Prescriptive Analytics for Reducing 30-day Hospital Readmissions after General Surgery

Dimitris Bertsimas,[1] Michael Lingzhi Li, [1] Ioannis Ch. Paschalidis,[2] Taiyao Wang[2]

[1]Operations Research Center,
Massachusetts Institute of Technology,
Cambridge, MA

[2]Center for Information and Systems Engineering,
Boston University,
Boston, MA

Corresponding Author:

Ioannis Ch. Paschalidis,
Department of Electrical and Computer Engineering,
Division of Systems Engineering,
and Department of Biomedical Engineering,
Boston University
8 Saint Mary's St.,
Boston, MA 02215
USA
e-mail: yannisp@bu.edu
http://sites.bu.edu/paschalidis
Tel: 617-353-0434

ABSTRACT

Introduction: New financial incentives, such as reduced Medicare reimbursements, have led hospitals to closely monitor their readmission rates and initiate efforts aimed at reducing them. In this context, many surgical departments participate in the American College of Surgeons National Surgical Quality Improvement Program (NSQIP), which collects detailed demographic, laboratory, clinical, procedure and perioperative occurrence data. The availability of such data enables the development of data science methods which predict readmissions and, as done in this paper, offer specific recommendations aimed at preventing readmissions.

Materials and Methods: This study leverages NSQIP data for 722,101 surgeries to develop predictive and prescriptive models, predicting readmissions and offering real-time, personalized treatment recommendations for surgical patients during their hospital stay, aimed at reducing the risk of a 30-day readmission. We applied a variety of classification methods to predict 30-day readmissions and developed two prescriptive methods to recommend pre-operative blood transfusions to increase the patient's hematocrit with the objective of preventing readmissions. The effect of these interventions was evaluated using several predictive models.

Results: Predictions of 30-day readmissions based on the entire collection of NSQIP variables achieve an out-of-sample accuracy of 87% (Area Under the Curve—AUC). Predictions based only on pre-operative variables have an accuracy of 74% AUC, out-of-sample. Personalized interventions, in the form of pre-operative blood transfusions identified by the prescriptive methods, reduce readmissions by 12%, on average, for patients considered as candidates for pre-operative transfusion (pre-operative hematoctic <30). The prediction accuracy of the proposed models exceeds results in the literature.

Conclusions: This study is among the first to develop a methodology for making specific, data-driven, personalized treatment recommendations to reduce the 30-day readmission rate. The reported predicted reduction in readmissions can lead to more than $20 million in savings in the U.S. annually.

Keywords: hospital readmissions; surgery; quality improvement; predictive analytics; prescriptive analytics; machine learning.

1. Introduction

The United States spends $3 trillion annually on healthcare, corresponding to more than 17% of the U.S. GDP and far exceeding the next-highest spender among high-income countries.[1] While many factors contribute to higher spending, hospital readmissions, defined as an additional admission to address the same issue within 30 days after discharge, are an important −and potentially preventable− source of excessive resource utilization.[2,3]

In an effort to reduce unnecessary costs, the Affordable Care Act of 2012 introduced financial penalties for hospitals with readmission rates above the national average. While these measures have so far concentrated on medical conditions (e.g., acute myocardial infarction, congestive heart failure, pneumonia) and common orthopedic procedures (e.g., hip and knee arthroplasty), the list could expand to include common general surgical procedures.

In anticipation of these changes, Surgical Departments have started to closely monitor their readmission rates, and establishing processes aimed at reducing them. Several authors have sought to determine common causes of readmission after general surgical procedures, and most appear to relate to pre-existing conditions[4–7] and complications after surgery.[8]

In 2005, the American College of Surgeons (ACS) established the National Surgical Quality Improvement Program (NSQIP), which collects detailed demographic, laboratory, clinical, procedure and perioperative occurrence data, currently for General Surgery, and eventually in several subspecialties. The availability of such data, enables the development of *data analytics* methods relevant to the readmission reduction efforts.

While earlier work has primarily focused on readmission *predictive methods*, there has only been limited attention given to specific interventions with the potential to reduce readmissions; and that has focused mostly on post-discharge care.[9–11] Earlier work on predictive methods for hospitalizations have been successful but focused on specific diseases.[12–14]

The objective of this work is to develop more direct *prescriptive methods* that offer specific treatment recommendations during the patients' hospital stay with the potential to reduce readmission risk. Our recommended interventions are driven by data; essentially, for each patient, we learn from data what

has been effective in preventing a readmission for other "similar" patients. While the methodologies we develop are general and can be applicable to any sort of interventions, we focus on in-hospital treatment because the NSQIP data we leverage contain only such variables. We further focus on the patients' pre-operative hematocrit because it is commonly measured, important for assessing readmission risk, and easily modulated through blood transfusion.

2. Materials and Methods

Data and pre-processing. The ACS-NSQIP dataset we use in our analysis contains over 300 variables on comorbidities, intra-operative events, and 30-day outcomes using prospective random sampling,[15] including: (i) baseline demographic and health care status characteristics (e.g., age, gender, race, BMI, smoking, diabetes, hypertension, admittance from the ER); (ii) procedure information (e.g., CPT codes, ICD9 codes, ASA classification, wound classification); (iii) pre-operative, intra-operative, and post-operative variables, such as hospital stay information, Surgical Site Infections (SSI, superficial/deep/organ space) and complications (e.g., pneumonia, infections, bleeding, thromboembolic events), and (iv) laboratories, including pre-operative and post-operative values.

