Surgical Quality Improvement Program: why is it what it is? AJS. Elsevier Inc; 2009 Nov 1;198(S):S19–S27.

19. Bilimora KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, et al. Development and Evaluation of the Universal ACS NSQIP Surgical Risk Calculator: A Decision Aid and Informed Consent Tool for Patients and Surgeons. Journal of the American College of Surgeons. American College of Surgeons; 2013 Nov 1;217(5):833–3.

20. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, et al. Development and Evaluation of the Universal ACS NSQIP Surgical Risk Calculator: A Decision Aid and Informed Consent Tool for Patients and Surgeons. Journal of the American College of Surgeons. 2013 Nov;217(5):833–3.

21. Szender JB, Frederick PJ, Eng KH, Akers SN, Lele SB, Odunsi K. Evaluation of the National Surgical Quality Improvement Program Universal Surgical Risk Calculator for a Gynecologic Oncology Service. International Journal of Gynecological Cancer. 2015 Mar;25(3):512–20.

22. Cologne KG, Keller DS, Liwanag L, Devraj B, Senagore AJ. Use of the American College of Surgeons NSQIP Surgical Risk Calculator for Laparoscopic Colectomy: How Good Is It and How Can We Improve It? Journal of the American College of Surgeons. American College of Surgeons; 2015 Mar 1;220(3):281–6.

23. Walter CJ, Collin J, Dumville JC, Drew PJ, Monson JR. Enhanced recovery in colorectal resections: a systematic review and meta-analysis. Colorectal Disease. 2009 May;11(4):344–53.

24. Wind J, Polle SW, Fung Kon Jin PHP, Dejong CHC, Meyenfeldt von MF, Ubbink DT, et al. Systematic review of enhanced recovery programmes in colonic surgery. Br J Surg. 2006;93(7):800–9.

25. Colorectal Resection and Primary Anastomosis in Patients Aged 70 and Older: Prospective Study. 2003 Nov 18;:1–5.

26. Retchin SM, Penberthy L, Desch C, Brown R, Clement , et all. Perioperative management of colon cancer under medicare risk programs. Arch Intern Med. 1997;157:1878-1884.

27. Kehlet H, Mogensen T. Hospital stay of 2 days after open sigmoidectomy with a multimodal rehabilitation programme. British Journal of Surgery. 1999, 86:227-230.

28. Gouvas N, Tan E, Windsor A, Xynos E, Tekkis PP. Fast-track vs standard care in colorectal surgery: a meta-analysis update. Int J Colorectal Dis. 2009 May 5;24(10):1119–31.

29. CS data collection, analysis and reporting. http://site.acsnsqip.org/program-specifics/data-collection-analysis-and-reporting.

30. Bodie G, Mayson K, Hong T, Panton N, Meneghetti A, Warnock GL, et al. Impact of An Enhanced Recovery After Surgery (ERAS) Program on Outcomes of Patients Undergoing Elective Colon Resection. Manuscript in preparation.

31. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. Journal of Clinical Epidemiology. 2005 May;58(5):475–83.

32. Frank E, Harrell JR, Kerry I, Lee and Daniel B. Mark.Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy and measuring and reducing errors. Statistics in Medicine. 1996; 15:361–387.

33. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Annals of Medicine. 1999 March 130; 6.

34. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009 May 28;338 (may28 1):b605–5.

35. Neuman MD, Charles LB. What we talk about when we talk about risk: refining surgery's hazards in medical thought. The Milbank Quarterly. 2012; 90(1) :135–159.

36. Janssen KJM, Moons KGM, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. Journal of Clinical Epidemiology. 2008 Jan;61(1):76–86.

