Published in final edited form as:

Surgery. 2018 December ; 164(6): 1209–1216. doi:10.1016/j.surg.2018.05.043.

Postoperative Bleeding Risk Prediction for Patients Undergoing Colorectal Surgery

David Chen, Ph.D.¹, Naveed Afzal, Ph.D.¹, Sunghwan Sohn, Ph.D.¹, Elizabeth B. Habermann, Ph.D.², James M. Naessens, Sc.D.², David W. Larson, M.D., M.B.A.³, and Hongfang Liu, Ph.D.¹

¹Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA

²Department of Health Science Research, Mayo Clinic, Rochester, MN, USA

³Department of Colorectal Surgery, Mayo Clinic, Rochester, MN, USA

Abstract

Background: There is limited consensus regarding risk factors for postoperative bleeding. The objective of this work was to investigate the capability of machine learning techniques in combination with practice-based longitudinal electronic medical record (EMR) data for identifying potential new risk factors for postoperative bleeding and predicting patients at high risk of postoperative bleeding.

Study Design: A retrospective study was conducted for patients who underwent colorectal surgery 1998-2015 at a single tertiary referral center. Various predictors were extracted from EMR. The outcome of interest was the occurrence of postoperative bleeding within 7 days of surgery. Logistic regression (LR) and gradient boosting machine (GBM) models were trained. Area under the receiver operating curve (AUROC) and area under the precision recall curve (AUPRC) were used to evaluate the performance to different models.

Results: Of 13,399 cases undergoing colorectal resection, 1,680 (12.5%) experienced postoperative bleeding. A total of 299 variables were evaluated. LR and GBM models returned AUROC of 0.735 and 0.822, and AUPRC of 0.287 and 0.423 respectively. In addition to well-known risk factors for postoperative bleeding, nutrition (ranked 3rd), weakness (ranked 5th), patient mobility (ranked 6th), and activity level (ranked 8th) were found to be novel predictors in the GBM model based on permutation importance

Conclusion: The study identified measures of functional capacity of patient as novel predictors of postoperative bleeding. The study demonstrated that risk of postoperative bleeding can be assessed, allowing for better utilization of human resources in addressing this important adverse event following surgery.

Hongfang Liu: Phone: 1-507-293-0057; Fax: 1-507-284-1516, Liu.hongfang@mayo.edu, Mayo Clinic, 200 First St. SW, Rochester, MN 55905

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

CRS; transfusion; bleeding; machine learning; risk assessment; adverse events

Introduction

Postoperative bleeding (POB) is a serious complication following colorectal surgery (CRS). Incidence of POB is reported to occur between 1-14% in all CRS cases¹⁻⁸. Although POB is rarely fatal by itself, POB is often the beginning of a cascade of serious complications, including ileus and anastomotic leak, which carry significant morbidity and mortality⁹. Furthermore, therapeutic transfusions are also not benign, and may result in transfusion related acute lung injury, circulatory overload, immunosuppression, infection, and more^{10,11}. Transfusions and related complications prolong hospitalizations adding an additional 2.5 days to the average length of stay and \$17,000 to the average cost of stay^{12,13}.

There have been extensive efforts to study the possible risk factors for bleeding related to CRS using traditional statistical methods. Various studies by Golda et al, Kirchhoff et al, and Benoist et al in smaller, more specific patient cohorts have identified various risk factors such as age, sex, hemoglobin levels, comorbidities, and additional procedures^{2,3,5}. Manilich *et al.* used more modern variable selection methods to identify comorbidities and surgical procedural properties which were important to predicting postoperative transfusion but did not report model performance⁶. Although many of these risk factors have been corroborated in multiple studies, there have been very limited efforts in developing an individualized risk prediction model. Potential reasons include the complexity of the underlying cause, and lack of standardization of practices. One recent study by Kim et al. developed a nomogram to predict perioperative blood transfusion after hepatopancreaticobiliary and CRS using traditional statistical methods and achieved an area under the operating curve of 0.756⁷.

Machine learning (ML) has been touted as a way to develop models that represent difficult nonlinear systems and deal with large number of potential variables available in modern electronic medical records (EMR). ML methods have been applied in many healthcare applications including prediction of cancer susceptibility, automated classification of clinical imaging, and predicting prognosis following transplant due to the availability of practice-based longitudinal datasets from the widespread adoption of electronic health records¹⁴⁻¹⁶. It has been shown that ML methods outperform classic statistical regression methods in addressing many healthcare problems, particularly in cases utilizing high-dimensional “big data”^{17,18}. Our objective of this study is to investigate the use of ML and EMR for identifying potential new risk factors for POB and predicting patients at high risk of POB immediately following CRS.

Methods

Data

This study was approved by the Mayo Clinic institutional review board. A retrospective dataset was constructed including 13,399 surgeries (demographics summarized in Table 1)

with a colorectal procedure as the primary procedure performed at Mayo Clinic Hospital, Methodist campus; a large academic medical center in Rochester, Minnesota, USA. The primary outcome was bleeding, which was defined by any type of gastrointestinal hemorrhage occurring within one week following CRS. The National Surgical Quality Improvement Program (NSQIP) identifies POB based on exceeding a threshold of transfusion¹⁹. Alternatively, the Center for Medicare and Medicaid Services uses Patient Safety Indicator 09 as a quality measure which identifies POB by diagnostic and procedural code combinations²⁰. However, it is well known that ICD-9 codes are often inaccurately coded and time-stamped²¹. To address the difficulty of using claims data to detect bleeding, an algorithm to detect bleeding events in the EMR validated by manual medical record review by Moriarty et al. was used to minimize false positive events²². Specifically, bleeding events were identified by cross-validating ICD-9 diagnosis and procedure codes with laboratory results (drop of hemoglobin concentration greater than 3 g/dL) and/or transfusion records.