The NSQIP dataset at our disposal included more than 2.2 million surgeries during 2011-2014. While the NSQIP program provided high-quality manually curated data obtained from trained data abstractors, the variable definitions change over time. Specifically, the definitions of the occurrences listed in Table 2 (e.g., sepsis, pneumonia, SSIs) have changed multiple times. To avoid comparisons among variables with a different meaning, we selected only surgeries that took place during 2014. We included only variables that were continuously monitored and used throughout this period; resulting in a total of 231/187 patient variables for post-operative/pre-operative analysis. Patients who died within 30 days of surgery without readmission were excluded. There were a total of 722,101 remaining patients, 39,641 of whom were readmitted within 30 days of discharge, resulting in a readmission rate of 5.49%.

For certain pre-operative lab variables, more than 80% of the entries were missing, and they were excluded from the study. For other variables which had missing data, we used a statistical method that uses k-nearest neighbors and clustering to find the most likely value for a missing value.[16] The variables were then further separated into two classes: pre-operative variables and post-operative

variables. Pre-operative variables are those that can be reliably known before the main surgical procedure starts, whereas post-operative variables (including complications) can only be known after the surgery has occurred. Variable scaling was used for all models, except for Optimal Classification Trees, to bring all values into the range [0,1]; specifically, all variables were normalized by subtracting the minimum and dividing by the range.

Predictive Methods. Two different classes of machine learning methods were used. The first class consists of *predictive* methods used to accurately predict the readmission outcome of a patient. Two different scenarios were evaluated: (i) predicting readmissions using pre-operative variables, and (ii) predicting readmissions using both pre-operative and post-operative variables.

We tested a variety of machine learning methods, including: Random Forests (RF),[17] Logistic Regression (LR), Support Vector Machines (SVM), Gradient Boosted Machines (GBM),[18] and Neural Networks (NN).[19] Logistic Regression aims to fit a regression between the features and the binary outcomes in the logit space. In this study, we fit the logistic regression with a L2-norm regularization term to induce robustness and help guard against data corruption.[20] Both RF and GBM assemble a large collection of classification trees that classifies by taking a majority vote of the individual trees. Linear SVM aims to find a separating hyperplane in the feature space to best separate the patients which were readmitted and those who were not. We implemented a variant of this algorithm, Sparse Linear SVM (SLSVM, see Appendix S1 for details), that chooses a sparse number of variables in the separating hyperplane. By only allowing the hyperplane to depend on a small number of features, we can understand what are the most important variables that separate those that are readmitted and those that are not, which improves interpretability.

To evaluate prediction quality, one typically considers two distinct performance metrics computed out-of-sample: the *false positive rate* (or one minus the *specificity* of the test) and the *true positive rate*, or *sensitivity* of the test. A Receiver Operating Characteristic (ROC) curve evaluates the performance of a binary classifier as the decision threshold is varied, and is formed by plotting the true positive rate against the false positive rate at different threshold settings. To have a single metric to compare different ROC curves, we will consider the Area Under the ROC Curve (AUC). An ideal prediction model has an AUC close to 1, whereas a random prediction would yield an AUC of 0.5.

Prescriptive Methods. The second class of methods we employ consists of *prescriptive* methods. We focus on the pre-operative hematocrit (HCT) and seek to modulate it in order to minimize the readmission rate. Operationally, this is achieved through blood transfusion before the surgery. Consistent with medical practice, the maximum change in the hematocrit level is limited to 9%, corresponding, roughly, to 3 standard (300cc) bags of blood, which can be considered as a safe upper limit for blood transfusion. Since any such intervention has to be applied before the surgery, the methods we develop will only use pre-operative variables (a total of 88 such variables in the dataset).

We introduce two methods we developed for this study: Prescriptive Support Vector Machines (P-SVM), and Optimal Prescriptive Trees (OPT). They seek to minimize a certain loss function over a set of "actionable" variables that can be controlled through treatments. In this study, the actionable variable is hematocrit, and there are only 4 treatments available: 0%, 3%, 6% and 9% increase in hematocrit, corresponding to 0 to 3 bags of blood transfused. The loss we aim to minimize is the readmission rate.

Before we present our prescriptive methods, we need to establish a baseline for the actionable variable under all treatments. This information would be used by one of our methods (OPT) to learn an effective treatment. In the NSQIP data, we utilize the TRANSFUS variable which indicates whether a pre-operative blood transfusion took place. However, there is no information on the amount of blood transfused. We formed our baseline treatment with the assumption that everyone who has a hematocrit value over 30 had at most 1 bag of blood transfused, as the common operative transfusion threshold is 30.[21] Then, we add additional bags of blood with decreasing hematocrit levels to bring the patient's hematocrit level above 30. The full table of assumed baseline treatment is shown in Table 1.

Table 1. Assumed baseline treatment.

Assumed Transfusion Factuals	**Condition in Data**
No blood transfusion	TRANSFUS=0
1 bag of blood	HCT>30 and TRANSFUS=1
2 bags of blood	27<HCT<30 and TRANSFUS=1
3 bags of blood	HCT<27 and TRANSFUS=1

The effect of the treatment suggested by our methods will be evaluated using several predictive methods discussed earlier. We rely on four different methods in order to ensure the stability of the result. Specifically, we will use: Random Forests (RF), Logistic Regression (LR), Gradient Boosted Trees (GBM), and Neural Networks (NN). There is evidence in the literature to suggest that pre-operative transfusion could potentially lead to adverse outcomes.[22–24] To ensure such effect is properly accounted for, we additionally consider second order effects of blood transfusion on other pre-operative variables other than hematocrit. Specifically, we fit a regression model of other variables on HCT, and then, calculate how they are affected by the transfusion. We use the modified variables to predict the final readmission outcome.