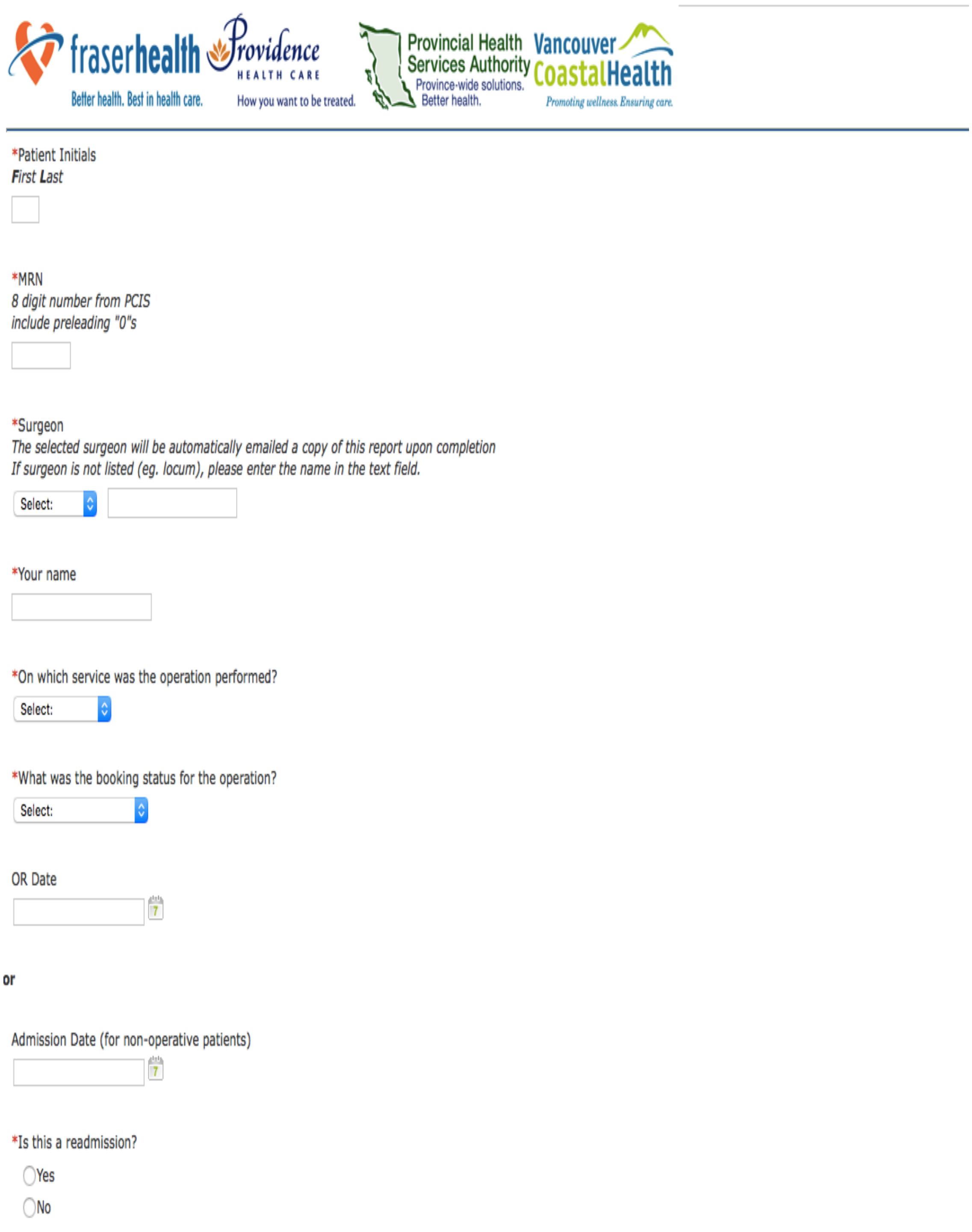
37. Steyerberg EW, Borsboom GJ, Van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Statistics in Medicine. 2004; 23:2567–2586.

38. Sacks GD, Dawes AJ, Ettner SL, Brook RH, Fox CR, Russell MM, et al. Impact of a Risk Calculator on Risk Perception and Surgical Decision Making. Annals of Surgery. 2016 May;:1–7.

39. Sacks GD, Dawes AJ, Ettner SL, Brook RH, Fox CR, Maggard-Gibbons M, et al. Surgeon perception of risk and benefit in the decision to operate. Annals of Surgery. 2016 May.

Appendices

Appendix A: UBC M&M Tool webpage



fraserhealth — Better health. Best in health care.
Providence HEALTH CARE — How you want to be treated.
Provincial Health Services Authority — Province-wide solutions. Better health.
Vancouver CoastalHealth — Promoting wellness. Ensuring care.

*Patient Initials
First Last

*MRN
8 digit number from PCIS
include preleading "0"s

*Surgeon
The selected surgeon will be automatically emailed a copy of this report upon completion
If surgeon is not listed (eg. locum), please enter the name in the text field.
Select:

*Your name

*On which service was the operation performed?
Select:

*What was the booking status for the operation?
Select:

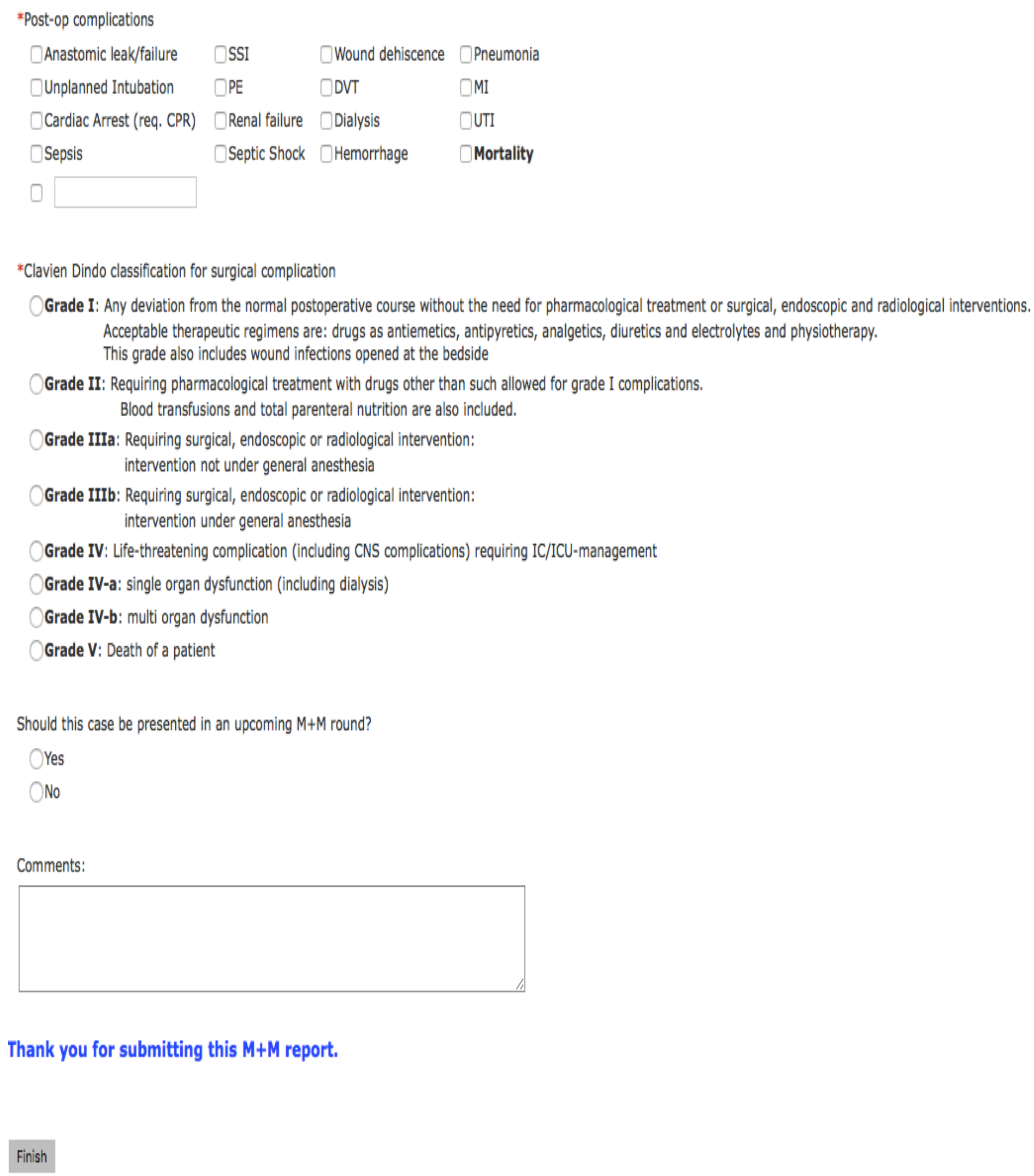
OR Date

or

Admission Date (for non-operative patients)

*Is this a readmission?
Yes
No

Figure (1) UBC M&M Tool website page



*Post-op complications

- ☐ Anastomic leak/failure ☐ SSI ☐ Wound dehiscence ☐ Pneumonia
- ☐ Unplanned Intubation ☐ PE ☐ DVT ☐ MI
- ☐ Cardiac Arrest (req. CPR) ☐ Renal failure ☐ Dialysis ☐ UTI
- ☐ Sepsis ☐ Septic Shock ☐ Hemorrhage ☐ **Mortality**
- ☐

*Clavien Dindo classification for surgical complication

- ◯ **Grade I**: Any deviation from the normal postoperative course without the need for pharmacological treatment or surgical, endoscopic and radiological interventions.
  Acceptable therapeutic regimens are: drugs as antiemetics, antipyretics, analgetics, diuretics and electrolytes and physiotherapy.
  This grade also includes wound infections opened at the bedside
- ◯ **Grade II**: Requiring pharmacological treatment with drugs other than such allowed for grade I complications.
  Blood transfusions and total parenteral nutrition are also included.
- ◯ **Grade IIIa**: Requiring surgical, endoscopic or radiological intervention:
  intervention not under general anesthesia
- ◯ **Grade IIIb**: Requiring surgical, endoscopic or radiological intervention:
  intervention under general anesthesia
- ◯ **Grade IV**: Life-threatening complication (including CNS complications) requiring IC/ICU-management
- ◯ **Grade IV-a**: single organ dysfunction (including dialysis)
- ◯ **Grade IV-b**: multi organ dysfunction
- ◯ **Grade V**: Death of a patient

Should this case be presented in an upcoming M+M round?