Variables Studied

A total of 117 variables falling under categories of demographics (4), patient provided information (19), symptoms (15), comorbidities (23), physiological measurements (11), laboratory tests (35), observational assessments (5), and operative factors (5), were directly extracted from EMR. All data types and variables are summarized in Table 2. Patient provided information was taken at time of admission. Symptoms, physiological values, laboratory tests, observational factors were abstracted from two weeks prior to surgery until start of surgery. Patient comorbidities were found from ICD-9 codes recorded within one year of surgery.

Definitions for symptoms and comorbidities ICD-9 codes can be found in supplementary materials. Observational assessments were taken from flowsheets with well-defined numeric scales. For example, nutrition, mobility, and activity level were found on general nursing assessments and assessed as defined by Braden Scale for prediction of pressure ulcers²³. Categorical physiological parameters (heart rhythm, respiratory effort, and respiratory pattern) were converted to binary variables of normal or abnormal. Although additional information can be included by more granular categorization, binary classification reduces variability introduced in the data documentation process.

Although many variables here such as demographics, comorbidities, and operative factors only have one valid value for each case, others such as physiological measurements may change throughout the course of a hospital stay. These temporally changing variables very well could represent changing risk of POB. However, representing these time-varying variables is not straight forward because different variables are updated at different frequencies (both between variables and between patients).

A method similar to the one used by Citi *et al* was used to convert the time-varying variables into scalar variables while preserving information contained in the dynamics of the signal²⁴. The minimum, mean, and maximum values were found for all physiological values for preoperative time window. The minimum, mean, and maximum changes of the physiological values were also measured to abstract their temporal change information. The first

measurement postoperation of temporal variables (e.g. first postoperative systolic blood pressure measurement) was also included. A total of 78 scalar variables were created from the 11 time-varying physiological measurements.

In addition to the 283 variables abstracted from EMR, another 16 were generated using MedTagger, an open source natural language processing (NLP) tool for extracting medical information from unstructured text^{25,26}. The 16 variables were chosen through consultation with subject matter experts. Clinical notes written between admission and surgery were used for extracting those variables. In total, 299 variables were created for the final dataset.

Model Training

The dataset extracted from EMR had a high rate of missingness. Variables that were missing more than 50% of data were excluded to ensure generalizability. The remaining missing data was imputed using a model-based imputation method. Imputation has been shown to yield better, more accurate and better fit due to more accessibility to a greater number of observations. Specifically, model based imputation has been shown to yield better fitted models compared to mean value or last value imputation because of better representation of the true distribution of the variables using model based imputation²⁷. Imputation was done using the “MissForest” package in R, which imputes missing values using a popular machine learning method, random forest²⁸.

Ability to predict POB was compared between LR and gradient boosting machine (GBM). GBM is a robust ML method similar to random forest which creates an ensemble of simple decision trees. Although each individual tree performs poorly, the ensemble decision of models gives a robust result by iteratively training decision trees to address misclassified cases. Unlike logistic regression or a single decision tree, ensemble methods are capable of capturing complex, nonlinear interactions between variables because individual trees can be grown using only a subset of variables. GBM has been widely applied to many clinical tasks with impressive results^{29–31}. All models were built and trained in R using the R base and “gbm” packages respectively. Both methods are robust to missing data and were trained from both the raw and imputed datasets.

Evaluation of each model was performed using 10 runs of 10-fold cross-validation to assess the predictive power of the model while avoiding overfitting. The dataset was semi-randomly divided into 10 subsets, with each subset containing roughly the same number of positive results (patients who experienced POB). A model was trained on 9 subsets and evaluated using the 10th. This process was repeated for all 10 subsets. The 10-fold cross-validation was then repeated with randomized sampling. Due to the highly unbalanced data, the observations with positive results of the training set were oversampled using the “SMOTE” package in R version 3.3.3. A model was learned using perioperative variables, and using only preoperative and intraoperative variables to determine the necessity of postoperative laboratory tests and observations.

Model Evaluation and Simplification

Area under the receiver operating curve (AUROC) and area under the precision recall curve (AUPRC) were used to assess the performance of different models. T-tests were also used to

evaluate if the curves are significantly different. Variable importance was evaluated using permutation importance (mean decrease in accuracy) from the random forest and the Boruta algorithm to identify important variables, reduce overfitting and redundancy, and simplify the models to improve generalizability³². In the Boruta algorithm, variable importance of predictors is iteratively compared with “shadow variables”, which are randomized and shuffled copies of predictors in the original dataset that have their correlations with the dependent variable removed. Predictors with variable importance significantly higher than the shadow variables are tagged as important, whereas those predictors which are not are tagged as unimportant. This is repeated until all variables are tagged as important or unimportant.

A final simplified ML model found empirically by reducing the number of variables until AUROC is significantly lower (measured by Venkatraman’s test comparing receiver operating characteristic curves) compared to the model developed using all preoperative and intraoperative predictors in the best performing model³³.

Results

Variable selection

A total of 12,402 unique patients undergoing 13,399 CRS procedures between 1998 and 2014 were selected. A total of 1,680 cases (12.5%) were found to experience POB. Overall, the 30-day postoperative mortality rate was 0.8% (109 cases).

Boruta variable selection of the training cohort identified the following 43 variables: age, BMI, procedure, surgical length, ASA score, laparoscopic, intraoperative complication, anemia, embolism, sepsis, ulcer, inflammation, injury, weakness, preoperative bleeding, malignant colon cancer, malignant rectal cancer, malignant other cancers, heart failure, coronary artery disease, kidney disease, colitis, hemophilia, activity level, fall risk score, mobility, nutrition, pain, diastolic blood pressure, systolic blood pressure, heart rate, O₂ saturation, respiratory rate, temperature, aspartate aminotransferase, hemoglobin level, leukocyte level, glucose, platelet, creatinine, potassium, sodium, and NLP abstracted abscess. The distribution and missingness of selected variables are shown in Table 3.