Prescriptive Support Vector Machines (P-SVM)[25] is an interpretable prescriptive method based on the interpretable SLSVM predictive method we discussed earlier. The method first trains a SLSVM to obtain a hyperplane in a sparse variable subspace that separates readmitted from non-readmitted patients. Fixing this hyperplane, a second optimization problem is formulated, seeking to select the value of the actionable variable (HCT) in order to minimize over the training set a linear combination of the readmission rate and a penalty for changes in the actionable variable. Essentially, this optimization problem sets a value of HCT for each readmitted patient in a way that balances the number of prevented readmissions with the percentage of HCT increase required to prevent them. A detailed mathematical formulation of the method is provided in Appendix S1.

Optimal Prescriptive Trees (OPT) is an interpretable prescriptive method based on Optimal Classification Trees (OCT). OCTs,[26] use integer programming to build a decision tree that optimizes the accuracy of predictions over the training set. A decision tree is interpretable because at each node we are only making a binary decision based on one feature, so the final decision is based upon a series of simple binary decisions. Such a tree, assigns each patient to a leaf node of the tree and makes a prediction for the patient by a majority vote of other patients assigned to the same leaf. OPTs similarly builds an optimal decision tree but with a modified objective, a linear combination of prediction accuracy and the readmission rate. A more detailed mathematical formulation of OPT is in Appendix S2.

Methods were evaluated in Python, Matlab, and Julia. For random forests, the number of trees grown was 500. Cross-validation was used to tune parameters of the methods.

3. Results

Sample Characteristics. For each patient, a total of 231 variables were extracted. Table 2 summarizes the baseline demographic and clinical characteristics of the 722,101 patients included in the study. We report the (unnormalized) mean values of the variables over all patients, readmitted patients, and non-readmitted patients, respectively, and only list 60 variables for which the difference between readmitted and non-readmitted patients was the most statistically significant. Specifically, for each variable we computed a two-tailed p-value using Welch's t-test, where the null hypothesis was that the two cohorts (readmitted and non-readmitted patients) have equal means. Hence, the smaller the p-value, the less likely it becomes that the variable means listed in Table 2 occurred by chance under the null hypothesis. We note that for indicator variables, the means reported correspond to the fraction of patients satisfying the condition.

Table 2. Most statistically significant differences in readmitted and non-readmitted patients.

Variable	All patients	Readmitted	Non-Readmitted	p-value
Estimated Probability of Morbidity	0.06	0.11	0.06	<0.000001
Pre-operative hematocrit	39.67	37.85	39.78	<0.000001
The American Society of Anesthesiology (ASA) Physical Status Classification	2.43	2.78	2.4	<0.000001
Estimated Probability of Mortality	0.01	0.02	0.01	<0.000001
Total operation time in minutes	111.31	148.79	109.14	<0.000001
Return to OR (binary)	0.03	0.24	0.02	<0.000001
Number of Superficial Wound Occurrences	0.02	0.08	0.01	<0.000001
Number of Deep Incisional SSI Occurrences	0.01	0.06	0	<0.000001
Number of Organ/Space SSI Occurrences	0.01	0.11	0.01	<0.000001
Number of Urinary Tract infection Occurrences	0.01	0.06	0.01	<0.000001
Number of Bleeding Transfusions Occurrences	0.06	0.13	0.05	<0.000001
Number of Sepsis Occurrences	0.02	0.1	0.01	<0.000001
Days from Operation to Discharge	2.77	4.51	2.67	<0.000001
OUTPATIENT (if surgical procedure was performed in an outpatient setting)	0.4	0.18	0.42	<0.000001
CPT_Muscl_29x: Casts and endoscopy/arthroscopy	0.03	0.01	0.03	<0.000001
Indicator for any morbidity/complications	0.12	0.49	0.1	<0.000001
no diagnosis of diabetes or diabetes controlled by diet alone.	0.85	0.77	0.85	<0.000001
Discharge Destination: Home	0.9	0.82	0.91	<0.000001

Pre-operative alkaline phosphatase	69.37	82.75	68.59	<0.000001
Pre-operative serum albumin	3.95	3.78	3.96	<0.000001
ICD9 550： Inguinal hernia	0.04	0.01	0.04	<0.000001
Work Relative Value Unit (a metric of surgical complexity)	16.35	19.75	16.16	<0.000001
Age	56.41	60.42	56.17	<0.000001
Hypertension requiring medication	0.45	0.57	0.44	<0.000001
Elective Surgery (binary)	0.8	0.69	0.81	<0.000001
Number of Pneumonia Occurrences	0.01	0.05	0.01	<0.000001
Bleeding disorders	0.04	0.09	0.04	<0.000001
Open wound/wound infection	0.03	0.07	0.03	<0.000001
Number of DVT/Thrombophlebitis Occurrences	0.01	0.04	0	<0.000001
CPT_Muscl_23x-25x: Shoulder arm wrist hand	0.03	0.01	0.03	<0.000001
CPT_CAT_2x: Musculoskeletal system	0.22	0.16	0.23	<0.000001
Discharge Destination: Skilled Care Not Home	0.06	0.11	0.05	<0.000001
Pre-operative serum creatinine	0.99	1.18	0.97	<0.000001
Pre-operative BUN	16.32	18.2	16.21	<0.000001
CPT_CAT_33x-37x: Cardiovascular system	0.07	0.12	0.06	<0.000001
Number of Pulmonary Embolism Occurrences	0	0.03	0	<0.000001
History of severe COPD	0.04	0.09	0.04	<0.000001
Disseminated cancer	0.02	0.06	0.02	<0.000001
Number of Wound Disruption Occurrences	0	0.03	0	<0.000001
Functional health status Prior to Surgery	0.03	0.07	0.03	<0.000001
Pre-operative serum sodium	138.82	138.46	138.87	<0.000001
Steroid use for chronic condition	0.04	0.07	0.03	<0.000001
Currently on dialysis (pre-op)	0.01	0.04	0.01	<0.000001
TRANST_Not transferred (admitted from home)	0.96	0.93	0.96	<0.000001
CPT_Digestive_441x: Intestines - excision	0.02	0.05	0.02	<0.000001
Number of Septic Shock Occurrences	0.01	0.03	0	<0.000001
Wound classification 4: Dirty/Infected	0.05	0.08	0.05	<0.000001
Surgical Specialty: Gynecology	0.07	0.05	0.08	<0.000001
CPT_Cardio_35x: Repairs bypasses etc.	0.04	0.07	0.03	<0.000001
Organ/Space SSI PATOS (Present at the Time of Surgery)	0	0.02	0	<0.000001
CPT_Digestive_48x: Pancreas	0.01	0.03	0.01	<0.000001
CPT_Digestive_49x: Abdomen Peritoneum and Omentum	0.11	0.08	0.12	<0.000001
No dyspnea	0.05	0.08	0.05	<0.000001
Pre-operative International Normalized Ratio (INR) of PT (Prothrombin Time) values	1.07	1.1	1.07	<0.000001
CPT_CAT_60x: Endocrine system	0.03	0.02	0.03	<0.000001

