Readmission Risk after Orthopedic Surgery

Selah Lynch, Riley Wong, Yancy Lo, Jason H. Moore, Eric Hume
Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Institute for Biomedical Informatics | Perelman School of Medicine University of Pennsylvania
Clinical Research Informatics Core

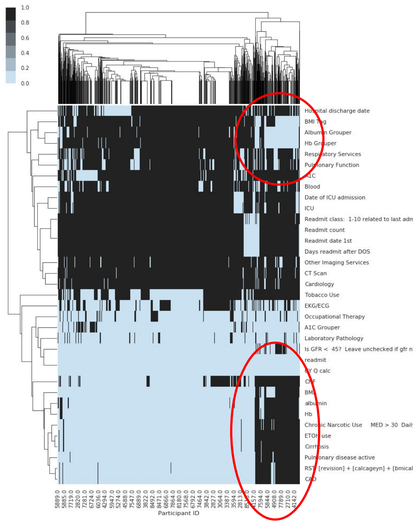
Goals

Can we improve upon the existing clinician model for predicting hospital readmission after joint replacement surgery?

Can we come up with potential interventions that might reduce hospital readmissions?

Cohort

Scheduled primary surgeries from 2014-2016.
Patients with < 75% missing data.
N = 3816
155(4.06%) readmits
2556(67%) training, 1260(33%) testing



Missingness of data, hierarchically clustered

Missing Data

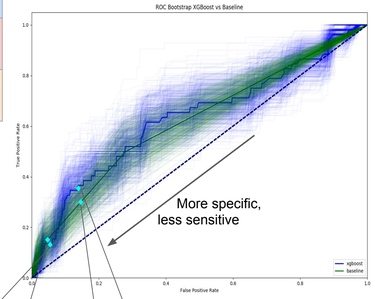
We had a lot of missing data and we used this visualization to examine patterns of missingness.

We noticed that some variables had been substituted for part of the patient population. From these, we constructed variables that would span the entire population.

Modeling

Model	Details	ROC AUC
Baseline	5 point scale:	0.593 (0.551, 0.637)
XGBoost	Max_depth: 1 Min child_weight:8	0.642 (0.561, 0.713)
Logistic Regression	C:0.251 Penalty: L1	0.638 (0.552, 0.718)
Random Forest	N_trees: 300 N_features: 20	0.615 (0.539, 0.690)

We tried multiple models. XGBoost had the best performance.



ROC Curve comparing xgboost to baseline, boostrapping overlayed

We compared our best model to the baseline model* that is currently used.

* 2*CHF + CAD + "GFR <= 45" + Dz_Pulmonary

XGBoost, follow 67 riskiest

	Actual Positive	Actual Negative	
Model Positive	7	60	67
Model Negative	45	1148	1193

Recall = 0.135
Precision = 0.104

Baseline, follow 67 riskiest

	Actual Positive	Actual Negative	
Model Positive	8	59	67
Model Negative	44	1149	1193

Recall = 0.154
Precision = 0.119

XGBoost, follow 189 riskiest

	Actual Positive	Actual Negative	
Model Positive	19	170	189
Model Negative	33	1038	1071

Recall = 0.365
Precision = 0.100

Baseline, follow 189 riskiest

	Actual Positive	Actual Negative	
Model Positive	12	177	189
Model Negative	40	1031	1071

Recall = 0.269
Precision = 0.074

Our model performance is better, but not overwhelmingly so. Its main benefit is the flexibility of its sensitivity.

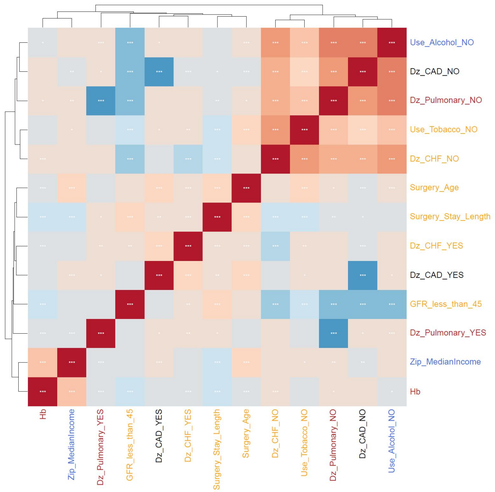
Feature Analysis

Baseline Features

- **Dz_CHF**
  - diagnosis of congestive heart failure
- **Dz_CAD**
  - diagnosis of coronary artery disease
- **GFR<45**
  - Glomerular filtration rate, a measure of kidney function
- **Dz_Pulmonary**
  - diagnosis of lung disease

Our New Features

- **Hb**
  - hemoglobin level from lab test
- **Albumin**
  - level from lab test
- **Surgery_Age**
  - Patient's age at the time of surgery
- **Surgery_Stay_Length**
  - Length of stay, in days
- **Use_Tobacco_NO**
  - The patient does not use tobacco



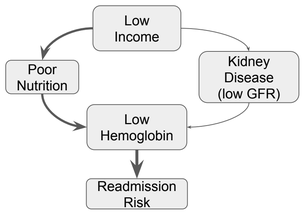
Predictive ML features correlation matrix, hierarchically clustered

We have confirmed existing knowledge such as that hospital readmission are associated with things such as heart conditions and age of patient.

We have discovered a new hypothesis that hospital readmissions may be associated with poor nutrition.

	Hemoglobin	GFR<45
GFR<45	-0.119 (<0.001)	
ZipIncome	0.245 (<0.001)	-0.038 (0.019)

Pearson correlation coefficient and p values between factors. The correlation between Hemoglobin and income is much stronger than what can be explained by kidney disease as measured by low GFR. Glomerular filtration rate (GFR) is an indicator of kidney health.



Hypothetical causality diagram of readmission risk