Model Performance

Table 4 shows the benefit of data imputation. Only the intraoperative and preoperative variables were used in building the models. Imputed data yields statistically better performance in both models, although the effect of imputation is much greater on the LR model compared to the GBM model.

Table 5 shows the performance difference of different subsets of variables. The GBM models outperforms the LR models with any subset of variables. There is a significant drop in performance when not including postoperative variables (p-values<0.001). There is a similar drop in AUROC between both classifier (~0.05), although the GBM model has a more significant drop in AUPRC (0.09 compared to 0.03 drop). Models learned from Boruta variable selection do not have reduced performance compared to models learned using all pre- and intra-operative variables (p-value of 0.352 and 0.724 of the LR and GBM models).

LR performance increases slightly more compared to GBM. Performance of LR and GBM models suffers when using only known variables using both measures of performance. Models developed from only known variables are significantly worse in both LR and GBM methods (p-value <0.001).

Table 6 shows a comparison of variable importance between the GBM and coefficients found using multivariate LR. Both models found the majority of known risk factors to be significant predictors (measured through statistical testing for LR model or permutation importance for GBM). There are several significant differences in variables which the LR model found to be significant but were fairly low importance (<5% relative to the variable of highest importance). These include inflammation, malignant cancers, coronary artery disease, diastolic blood pressure, laparoscopic procedure, and intraoperative complications. Conversely, the GBM model found systolic blood pressure, respiratory rate, hemoglobin levels, aspartate aminotransferase levels, activity level, mobility, and nutrition to be important variables (variable importance >5% relative to most important variable) whereas the LR model did not find them to be significant.

Figure 1 shows the results of stepwise model simplification. The AUROC of the GBM model increases as variables are added up to the 10th ranked variable, subsequently which the AUROC plateaus and does not significantly change. The AUROC of the LR model has a similar initial improvement in AUROC compared to the GBM as the first 5 variables are added. However, additional variables cause large variation in AUROC. The final simplified model was found to contain anemia, hemophilia, nutrition, surgical length, weakness, mobility, heart failure, activity level, kidney disease, and preoperative bleeding. The AUROC of this final simplified model was found to be 0.812 and 0.734 with the GBM and LR classifiers respectively.

Discussion

POB following CRS is a serious complication associated with substantial increases in cost of care, length of stay, and risk of morbidity and mortality¹. However, the relatively low rate of POB and limited knowledge of causes and predictors of POB makes postoperative surveillance attention intensive and expensive. Furthermore, assessment of multiple risk factors (particularly when the risk factors are nonlinearly related) is difficult for providers to properly integrate during busy post-surgical care periods. The model developed in this work, however, can be used immediately following surgery to identify patients at-risk for POB.

The presented model may improve the value of postoperative care. Providers may choose to not order blood tests for patients at low-risk, therefore reducing overall cost of care, while patients at high-risk could also be targeted for more frequent evaluation and lab tests. In addition, this directed allocation of attention can be particularly beneficial to reducing the time to identification of POB.

The results presented here demonstrate the utility of machine learning methods in predicting POB in CRS patients using data from EMR. Unlike many prior studies, variables were extracted from the EMR with few limits on the scope. Compared to well annotated and

carefully planned clinical registry data (such as ACS-NSQIP), this methodology allows for a more unbiased, data driven approach to research. With this approach, a large number of patients may be included with minimum overhead, whereas adding additional patients to clinical registries can be prohibitively expensive in both time and money.

However, casting such a wide net for data also results in a large number of studied variables, which can be difficult to manage using traditional statistical methods, as shown by poor LR performance when including all variables in this study. This limitation of traditionally used LR is significant given the large number of variables available in modern EMR.

Furthermore, the LR classifier does not achieve comparable performance even following variable selection, suggesting that the underlying cause of POB is complex and nonlinear. In particular, using a GBM classifier without postoperative variables (AUC=0.819) performed statistically better ($p>0.001$) compared to the LR classifier with postoperative variables (AUC=0.773).

The data used in this work reinforces well-established risk factors: anemia, hemophilia, history of bleeding disorders, surgery length, heart failure, and kidney disease^{1,6,7,34}. Although not included in the final simplified model, Boruta variable selection did find age, ulcers, open surgery, and malignant cancer as significant predictors³⁵. These results are correlated with significant variables in the LR model. Previously identified risk factors such as diabetes⁶, COPD⁵, and prothrombin time⁷ were not found to be important predictors for POB in this study. Cohorts between those studies and the current one were significantly different. Diabetes with and without complications were combined in the current dataset whereas only diabetes with complications was found to be predictive of POB. The study which included COPD as a risk factor was specific for anastomotic bleeding. Prothrombin time was excluded from the analysis in this work due to high level of missing.

The LR model gives greater weights to known variables compared to variable importance in GBM model. The use of regression analysis to identify these risk factors likely had an impact in the correlation. In contrast to comorbidities and symptoms, LR found very few significant predictors for POB from physiological measurements. Only diastolic blood pressure, glucose levels, and potassium levels were found to be significant. LR is more sensitive to pre-specifying the nature of the relationship (e.g., linear, quadratic or exponential increase in risk with increasing values) than GBM. Another possible cause for this is that there is high collinearity between several of the predictors (e.g. anemia, heart rate, respiratory rate, O₂ saturation, and hemoglobin levels). The association between anemia and weakness is significant (chi-square p-value < 0.001).