Number of Progressive Renal Insufficiency Occurrences	0	0.02	0	<0.000001
Number of Myocardial Infarction Occurrences	0	0.02	0	<0.000001
CPT_CAT_4x: Digestive system	0.41	0.46	0.4	<0.000001
Number of Unplanned Intubation Occurrences	0.01	0.02	0.01	<0.000001
Discharge Destination: Rehab	0.03	0.05	0.03	<0.000001

Accuracy of predictions. For the predictive task, we evaluated the methods across three distinct splits of the data into a training and a test dataset. Each split, randomly selects 80% of the data to form the training set and keeps the remaining 20% as the test set, on which model performance is evaluated. The mean (Avg.) and standard deviation (Std.) of AUC for each predictive method is reported in Table 3; the top table considers predictions using only pre-operative (PRE-op) variables, while the bottom table evaluates models using pre-operative and post-operative variables (POST-op).

Table 3. Performance of predictive models.

PRE-op					
Methods	**Split I**	**Split II**	**Split III**	**Avg.**	**Std.**
L2LR	72.55%	72.61%	72.97%	72.71%	0.23%
SLSVM	72.51%	72.58%	72.91%	72.67%	0.21%
RF	73.39%	73.24%	73.59%	73.41%	0.18%
GBM	73.49%	73.51%	73.78%	73.59%	0.16%
NN	72.50%	72.74%	73.18%	72.81%	0.34%

POST-op					
Methods	**Split I**	**Split II**	**Split III**	**Avg.**	**Std.**
L2LR	84.20%	84.36%	84.64%	84.40%	0.22%
SLSVM	84.25%	84.38%	84.68%	84.44%	0.22%
RF	85.24%	85.34%	85.67%	85.41%	0.22%
GBM	87.06%	87.32%	87.80%	87.39%	0.38%
NN	83.03%	83.06%	84.00%	83.36%	0.55%

Table 3 suggests that readmission predictions are less accurate when using only pre-operative variables. Using all variables, predictions are very strong, achieving an average AUC above 87% (using GBM). Low standard deviations across different splits for all methods, imply that the predictive power is not greatly impacted by the choice of the training data subset. A subgroup analysis using only the general surgery class is contained in Appendix S3. The results show that models trained on only the subgroup achieved a lower AUC, which suggests that the models trained on the general dataset exhibits favorable cross-learning behavior.

Effectiveness of prescriptions. Predicting readmission is only one step toward preventing readmissions. For the prescriptive task as well, we evaluated the P-SVM and OPT methods across the 3 distinct splits of the data. For each split, we train P-SVM and OPT in the training set and then apply the method to obtain a recommended number of bags of blood to be transfused for each patient whose HCT is less than 30 (HCT<30) in the test set. We evaluate the outcome for each test patient using four different predictive methods: L2LR, RF, GBM, and NN. For each predictive model, we chose a threshold so that the predicted readmission rate equals the ground truth readmission rate in the training dataset. Such a threshold gives a specificity >96% for all of our models as shown in Appendix S4. To account for the effects of transfusion on other variables, we modify variables highly correlated with HCT (absolute value>0.1) for each test patient under transfusion using the regression model constructed against HCT. Such variables are: pre-op creatinine, international normalized ratio, prothrombin time, albumin, mortality probability (MORTPROB), and morbidity probability (MORBPROB). We used generalized linear regression models to predict the effect on MORTPROB and MORBPROB since they have bounded values (in [0,1]).

We report in Table 4 the percentage of readmissions prevented in the test set, defined as the ratio (in %) of (i) the number of patients with HCT<30 originally predicted to be readmitted (assuming no treatment) and now predicted not to be readmitted (after treatment), over (ii) the number of patients with HCT<30 predicted to be readmitted (assuming no treatment). We also report the average number of bags of blood per patient under the recommended treatment.

The first column of Table 4 lists the predictive models used to evaluate the effect of treatment, the 2nd and 4rd columns show the percentage of readmissions prevented using the OPT and P-SVM prescriptions, the 3th and 5th columns show the average number of bags per patient when using OPT

and P-SVM prescriptions, and the last column reports a baseline percentage of readmissions prevented, assuming any patient (with HCT<30) in the test set gets 1 bag of blood.

