Machine Learning Generated Risk Model to Predict Unplanned Hospital Admission in Heart Failure

O'Donovan F[1], Brecht T[1], Kekeh C[1], Su Z[1], Boussios C[1], Menon V[1], Gliklich R[1], Fonarow GC[2]

[1]OM1, Inc, Cambridge, MA 02138; [2]Division of Cardiology, David Geffen School of Medicine at UCLA, Los Angeles, California 90095

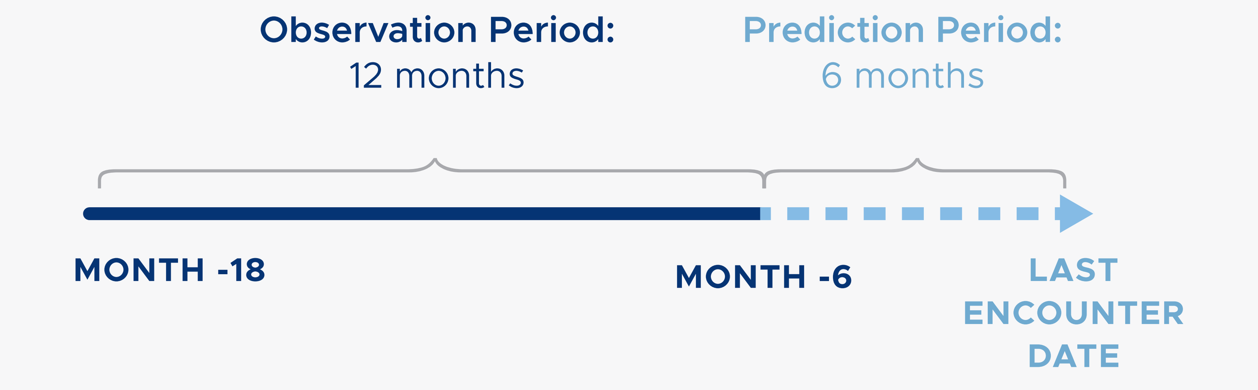
OM1

Introduction

Heart failure (HF) is a leading cause of hospitalization. There are few tools to accurately identify patients at high risk for unplanned admission in the outpatient setting. We used machine learning (ML) on outpatient electronic medical records and medical claims to develop a HF specific predictive model.

Methods

The OM1™ Cardiology data warehouse contains deep clinical and claims data on patients seen in cardiology practices across the US. HF Patients with at least 18 months of data (July 2013 to December 2016) were included, with the last 6 months serving as the prediction period. The outcome was unplanned admission due to HF during the prediction period.



REFERENCES: [1] Excess Days in Acute Care after Hospitalization for Acute Myocardial Infarction (AMI) (Version 1.1), submitted by Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation. Prepared for Centers for Medicare & Medicaid Services (CMS), March 2016.

DISCLOSURE INFORMATION: The authors have no conflicts of interest to disclose.

Methods (continued)

ML methods (Random Forests, Xgboost, and Treenet) were used to examine the association between predictive variables collected in the initial observation period and the binary outcome (unplanned admission versus no unplanned admission) and to identify high value predictors. We followed the definition of unplanned admission as described by Centers for Medicare & Medicaid Services [1]. To facilitate interpretation and comparison with literature, important predictors identified via ML were entered into a multivariable logistic regression model to derive HF risk scores. Patients were randomly assigned to training (70%) and validation (30%) sets.

One of the key predictive features is the OM1 Medical Burden Index, which is a standardized measure of the combined effect of current and prior conditions and treatments on current health status, on 0-1000 scale. It has been generated from extensive analysis of OM1's longitudinal patient cohort (n>175M).

Results

A total 48,761 patients with 2,770 unplanned admissions were included in the analysis; median age was 71 years, 49% were men, and 54% were white (**Table 1**). The top 5 predictors from the training set, as determined by machine learning, were OM1 medical burden index (ML derived scores on a 0-1000 scale, leveraging all available prior history), number of services in the previous year, number of hospitalizations in the previous year, time since last hospitalization, and Charlson comorbidity. These 5 predictors were used to derive HF risk scores in the validation set of 14,546 patients (**Table 2**). Using the OM1 medical burden index as the single predictor, the model had a C statistic of 0.72. Adding the other top 4 predictors increased the C statistic to 0.78. Using all available data only slightly improved the C statistic to 0.79. All C statistics reported here were from the validation set (**Figure 1**).

Results (continued)

In the validation set, 825 (5.7%) were hospitalized during the 6-month prediction period. Using the risk score as a single predictor, our model correctly predicted outcomes for 12,189 (84%) patients. All top 5 predictors identified by machine learning were statistically significant in the regression model (**Figure 2**). The odds ratios in Figure 2 represent one unit increase of OM1 medical burden index on a log scale, every 20 services in the past 12 months, time since last hospitalization in quarters, and Charlson comorbidity index. Fixing specificity at 50%, the machine learning model had high sensitivity (85%) in predicting unplanned hospitalization in the next 6 months (**Figure 3**).