One disadvantage of Boruta variable selection method is it does not identify collinearity between variables. Collinearity between variables can cause unreliable and unstable estimates of regression coefficients, resulting in poor estimates of their significance. Indeed, removing many of these variables has a positive result on LR classifier performance (as shown in table 5). In contrast, many ML methods are not negatively affected by multicollinearity between variables. GBM solves multicollinearity problems by including individual decision trees which include only one variable or the other³⁶. This is exemplified

by the high variance in AUROC with differing variables in the LR model compared to the stability of the GBM model.

In addition to the established risk factors, weakness, nutrition, mobility, and activity level were found to be important variables in the GBM model. Weakness was defined as general muscle weakness found using the ICD-9 code for general muscle weakness. Many risk scores such as Modified Frailty Index attempt to assess the patient's physical state have been shown to correlate with poor patient outcomes and other complications^{37,38}. Weakness being a strong predictor of POB along with anemia may also suggest that symptomatic presentation of anemia is more informative of adverse outcomes rather than anemia itself. A generalized linear model which included an interaction term between anemia and weakness found the interaction term to be significant (p-value < 0.001), supporting this hypothesis. Also, malnourishment has been previously found to be a risk factor for postoperative morbidity and mortality compared to well-nourished patients³⁹. The 4 predictors together support the idea that a patient's functional condition is a key factor for adverse outcomes.

One significant limitation of this work is the severity of POB was not taken into account. Cases with subclinical bleeding which do not require transfusion were also included as POB. Although transfusion or subsequent reoperation are considered clinically significant endpoints for bleeding, guidelines for blood transfusion vary between procedures and patient groups^{40,41}. Another limitation is the lengthy study period; the clinical definition of various comorbidities such as diabetes has changed over time. In particular, clinicians have become less aggressive with prescription of transfusion of blood products over time as the understanding of transfusion costs and related complications has evolved⁴². As such, the learned classifier may not reflect the correct distribution of preoperative bleeding cases according to current clinical practice.

Conclusion

Machine learning methods successfully identified risk factors for POB following CRS. These models, relying on preoperative and intraoperative variables, potentially allow for targeted surveillance and reduction of costs related to laboratory tests and human monitoring. This work can contribute to both evaluation of future CRS and gives insight on possible physiological pathways of POB.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding Source(s): NIBIB R01 EB19403

Abbreviations

(POB)	postoperative bleeding
(CRS)	colorectal surgery

(EMR)	electronic medical record
(ASA)	American society of anesthesiologist
(ML)	machine learning
(NLP)	natural language processing
(BMI)	body mass index
(AUROC)	area under the receiver operating curve
(AUPRC)	area under the precession recall curve
(LR)	logistic regression
(GBM)	gradient boosting machine
(BP)	blood pressure
(CA)	cancer
(CAD)	coronary artery disease
(COPD)	chronic obstructive pulmonary disease

References

1. Tasu JP, Vesselle G, Herpe G, et al. Postoperative abdominal bleeding. Diagnostic and Interventional Imaging 2015;96:823–31. [PubMed: 26078019]
2. Kirchhoff P, Dincler S, Buchmann P. A multivariate analysis of potential risk factors for intra-and postoperative complications in 1316 elective laparoscopic colorectal procedures. Annals of surgery 2008;248:259–65. [PubMed: 18650636]
3. Benoist S, Panis Y, Alves A, Valleur P. Impact of obesity on surgical outcomes after colorectal resection. The American journal of surgery 2000;179:275–81. [PubMed: 10875985]
4. McCloskey CA, Wilson MA, Hughes SJ, Eid GM. Laparoscopic colorectal surgery is safe in the high-risk patient: A NSQIP risk-adjusted analysis. Surgery 2007;142:594–7.e2. [PubMed: 17950353]
5. Golda T, Zepa C, Kreisler E, Trenti L, Biondo S. Incidence and management of anastomotic bleeding after ileocolic anastomosis. Colorectal Disease 2013;15:1301–8. [PubMed: 23710632]
6. Manilich E, Vogel J, Kiran R, Church J, Seyidova-Khoshknabi D, Remzi F. Key factors associated with postoperative complications in patients undergoing colorectal surgery. Diseases of the Colon & Rectum 2013;56:64–71. [PubMed: 23222282]
7. Kim Y, Bagante F, Gani F, et al. Nomogram to predict perioperative blood transfusion for hepatopancreaticobiliary and colorectal surgery. British Journal of Surgery 2016;103:1173–83. [PubMed: 27222214]
8. Yasui M, Ikeda M, Miyake M, et al. Comparison of bleeding risks related to venous thromboembolism prophylaxis in laparoscopic vs open colorectal cancer surgery: a multicenter study in Japanese patients. Am J Surg 2017;213:43–9. [PubMed: 26772140]
9. Boccola MA, Buettner PG, Rozen WM, et al. Risk Factors and Outcomes for Anastomotic Leakage in Colorectal Surgery: A Single-Institution Analysis of 1576 Patients. World Journal of Surgery 2011;35:186–95. [PubMed: 20972678]
10. Maxwell MJ, Wilson MJ. Complications of blood transfusion. Continuing Education in Anaesthesia, Critical Care & Pain 2006;6:225–9.