Table 4. Percentage reduction of readmissions due to increase in pre-operative hematocrit.

Split I	OPT	Average bags for patients (HCT<30)	PSVM	Average bags for patients (HCT<30)	decrease_1bag
LR	9.27%	0.97	9.49%	1.06	6.78%
RF	14.50%	0.97	14.03%	1.06	9.03%
GBM	4.81%	0.97	3.78%	1.06	3.19%
NN	16.37%	0.97	16.50%	1.06	8.03%

Split II	OPT	Average bags for patients (HCT<30)	PSVM	Average bags for patients (HCT<30)	decrease_1bag
LR	9.27%	0.95	9.63%	1.08	5.97%
RF	13.08%	0.95	11.30%	1.08	8.24%
GBM	5.34%	0.95	4.20%	1.08	2.84%
NN	18.96%	0.95	19.08%	1.08	9.22%

Split III	OPT	Average bags for patients (HCT<30)	PSVM	Average bags for patients (HCT<30)	decrease_1bag
LR	10.25%	0.98	10.65%	1.07	6.94%
RF	15.84%	0.98	14.60%	1.07	8.89%
GBM	8.66%	0.98	5.72%	1.07	4.41%
NN	19.51%	0.98	18.37%	1.07	8.46%

We observe that across the different ground truths and splits of the data, the two methods significantly decrease the readmitted patients, on average. For OPT, the average decrease across all splits is 12.15%,

while it is 11.45% for P-SVM. Moreover, the average needed bags is about 1 bag, which roughly corresponds to 300cc of blood.

4. Discussion

An analysis of the most statistically significant differences in readmitted vs. non-readmitted patients, reveals (cf. Table 2) that the former tend to be patients who underwent vascular surgery, or surgeries involving the pancreas. In contrast, surgeries involving the endocrine system, or Abdomen, Peritoneum, and Omentum are less likely to lead to a readmission. Furthermore, readmitted patients tend to have more complications (e.g., septic shock, bleeding, pneumonia, organ/space/deep incisional SSIs, renal insufficiency) and show higher incidence of return to OR and unplanned intubation.

The predictive models we tested lead to very accurate predictions of 30-day readmissions, exceeding 87% AUC with GBM when using all (pre-op and post-op) variables. Using only pre-operative variables, AUC is 74%, on average (with GBM). These results outperform earlier models, such as the LACE index,[27] which has an AUC of 68.4%, and more recent models,[28] which yield a 72% AUC in 2 days after admission, and 78-81% at discharge.

In terms of specific actionable interventions, we developed prescriptive methods based on the pre-operative predictive models and examined the potential of reducing readmissions by increasing pre-operative HCT levels. We have shown that across a wide variety of different ground truths, two separate prescriptive methods (OPT and P-SVM) are able to prescribe blood transfusion treatments that reduce predicted readmissions for patients with HCT<30, with the decrease ranging from 4.81% to 19.51% for OPT and 3.78% to 19.08% for P-SVM, and with transfusions in the range of 300cc of blood per patient on average. To put the achievable readmission reductions into context, if one could reduce by the mean percentage we achieved (12%) all 30-day readmissions of patients with HCT<30 across the U.S. (over 10,000 per year), the cost savings would amount to $20.3 million on an annual basis.[29]

A further potential use of our model is to decrease the length of hospital stay for patients with low-risk of readmission. We can choose a threshold for our models to have high specificity and thus able to accurately identify those that are at low risk of readmission (e.g., the threshold so that the predicted readmission rate equals the ground truth readmission rate in the training dataset as in Appendix S4).

Moreover, the machine learning methods employed are interpretable. Due to its sparse nature, P-SVM produces a small number of predictive variables which provide an explanation as to why a specific patient has been predicted as having a high readmission risk. For this particular study, P-SVM chose inpatient status, ASA Classification, and mortality probability as some of the variables with most explanatory power on readmission, corresponding to an intuitive understanding that patients that have more critical conditions going into the surgery are more likely to be readmitted. OPTs, additionally, offer the ability to examine every prediction or prescription and identify the specific path through the decision tree that led to the decision. Figure 1 depicts a part of a prescriptive tree. At every branching step, a binary decision is made. For example, at the top node, the patients are split into two classes, smoking or non-smoking. Several splits lead to the leaves of the tree (colored red and blue in this case) in which all patients corresponding to a leaf are prescribed a certain treatment. Prescription 1 corresponds to no transfusion and Prescription 2 corresponds to transfusing one bag of blood. Interpretable decision trees enable doctors and experts to understand the proposed decisions and, potentially, further improve the model based on the binary decisions it makes.

[Figure 1 goes here]

Figure 1: An instance of an Optimal Prescriptive Tree.

A limitation of our study is the lack of the full ground truth for the impact of the proposed interventions, namely, whether they prevent readmissions. That is impossible to ascertain without performing a randomized clinical trial. Instead, we use a number of strong predictive models to evaluate the impact of the derived prescriptions and observe consistent readmission rate reduction across these models. We hope that this work motivates clinical trials that could confirm our findings. Furthermore, beyond the short-term readmission outcome, it would be beneficial to understand how these measures can impact the long-term oncologic outcome, which is equally if not more important. However, the NSQIP database unfortunately does not track patients long term, and thus we do not have the required data to conduct such an analysis in the paper.

5. Conclusions

We leveraged a large national dataset of surgical patients with the goal of reducing 30-day readmissions. We developed both predictive and prescriptive machine learning models. The former predict 30-day readmissions and identify the most discriminative variables. The latter, build on the

predictive models and can offer specific recommendations on actionable decisions to reduce readmissions. We focused on pre-operative hematocrit and showed how to make personalized recommendations to increase its value, when needed, through blood transfusion.