- ◯ Yes
- ◯ No

Comments:

Thank you for submitting this M+M report.

Finish

Figure (2) UBC M&M Tool website page

Appendix B: ACS NSQIP Surgical Risk Calculator software example

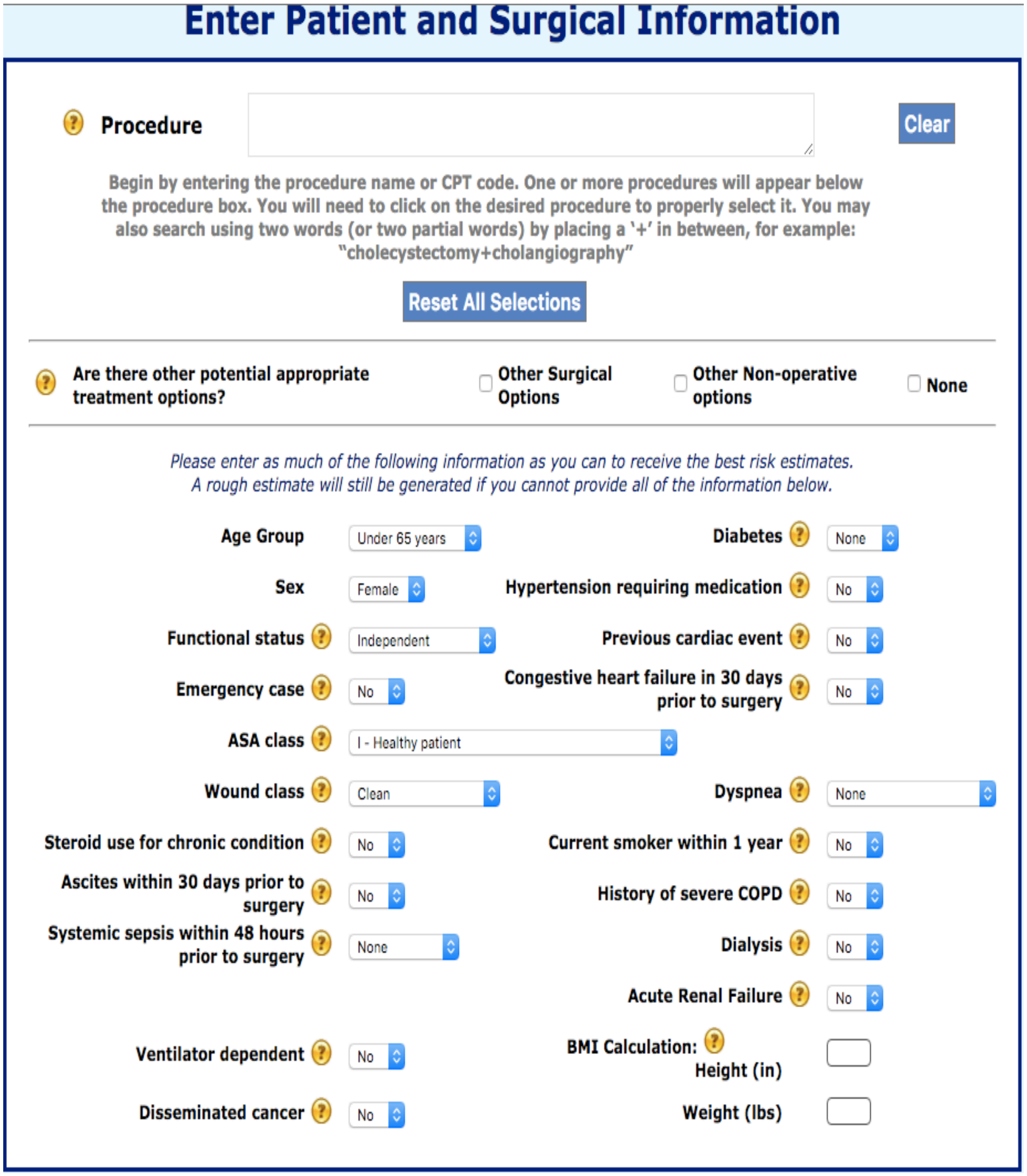


Figure (1) Data required in the RC software

Figures (1) & (2) Screenshots of the ACS NSQIP Surgical Risk Calculator (http://riskcalculator.facs.org). (A) Risk factor entry screen. (B) Example of report screen.