11. Tartter P Blood transfusion and infectious complications following colorectal cancer surgery. *British journal of surgery* 1988;75:789–92. [PubMed: 3167530]
12. Callum JL, Waters JH, Shaz BH, Sloan SR, Murphy MF. The AABB recommendations for the Choosing Wisely campaign of the American Board of Internal Medicine. *Transfusion* 2014;54:2344–52. [PubMed: 25100209]
13. Morton J, Anastassopoulos KP, Patel ST, et al. Frequency and Outcomes of Blood Products Transfusion Across Procedures and Clinical Conditions Warranting Inpatient Care: An Analysis of the 2004 Healthcare Cost and Utilization Project Nationwide Inpatient Sample Database. *American Journal of Medical Quality* 2010;25:289–96. [PubMed: 20530223]
14. Delen D, Oztekin A, Kong ZJ. A machine learning-based approach to prognostic analysis of thoracic transplantations. *Artificial Intelligence in Medicine* 2010;49:33–42. [PubMed: 20153956]
15. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal* 2015;13:8–17. [PubMed: 25750696]
16. Zacharaki EI, Kanas VG, Davatzikos C. Investigating machine learning techniques for MRI-based classification of brain neoplasms. *International journal of computer assisted radiology and surgery* 2011;6:821–8. [PubMed: 21516321]
17. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *Jama* 2013;309:1351–2. [PubMed: 23549579]
18. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44. [PubMed: 26017442]
19. Project ACoSACoSNSQI. User Guide for the 2015 Participant Use Data File. In: Project ACoSACoSNSQI, ed. 2016.
20. Indicators AQ. Patient Safety Indicator 09 (PSI 09) Perioperative Hemorrhage or Hematoma Rate: U.S. Department of Health and Human Services; 2017.
21. Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T. Comorbidities, complications, and coding bias: Does the number of diagnosis codes matter in predicting in-hospital mortality? *JAMA* 1992;267:2197–203. [PubMed: 1556797]
22. Moriarty JP, Daniels PR, Manning DM, et al. Going Beyond Administrative Data Retrospective Evaluation of an Algorithm Using the Electronic Health Record to Help Identify Bleeding Events Among Hospitalized Medical Patients on Warfarin. *American Journal of Medical Quality* 2016;1062860616660757.
23. Bergstrom N The Braden Scale for predicting pressure sore risk. *Nurs res* 1987;36:205–10. [PubMed: 3299278]
24. Citi L, Barbieri R. PhysioNet 2012 Challenge: Predicting mortality of ICU patients using a cascaded SVM-GLM paradigm Computing in Cardiology (CinC), 2012; 2012: IEEE p. 257–60.
25. Liu H, Bielinski SJ, Sohn S, et al. An information extraction framework for cohort identification using electronic health records. *AMIA Summits on Translational Science Proceedings* 2013;2013:149.
26. Torii M, Waghlikar K, Liu H. Using machine learning for concept extraction on clinical documents from multiple data sources. *Journal of the American Medical Informatics Association* 2011;18:580–7. [PubMed: 21709161]
27. Graham JW. Missing data analysis: Making it work in the real world. *Annual review of psychology* 2009;60:549–76.
28. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012;28:112–8. [PubMed: 22039212]
29. Motwani M, Dey D, Berman DS, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *European heart journal* 2016;38:500–7.
30. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PloS one* 2017;12:e0174944. [PubMed: 28376093]
31. Chekroud AM, Zotti RJ, Shehzad Z, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry* 2016;3:243–50. [PubMed: 26803397]
32. Kursu MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw* 2010;36:1–13.

33. Venkatraman ES, Begg CB. A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment. *Biometrika* 1996;83:835–48.
34. Kirchhoff P, Clavien P- A, Hahnloser D. Complications in colorectal surgery: risk factors and preventive strategies. *Patient Safety in Surgery* 2010;4:5. [PubMed: 20338045]
35. Moghadamyeghaneh Z, Mills SD, Pigazzi A, Carmichael JC, Stamos MJ. Risk factors of postoperative upper gastrointestinal bleeding following colorectal resections. *J Gastrointest Surg* 2014;18:1327–33. [PubMed: 24841442]
36. Nassif AB, Capretz LF, Ho D, Azzeh M. A treeboost model for software effort estimation based on use case points. *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*; 2012: IEEE p. 314–9.
37. Keller DS, Bankwitz B, Nobel T, Delaney CP. Using frailty to predict who will fail early discharge after laparoscopic colorectal surgery with an established recovery pathway. *Diseases of the Colon & Rectum* 2014;57:337–42. [PubMed: 24509456]
38. Robinson TN, Wu DS, Pointer L, Dunn CL, Cleveland JC, Moss M. Simple frailty score predicts postoperative complications across surgical specialties. *The American Journal of Surgery* 2013;206:544–50. [PubMed: 23880071]
39. Dempsey D, Mullen J, Buzby G. The link between nutritional status and clinical outcome: can nutritional intervention modify it? *The American journal of clinical nutrition* 1988;47:352–6. [PubMed: 3124596]
40. Carson JL, Grossman BJ, Kleinman S, et al. Red blood cell transfusion: a clinical practice guideline from the AABB. *Annals of internal medicine* 2012;157:49–58. [PubMed: 22751760]
41. Carson JL, Guyatt G, Heddle NM, et al. Clinical practice guidelines from the AABB: red blood cell transfusion thresholds and storage. *Jama* 2016;316:2025–35. [PubMed: 27732721]
42. Goodnough LT, Levy JH, Murphy MF. Concepts of blood transfusion in adults. *Lancet* 2013;381:1845–54. [PubMed: 23706801]

Figure 1:

Change in predictive power (AUROC) by variable. Variables are added based on variable importance rank (refer to table 6). GBM and LR models are represented by circles and triangles respectively.

Table 1:

Characteristics of patients with POB compared to those without. Percentages indicate incidence rate within a group.