Prediction accuracy with our methods exceeds 87% using the entire collection of NSQIP variables and 74% using only variables known pre-operatively. The proposed prescriptions/interventions can reduce the predicted readmissions by 11.45%-12.15% for patients with HCT<30, on average. Beyond improving patient outcomes, this reduction can lead to more than $20 million in annual savings in the U.S.

Acknowledgments

The authors would like to thank Dr. George Velmahos at the Massachusetts General Hospital and Dr. George Kasotakis at the Boston Medical Center for useful discussions and for providing access to the NSQIP data.

References

1. Squires D, Anderson C. US health care from a global perspective: spending, use of services, prices, and health in 13 countries. Issue Brief Commonw Fund. 2015;15: 1–15.

2. James J, others. Medicare hospital readmissions reduction program. Health Aff (Millwood). 2013;34: 1–5.

3. Readmissions Reduction Program. Centers for Medicare and Medicaid Services; 2012. Available: http://www.cms.gov/ Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/ Readmissions-Reduction-Program.html

4. Gonzalez AA, Cruz CG, Dev S, Osborne NH. Indication for Lower Extremity Revascularization and Hospital Profiling of Readmissions. Ann Vasc Surg. 2016;35: 130–137. doi:10.1016/j.avsg.2016.01.054

5. Tosoian JJ, Hicks CW, Cameron JL, Valero V, Eckhauser FE, Hirose K, et al. Tracking early readmission after pancreatectomy to index and nonindex institutions: a more accurate assessment of readmission. JAMA Surg. 2015;150: 152–158. doi:10.1001/jamasurg.2014.2346

6. Kimbrough CW, Agle SC, Scoggins CR, Martin RCG, Marvin MR, Davis EG, et al. Factors predictive of readmission after hepatic resection for hepatocellular carcinoma. Surgery. 2014;156: 1039–1046. doi:10.1016/j.surg.2014.06.057

7. Cai X, Perez-Concha O, Coiera E, Martin-Sanchez F, Day R, Roffe D, et al. Real-time prediction of mortality, readmission, and length of stay using electronic health record data. J Am Med Inform Assoc. 2016;23: 553–561. doi:10.1093/jamia/ocv110

8. Weller GB, Lovely J, Larson DW, Earnshaw BA, Huebner M. Leveraging electronic health records for predictive modeling of post-surgical complications. Stat Methods Med Res. 2018;27: 3271–3285. doi:10.1177/0962280217696115

9. Chakravarthy V, Ryan MJ, Jaffer A, Golden R, McClenton R, Kim J, et al. Efficacy of a Transition Clinic on Hospital Readmissions. Am J Med. 2017. doi:10.1016/j.amjmed.2017.08.037

10. Lasater KB, McHugh MD. Reducing Hospital Readmission Disparities of Older Black and White Adults After Elective Joint Replacement: The Role of Nurse Staffing. J Am Geriatr Soc. 2016;64: 2593–2598. doi:10.1111/jgs.14367

11. Vest JR, Kern LM, Silver MD, Kaushal R. The potential for community-based health information exchange systems to reduce hospital readmissions. J Am Med Inform Assoc. 2015;22: 435–442. doi:10.1136/amiajnl-2014-002760

12. Dai W, Brisimi TS, Adams WG, Mela T, Saligrama V, Paschalidis IC. Prediction of hospitalization due to heart diseases by supervised learning methods. Int J Med Inf. 2015;84: 189–197.

13. Brisimi TS, Xu T, Wang T, Dai W, Adams WG, Paschalidis IC. Predicting Chronic Disease Hospitalizations from Electronic Health Records: An Interpretable Classification Approach. Proc IEEE. 2018;106: 690–707. doi:10.1109/JPROC.2017.2789319

14. Brisimi TS, Xu T, Wang T, Dai W, Paschalidis IC. Predicting diabetes-related hospitalizations based on electronic health records. Stat Methods Med Res. 2019;28: 3667–3682. doi:10.1177/0962280218810911

15. Ingraham AM, Richards KE, Hall BL, Ko CY. Quality Improvement in Surgery: the American College of Surgeons National Surgical Quality Improvement Program Approach. Adv Surg. 2010;44: 251–267. doi:10.1016/j.yasu.2010.05.003

16. Bertsimas D, Pawlowski C, Zhuo YD. From predictive methods to missing data imputation: an optimization approach. J Mach Learn Res. 2017;18: 7133–7171.

17. Breiman L. Random forests. Mach Learn. 2001;45: 5–32.

18. Friedman JH. Stochastic gradient boosting. Comput Stat Data Anal. 2002;38: 367–378.

19. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016.

20. Chen R, Paschalidis IC. A Robust Learning Approach for Regression Models Based on Distributionally Robust Optimization. J Mach Learn Res. 2018;19: 1–48.

21. Wang JK, Klein HG. Red blood cell transfusion in the treatment and management of anaemia: the search for the elusive transfusion trigger. Vox Sang. 2010;98: 2–11.

22. Whitlock EL, Kim H, Auerbach AD. Harms associated with single unit perioperative transfusion: retrospective population based analysis. bmj. 2015;350: h3037.

23. Hollis RH, Singletary BA, McMurtrie JT, Graham LA, Richman JS, Holcomb CN, et al. Blood transfusion and 30-day mortality in patients with coronary artery disease and anemia following noncardiac surgery. JAMA Surg. 2016;151: 139–145.

24. Goel R, Patel EU, Cushing MM, Frank SM, Ness PM, Takemoto CM, et al. Association of perioperative red blood cell transfusions with venous thromboembolism in a North American registry. JAMA Surg. 2018;153: 826–833.