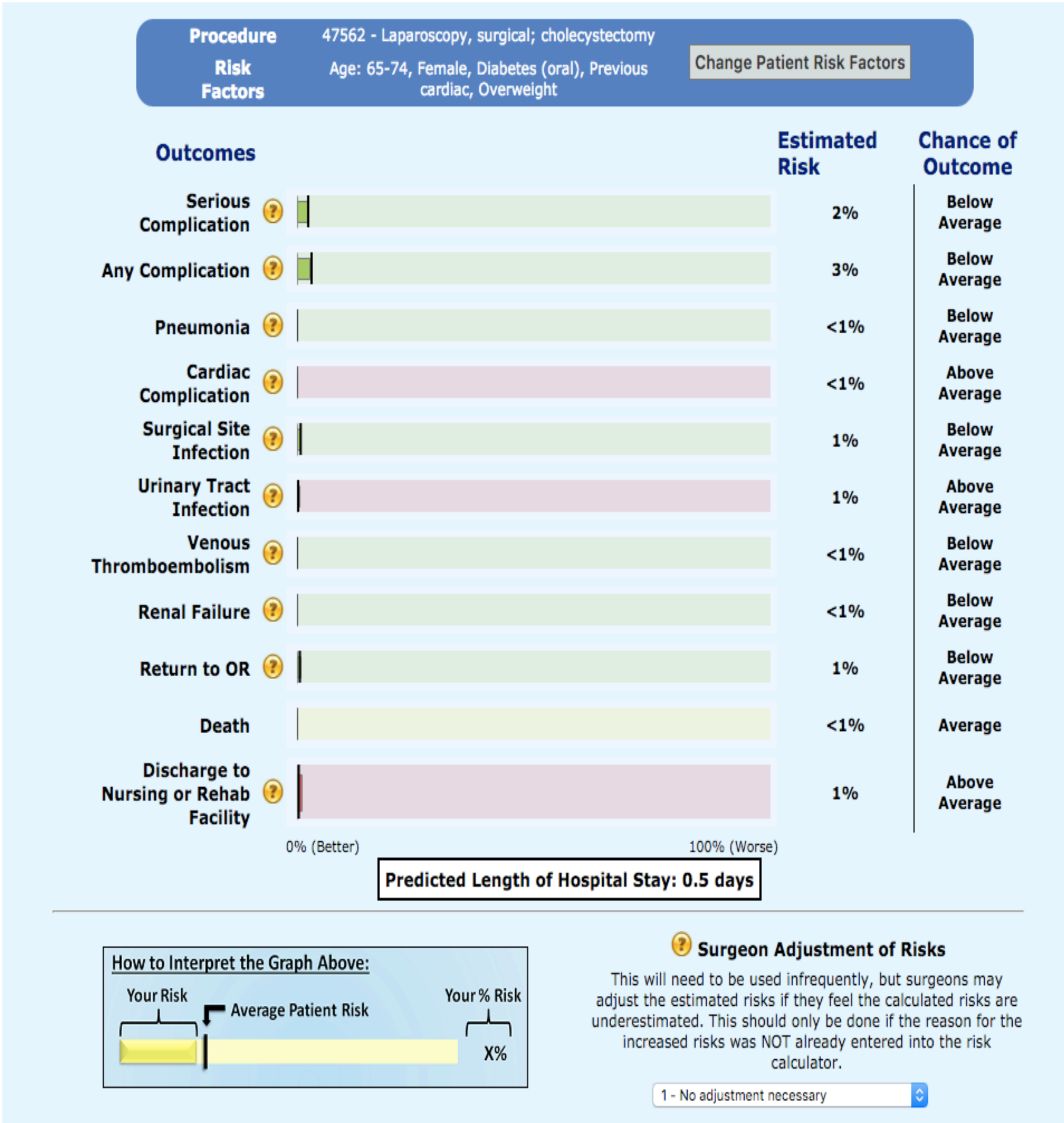


Figure (2) Example of report screen

Figures (1) & (2) Screenshot of the ACS NSQIP Surgical Risk Calculator (http:// riskcalculator.facs.org)(A) Risk factor entry screen. (B) Example of report screen.