	No POB n=11,719	POB n=1,680	p-value
Demographics			
mean age	57.6 ± 17.8	61.5 ± 16.7	<0.001
male sex	6,206 (53.0%)	865 (51.5%)	0.271
BMI	30.3 ± 11.0	27.8 ± 8.6	0.656
Race			
white	10,612 (90.6%)	1,447 (86.1%)	<0.001
black	49 (0.4%)	18 (1.1%)	<0.001
asian	74 (0.6%)	14 (0.8%)	0.433
other	164 (1.2%)	35 (2.1%)	0.041
alcohol use	723 (6.2%)	123 (7.3%)	0.078
smoker	2,097 (17.9%)	277 (16.5%)	0.168
drug use	180 (1.5%)	39 (2.3%)	0.023
Indications for surgery			
rectal cancer	2,060 (17.6%)	321 (19.1%)	0.134
colon cancer	5,098 (43.5%)	654 (38.9%)	<0.001
inflammatory bowel disease	3,325 (28.4%)	461 (27.4%)	0.444
other	2,916 (24.9%)	244 (14.5%)	<0.001
Surgical characteristics			
partial colectomy	4064 (34.7%)	728 (43.3%)	<0.001
total colectomy	4531 (38.7%)	626 (37.3%)	0.281
proctectomy	5553 (47.4%)	650 (38.7%)	<0.001
laparoscopic	3,919 (33.4%)	206 (12.3%)	<0.001
elective surgery	565 (4.8%)	211 (12.6%)	<0.001
median ASA score	2 ± 0.71	3 ± 0.79	<0.001
Outcomes			
mean hospital stay	6.5 ± 4.7	13.4 ± 13.8	<0.001
death	52 (0.4%)	57 (3.4%)	<0.001

Table 2:

List of variables.

Variable Types	Variables
Demographics	age ^{††} , BMI [†] , gender [†] , race
Patient provided information	alcohol usage, anesthesia complication, bleeding complication, difficulty concentrating, difficulty dressing, difficulty eating, difficulty moving, difficulty performing daily chores, difficulty sleeping, difficulty swallowing, difficulty taking medications, difficulty urinating, difficulty using the toilet, difficulty walking, difficulty with housekeeping, difficulty with leaking bladder, difficulty with pain, drug usage, smoking
Symptoms	abdominal pain, altered elimination, anemia ^{††} , angina, bleeding ^{††} , edema, hernia, inflammation [†] , injury [†] , nausea, sepsis [†] , transfusion, ulcer [†] , vertigo, weakness [†]
Comorbidities	benign rectal cancer, benign colon cancer, benign other cancer, cardiomyopathy ^{††} , cholangitis, cholecystitis, colitis [†] , chronic obstructive pulmonary disease, coronary artery disease ^{††} , cystitis, diabetes, diverticulitis [†] , embolism [†] , heart failure ^{††} , hemophilia ^{††} , hepatitis, hypertension, kidney disease [†] , malignant colon cancer ^{††} , malignant other cancer ^{††} , malignant rectal cancer ^{††} , malnutrition, nephritis, pancreatitis
Physiological Parameters	blood pressure (diastolic) [†] , blood pressure (systolic) [†] , central venous pressure, heart rate [†] , heart rhythm, mean arterial pressure, O ₂ saturation [†] , respiratory effort, respiratory pattern, respiratory rate [†] , temperature [†]
Laboratory Tests	alanine, albumin, alkaline phosphatase, anion, aspartate aminotransferase [†] , basophil, bicarbonate, bilirubin (direct), bilirubin (total), blood urea nitrogen, calcium, chloride, creatinine [†] , glomerular filtration rate, glucose [†] , hematocrit [†] , hemoglobin ^{††} , ionized calcium, leukocyte [†] , lymphocyte, magnesium, mean corpuscular volume, monocyte, neutrophil, osmolality, pH, phosphorus, platelet [†] , potassium [†] , prothrombin time, protein, sodium [†] , triglycerides, troponin
Observational Assessments	activity level [†] , ASA score [†] , fall risk score [†] , mobility [†] , nutrition [†] , pain [†]
NLP	abscess [†] , anemia, antibiotic treatment, anticoagulants, ascites, cellulitis, leaking wound, prophylaxis, purulent, steroid usage, tender abdomen, transfusion, wound infected
Medications	anticoagulants
Operational	intra-operational complications [†] , laparoscopic approach [†] , procedure ^{††} , surgery length ^{††}
Postoperative	abdominal assessment, activity level, activity tolerance, altered elimination, anion, bicarbonate, blood pressure (diastolic), blood pressure (systolic), blood urea nitrogen, bowel sounds, chloride, creatinine, glucose, heart rate, hematocrit, hemoglobin, incision condition, incision drainage amount, leukocyte, lymphocyte, magnesium, mobility, nausea, nutrition, O ₂ saturation, pain, pH, platelet, potassium, respiratory rate, sodium, temperature, urine clarity, urine color, vertigo

[†]Denotes known risk factors for POB^{††}Denotes variables which are found significant using Boruta method.^{†††}Denotes simplified variables found empirically.

Table 3:

Univariate analysis of patient characteristics of variables.