25. Wang T, Paschalidis IC. Prescriptive Cluster-Dependent Support Vector Machines with an Application to Reducing Hospital Readmissions. European Control Conference. Napoli, Italy; 2019.

26. Bertsimas D, Dunn J. Optimal classification trees. Mach Learn. 2017;106: 1039–1082.

27. Walraven C van, Dhalla IA, Bell C, Etchells E, Stiell IG, Zarnke K, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. Can Med Assoc J. 2010;182: 551–557. doi:10.1503/cmaj.091117

28. Graham LA, Mull HJ, Wagner TH, Morris MS, Rosen AK, Richman JS, et al. Comparison of a Potential Hospital Quality Metric With Existing Metrics for Surgical Quality–Associated Readmission. JAMA Netw Open. 2019;2: e191313–e191313.

29. Bailey MK, Weiss AJ, Barrett ML, Jiang HJ. Characteristics of 30-day all-cause hospital readmissions, 2010-2016 [Statistical Brief# 248]. Rockv MD Agency Healthc Res Qual Retrieved Www Hcup-Us Ahrq Govreportsstatbriefssb248-Hosp-Readmissions-2010-2016 Jsp. 2019.

APPENDICES

APPENDIX S1

Prescriptive Support Vector Machines (P-SVM)

Predictions. Prescriptive Support Vector Machines (P-SVM) is a prescriptive method that is based on Sparse Linear SVM (SLSVM). To formulate the SLSVM problem, let $(x_i^+, y_i^+),\ i = 1, \dots, N^+$, denote the $(D+1)$ −dimensional positive samples, where x_i^+ is the D −dimensional vector of variables for sample i and $y_i^+ = 1$ the class label. Similarly, $(x_j^-, y_j^-),\ j = 1, \dots, N^-$, denote the negative samples (patients who are not re-admitted within 30 days) with $y_i^- = -1$. Let $(\boldsymbol{\beta}, \beta_0)$ be the vector orthogonal to the SVM hyperplane. Let also M be a parameter controlling the level of sparsity. Training a classifier amounts to selecting $(\boldsymbol{\beta}, \beta_0)$ so that the margin of the hyperplane is maximized:

$$min_{\boldsymbol{\beta},\beta_0} \frac{1}{2}\|\boldsymbol{\beta}\|^2 + \lambda^+ \sum_{i=1}^{N^+} \xi_i + \lambda^- \sum_{j=1}^{N^-} \zeta_j$$

$$s.t. \sum_{d=1}^{D} |\beta_d| \le M,$$

$$\xi_i \ge 1 - y_i^+ \beta_0 - \sum_{d=1}^{D} y_i^+ \beta_d x_{i,d}^+, \qquad \forall i = 1, \dots, N^+,$$

$$\zeta_j \ge 1 - y_j^- \beta_0 - \sum_{d=1}^{D} y_j^- \beta_d x_{j,d}^-, \qquad \forall j = 1, \dots, N^-,$$

$$\xi_i, \zeta_j \ge 0,\ \forall i = 1, \dots, N^+, j = 1, \dots, N^-.$$

This is a convex quadratic optimization problem and can be solved very efficiently for large training sets involving thousands of patients. Let $(\boldsymbol{\beta}, \beta_0)$ be an optimal solution of the problem above. Then, for a patient represented with a vector of variables $\boldsymbol{x}$ we compute $\beta_0 + \sum_{d=1}^{D} x_d \beta_d$ and compare it with some threshold. If this value is above the threshold, we predict that the patient will be re-admitted. Otherwise, we predict it will not. The threshold can be set using cross-validation given a desirable false positive probability.

Prescriptions. Fixing the hyperplane $(\boldsymbol{\beta}, \beta_0)$, we next consider each patient i in the training set and seek

to optimize the value of "actionable" variables $x_{i,d}^{+}$, for $d \in \mathrm{C}$, where C is the index set of actionable variables, so as to "flip" the patient to the negative side of the hyperplane. To that end, we solve the following convex optimization problem. The objective is a linear combination of a penalty for not placing the patient on the negative side of the hyperplane and a penalty for altering the values of the variables characterizing the patient:

$$min_{\xi,\boldsymbol{y}}\ \lambda\xi + ||\boldsymbol{y} - \boldsymbol{x}_i^{+}||_p^p$$

$$s.t.\ \xi - 1 \geq \beta_0 + \sum_{d=1}^{D} \beta_d y_d,$$

$$y_d = x_{i,d}^{+}, \qquad \forall d \notin \mathrm{C},$$

$$\xi \geq 0,$$

$$L_d \leq y_d \leq U_d,\ \forall d \in \mathrm{C},$$

where $||.||_p$denotes the p-norm, $L_d$and U_d are lower and upper bounds on the actionable variables, and λ is a parameter trading-off the two penalty terms in the objective. The parameter λ can be determined by validating the performance of the prescription determined by the above formulation in a validation dataset. After we fix λ, we can solve the above problem for each patient in the test set to determine the optimal value of the actionable variables.

APPENDIX S2

Optimal Prescriptive Trees (OPT)

We motivate and present the Optimal Prescription Tree (OPT) algorithm that trains prescriptive trees to directly minimize the personalization risk.