Risk Factor	Missing %	No POB (n=11,719)	POB (n=1,680)	p-value	Odds Ratio (99.75% CI)
Demographics and Outcomes					
age [‡]	0.00%	57.6 ± 17.8	61.5 ± 16.7	<0.001	
gender (M) [‡]	0.00%	6,206 (53.0%)	865 (51.5%)	0.271	0.94 (0.85, 1.04)
BMI	7.40%	27.8 ± 11.0	27.7 ± 8.6	0.657	
Symptoms					
anemia ^{‡¶}	0.00%	3,886 (33.2%)	1,176 (70.0%)	<0.001	4.70 (4.21, 5.26)
inflammation	0.00%	6,971 (60.0%)	899 (53.5%)	<0.001	0.78 (0.71, 0.87)
injury	0.00%	56 (0.5%)	17 (1.0%)	0.009	2.13 (1.23, 3.67)
preop-bleeding ^{‡¶}	0.00%	160 (1.4%)	207 (12.3%)	<0.001	10.2 (8.20, 12.5)
sepsis	0.00%	161 (1.4%)	169 (10.1%)	<0.001	8.03 (6.43, 10.0)
ulcer	0.00%	2,760 (23.6%)	336 (20.0%)	<0.001	0.81 (0.71, 0.92)
weakness	0.00%	1,159 (9.9%)	401 (23.9%)	<0.001	2.86 (2.51, 3.25)
Comorbidities					
any cancer	0.00%	7,449 (63.6%)	1,130 (67.3%)	0.003	1.18 (1.07, 1.31)
malignant rectal ca [‡]	0.00%	2,438 (20.8%)	506 (30.1%)	<0.001	1.64 (1.46, 1.84)
malignant colon ca [‡]	0.00%	3,791 (32.3%)	532 (31.7%)	0.595	0.97 (0.87, 1.08)
malignant other ca [‡]	0.00%	3,687 (31.5%)	795 (47.3%)	<0.001	1.96 (1.76, 2.17)
benign rectal ca	0.00%	721 (6.2%)	70 (4.2%)	0.002	0.66 (0.52, 0.85)
benign colon ca	0.00%	3,625 (30.9%)	371 (22.1%)	<0.001	0.63 (0.56, 0.71)
benign other ca	0.00%	1,446 (12.3%)	250 (14.9%)	0.004	1.24 (1.07, 1.44)
cardiomyopathy [‡]	0.00%	264 (2.3%)	55 (3.3%)	0.013	1.47 (1.09, 1.97)
colitis	0.00%	4,034 (34.4%)	574 (34.2%)	0.858	0.99 (0.89, 1.10)
CAD [‡]	0.00%	1,842 (15.7%)	419 (24.9%)	<0.001	1.78 (1.58, 2.01)
diabetes [‡]	0.00%	2,086 (17.8%)	343 (20.4%)	0.010	1.18 (1.04, 1.35)
diverticulitis	0.00%	3,438 (29.3%)	417 (24.8%)	<0.001	0.80 (0.71, 0.89)
embolism	0.00%	788 (6.7%)	250 (14.9%)	<0.001	2.43 (2.08, 2.82)
heart failure [‡]	0.00%	599 (5.1%)	177 (10.5%)	<0.001	2.19 (1.83, 2.61)
hemophilia [‡]	0.00%	466 (4.0%)	245 (14.6%)	<0.001	4.12 (3.50, 4.86)
hypertension	0.00%	4,688 (40.0%)	822 (48.9%)	<0.001	1.44 (1.30, 1.59)
kidney disease	0.00%	778 (6.6%)	203 (12.1%)	<0.001	1.93 (1.64, 2.28)
Physiological Values					
BP (diastolic)	7.20%	68.1 ± 11.93	65.0 ± 13.17	<0.001	
BP (systolic)	7.20%	81.7 ± 23.3	78.5 ± 23.2	<0.001	
heart rate	9.70%	82.3 ± 16.8	87.0 ± 19.5	<0.001	
O ₂ saturation	45.60%	96.0 ± 3.3	94.7 ± 4.6	<0.001	

Risk Factor	Missing %	No POB (n=11,719)	POB (n=1,680)	p-value	Odds Ratio (99.75% CI)
respiratory rate	36.90%	17.6 ± 5.4	18.9 ± 5.9	<0.001	
temperature	15.20%	36.8 ± 0.53	36.9 ± 0.69	<0.001	
Laboratory Tests					
AST	49.60%	25.6 ± 19.3	29.5 ± 29.4	<0.001	
creatinine	33.00%	1.09 ± 0.55	1.18 ± 0.69	<0.001	
glucose	43.30%	112.6 ± 51.7	126.5 ± 65.9	<0.001	
hematocrit	30.80%	36.8 ± 6.2	33.8 ± 6.3	<0.001	
hemoglobin [‡]	30.40%	12.1 ± 2.5	11.2 ± 2.4	<0.001	
leukocyte	30.70%	8.08 ± 3.94	8.85 ± 5.49	<0.001	
platelet	30.80%	280.2 ± 107.5	265.8 ± 123.1	<0.001	
potassium	39.50%	4.20 ± 0.53	3.98 ± 0.60	<0.001	
sodium	41.00%	139.9 ± 2.9	139.2 ± 9.4	<0.001	
Observational Assessment					
activity level [¥]	41.30%	3.72 ± 0.66	1.65 ± 0.86	<0.001	
ASA score	0.30%	2.25 ± 0.71	2.50 ± 0.79	<0.001	
fall risk score	45.60%	1.55 ± 1.43	1.98 ± 1.90	<0.001	
mobility [¥]	41.30%	3.73 ± 0.52	2.58 ± 0.72	<0.001	
nutrition [¥]	0.00%	3.18 ± 0.74	2.89 ± 0.79	<0.001	
pain	44.50%	2.12 ± 3.00	3.08 ± 3.51	<0.001	
Medications					
anticoagulants	0.00%	4,895 (41.8%)	755 (44.9%)	0.015	1.14 (1.03, 1.26)
NLP					
abscess	0.00%	793 (6.8%)	171 (10.2%)	<0.001	1.56 (1.31, 1.86)
Operational Variables					
complications	0.00%	313 (2.7%)	134 (8.0%)	<0.001	3.16 (2.56, 3.89)
laparoscopic	0.00%	3,919 (33.4%)	206 (12.3%)	<0.001	0.28 (0.24, 0.32)
procedure [‡]	0.00%				
4414X		2,803 (23.9%)	495 (29.5%)	<0.001	1.33 (1.19, 1.49)
4415X		1,261 (10.8%)	233 (13.9%)	<0.001	1.34 (1.15, 1.55)
4416X		2,108 (18.0%)	302 (18.0%)	1	1.00 (0.87, 1.14)
4420X		2,563 (21.9%)	110 (6.5%)	<0.001	0.25 (0.21, 0.31)
4421X		1,103 (9.4%)	78 (4.6%)	<0.001	0.47 (0.37, 0.59)
4511X		1,548 (13.2%)	427 (25.4%)	<0.001	2.24 (1.98, 2.53)
4512X		59 (0.5%)	13 (0.8%)	0.215	1.54 (0.84, 2.82)
4539X		274 (2.3%)	22 (1.3%)	<0.001	0.55 (0.36, 0.86)
surgical length ^{‡,¥}	0.00%	194.1 ± 92.8	277.3 ± 162.8	<0.001	

[‡]Denotes known risk factors for POB.

[¥]Denotes variables which were included in the final simplified model. Percentages indicate incidence rate within a group.

Table 4:

Comparison of the effects of imputation on AUROC and AUPRC.

	<u>Logistic Regression</u>			<u>Gradient Boosting Machine</u>		
	Missing	Imputed	p-value	Missing	Imputed	p-value
AUROC	0.712 ± 0.005	0.773 ± 0.005	<0.001	0.802 ± 0.002	0.819 ± 0.001	0.011
AUPRC	0.262 ± 0.007	0.311 ± 0.006	<0.001	0.419 ± 0.004	0.422 ± 0.004	0.009

Table 5:

Comparison of models developed using various subsets of variables. Selected denotes the subset of variables found using Boruta variable selection method. Known denotes the subset of variables which were previously found risk factors.

Variables	<u>LR</u>		<u>GBM</u>	
	AUROC	AUPRC	AUROC	AUPRC
Perioperative	0.773 ± 0.005	0.311 ± 0.006	0.865 ± 0.001	0.513 ± 0.003
Pre + Intraoperative	0.720 ± 0.003	0.279 ± 0.006	0.819 ± 0.001	0.422 ± 0.004
Selected	0.735 ± 0.004	0.287 ± 0.008	0.822 ± 0.001	0.423 ± 0.003
Known	0.672 ± 0.005	0.233 ± 0.009	0.752 ± 0.002	0.347 ± 0.004

Table 6:

Comparison of variable importance between different models which were trained using all preoperative and intraoperative variables which were found to be important using Boruta variable selection. The beta coefficient and p-value are shown for LR. The variable importance for GBM is based on permutation importance.

Risk Factor	LR β (p-value)	GBM (Rank)
anemia [‡]	1.252 (<0.001)	100 (1)
hemophilia [‡]	1.464 (<0.001)	47.4 (2)
nutrition	-0.114 (0.343)	41.7 (3)
surgical length [‡]	0.006 (<0.001)	36.3 (4)
weakness	0.776 (<0.001)	30.2 (5)
mobility	-0.368 (0.284)	25.6 (6)
heart failure [‡]	1.019 (<0.001)	18.4 (7)
activity level	-0.050 (0.320)	13.2 (8)
kidney disease [‡]	0.908 (0.004)	12.0 (9)
preop-bleeding [‡]	2.230 (<0.001)	8.1 (10)
aspartate aminotransferase	0.013 (0.131)	7.9 (11)
BP (systolic)	-0.001 (0.141)	6.5 (12)
hemoglobin [‡]	-0.020 (0.413)	6.4 (13)
respiratory rate	-0.022 (0.148)	6.3 (14)
ASA score	0.027 (0.321)	6.3 (15)
age [‡]	0.006 (0.028)	6.1 (16)
sepsis	1.711 (<0.001)	5.7 (17)
abscess	0.351 (<0.001)	4.0 (18)
procedure [‡]	-0.527 (<0.001)	3.5 (19)
fall risk score	-0.035 (0.217)	2.9 (20)
embolism	0.644 (0.171)	2.5 (21)
pain scale	0.122 (0.276)	2.5 (22)
laparoscopic	-0.356 (<0.001)	2.4 (23)
CAD [‡]	0.431 (<0.001)	2.2 (24)
sodium	-0.004 (0.237)	2.1 (25)
heart rate	-0.001 (0.189)	1.7 (26)
O ₂ saturation	-0.052 (0.078)	1.6 (27)
complications	1.038 (<0.001)	1.2 (28)
temperature	-0.014 (0.281)	1.2 (29)
creatinine	-0.141 (0.198)	1.0 (30)
glucose	-0.002 (0.001)	0.9 (31)
potassium	-0.406 (0.006)	0.9 (32)
BP (diastolic)	-0.008 (0.009)	0.8 (33)

Risk Factor	LR β (p-value)	GBM (Rank)
other mal ca [‡]	0.337 (0.012)	0.6 (34)
rectal mal ca [‡]	0.414 (0.007)	0.3 (35)
leukocyte	-0.042 (0.399)	0.1 (36)
platelet	4.4×10^{-4} (0.179)	0.1 (36)
BMI	7.7×10^{-6} (0.535)	<0.1 (38)
any cancer	-0.161 (0.113)	<0.1 (39)
colon mal ca [‡]	-0.177 (0.032)	<0.1 (39)
colitis	0.230 (0.436)	<0.1 (39)
inflammation	0.024 (0.007)	<0.1 (39)
injury	-0.317 (0.424)	<0.1 (39)
ulcer	0.012 (0.117)	<0.1 (39)

[‡]Denotes known risk factors for POB.