Personalization Risk. We consider data such that Y is the outcome for each patient, T the choice of treatment and $\boldsymbol{X}$ the feature vector. First, we establish the convention that the smaller the outcome the better. Hence, we would like to minimize the expected outcome $E[Y(\tau(\boldsymbol{X}))]$ with respect to a prescriptive rule $\tau(\boldsymbol{X})$. For a given dataset, the discretization of the expected value is thus:

$$R(\tau) = \sum_{i=1}^{n} Y_i\ 1[\tau(\boldsymbol{X}_i) = T_i],$$

where $1[\cdot]$ denotes the indicator function. We call $R(\tau)$ the *personalization risk*. However, in

observational data we only observe the outcome for the treatment that was assigned to the patient in the data, and do not know what the outcome would be if a different treatment were prescribed. This leads to the corrected expression for the personalization risk:

$$R(\tau) = \sum_{i=1}^{n} \left(Y_i \, 1[\tau(\boldsymbol{X}_i) = T_i] + \sum_{t \neq T_i} \hat{Y}_i(t) \, 1[\tau(X_i) = t] \right),$$

where $\hat{Y}_i(t)$ denotes the unknown counterfactual outcome that would have been observed if patient i were to be assigned treatment t.

Then, to further control for accuracy, we account for the quality of the counterfactual estimates. However, since we only know the true value of the outcome for one particular counterfactual, we will only control the quality of those. This leads to the squared loss term:

$$\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i(T_i) \right)^2,$$

where T_i is the treatment corresponding to the point $(\boldsymbol{X}_i, Y_i)$ in the data. Forming the linear combinations of these two terms, we obtain the final objective:

$$R(\tau) = \mu \sum_{i=1}^{n} \left(Y_i \, 1[\tau(\boldsymbol{X}_i) = T_i] + \sum_{t \neq T_i} \hat{Y}_i(t) \, 1[\tau(\boldsymbol{X}_i) = t] \right) + (1 - \mu) \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i(T_i) \right)^2,$$

where μ is the *prescription factor*, a hyperparameter that controls the trade-off between the prescription error and the prediction error. Minimizing this function above over τ is the basis of Optimal Prescriptive Trees.

Prescription Predictions. To minimize over τ, we seek a decision rule that takes the form of a prescriptive tree, that is, a decision tree that in each leaf prescribes a common treatment for all samples. Our approach is to estimate the counterfactual outcomes using this prescriptive tree during the training process, and therefore jointly optimize the counterfactual estimation and minimization of personalization risk.

Observe that a decision tree divides the training data into clusters where the samples are similar. We propose using these clusters as the basis for our counterfactual estimation. More concretely, we will estimate the counterfactual $\hat{Y}_i(t)$ using the outcomes Y_j for all samples j with $T_j = t$ that fall into the

same leaf of the tree as sample i. An immediate method for estimation is to simply use the mean outcome of the relevant samples in this cluster, giving the following expression for $\hat{Y}_i(t)$:

$$\hat{Y}_i(t) = \frac{1}{|\{j: \boldsymbol{X}_j \in \Lambda_{l(i)}, T_j = t\}|} \sum_{\{j: \boldsymbol{X}_j \in \Lambda_{l(i)}\}} Y_j,$$

where $\Lambda_{l(i)}$ represents the leaf of the prescription tree that $\boldsymbol{X}_i$ falls into. Then, using this expression, we want to find a prescriptive tree τ such that it solves the following problem:

$$R(\tau) = \mu \sum_{i=1}^{n} \left(Y_i \, 1[\tau(\boldsymbol{X}_i) = T_i] + \sum_{t \neq T_i} \frac{1}{|\{j: \boldsymbol{X}_j \in \Lambda_{l(i)}, T_j = t\}|} \sum_{\{j: X_j \in \Lambda_{l(i)}\}} Y_j \, 1[\tau(\boldsymbol{X}_i) = t] \right)$$
$$+ (1-\mu) \sum_{i=1}^{n} \left(Y_i - \frac{1}{|\{j: \boldsymbol{X}_j \in \Lambda_{l(i)}, T_j = T_i\}|} \sum_{\{j: \boldsymbol{X}_j \in \Lambda_{l(i)}\}} Y_j \right)^2.$$

We then use local-search methods to optimize the splits of the prescriptive trees over this objective, with each one starting from different random splits. Specifically, we initialize many prescriptive trees with random splits, and iteratively apply the following steps:

1. We randomly select a tree τ_i.
2. We randomly select a node n_{ij} within the selected tree τ_i.
3. We optimize the split at node n_{ij} keeping all other splits constant by minimizing the function above.
4. We return to Step 1 and repeat.

The process is completed when all nodes are individually optimized. Then, the best tree is selected through validation. The hyperparameter μ is also chosen through the validation set.

APPENDIX S3

To illustrate the behavior of our models under a subgroup analysis, we selected the largest subgroup, general surgeries, and retrained our models using only these surgeries. We further utilized the same splitting scheme for train/validation/test, and ran each model 3 times across 3 random splits. The average and standard deviation of the results across these 3 runs are listed in Table S1.

Table S1. Performance of predictive models on the subgroup of general surgeries.

	POST-op		**PRE-op**	

Methods	Avg.	Std.	Avg.	Std.
L2LR	82.91%	0.13%	72.42%	0.20%
SLSVM	82.89%	0.10%	72.38%	0.18%
RF	84.10%	0.12%	72.91%	0.28%
GBM	85.21%	0.21%	73.71%	0.34%
NN	81.90%	0.25%	72.56%	0.41%

APPENDIX S4

Table S2 reports the average specificity of our models across the three splits using the threshold such that the predicted readmission rate equals the ground truth readmission rate in the training dataset.

Table S2. Average specificity of the predictive models.

Methods	PRE-op specificity	POST-op specificity
L2LR	95.20%	96.71%
RF	95.17%	96.85%
SLSVM	95.19%	96.60%
GBM	95.30%	97.01%
NN	94.89%	96.80%