Table 1. Patient characteristics during the 12-month observation period, stratified by the admission status during the 6-month prediction period

Patient Characteristics		Admission N=2,770	No admission N=45,991	Total N=48,761
Gender n (%)	Female	1,441 (52%)	23,385 (51%)	24,826 (51%)
	Male	1,329 (48%)	22,606 (49%)	23,935 (49%)
Age (years)	Mean (SD)	68 (13)	70 (12)	70 (12)
	Median (Q1-Q3)	69 (69-78)	71 (62-79)	71 (62-79)
OM1 medical burden index	Mean (SD)	24 (42)	10 (23)	11 (25)
	Median (Q1-Q3)	13 (5-27)	4 (1-11)	5 (1-11)
Charlson comorbidity index	Mean (SD)	3.9 (2.2)	2.7 (1.8)	2.8 (1.9)
	Median (Q1-Q3)	4 (2-5)	2 (1-4)	2 (1-4)

Table 2. The OM1 medical burden index is the single most predictive variable in univariate analysis

Variable name	C statistic[1]	Odds ratio	95% CI
OM1 medical burden index	0.72	1.14	1.08-1.21
Number of services in the past 12 months	0.71	1.04	1.02-1.06
Number of hospitalizations in the past 12 months	0.70	1.22	1.18-1.25
Time since last hospitalization	0.70	1.16	1.05-1.31
Charlson comorbidity index	0.65	1.08	1.03-1.12

[1] With each variable included as the only predictor in the model

Figure 1. Addition of other machine learning-generated predictive features to the OM1 medical burden index improved the predictive performance (C statistic: 0.72 to 0.79)

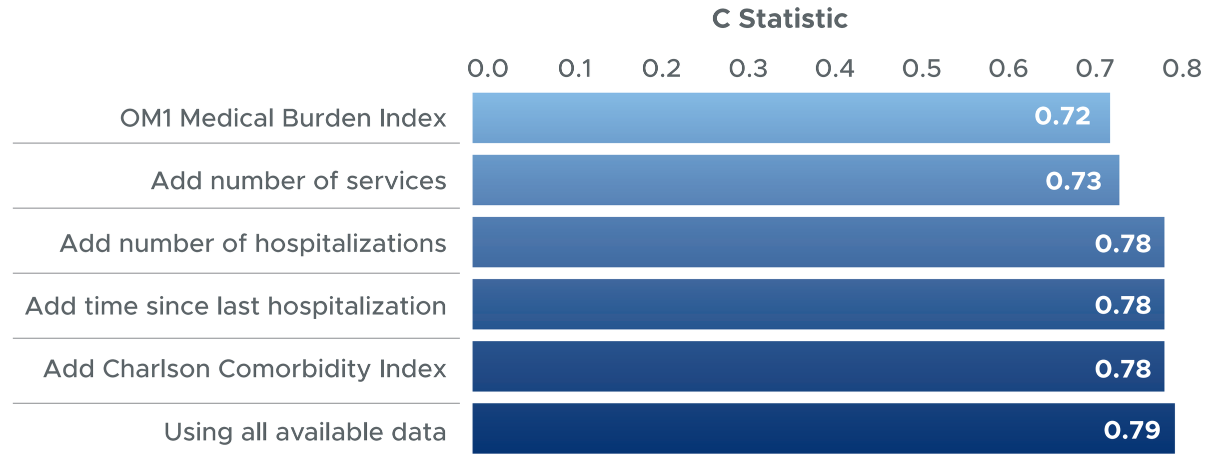


Figure 2. Odds ratios for the most influential predictive features in multivariable analysis

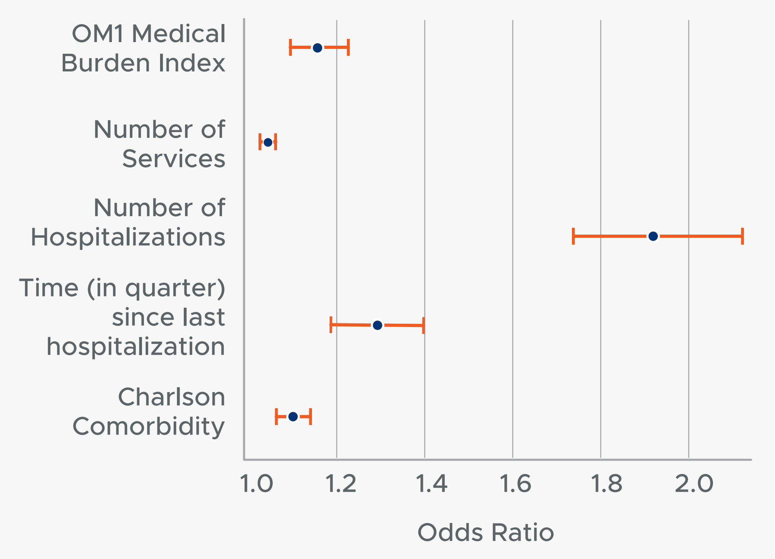


Figure 3. Receiver operating characteristics of the machine learning model (C statistic: 0.79)

Conclusions

We demonstrated the utility of machine learning in leveraging variables readily available in an outpatient EMR and medical claims to predict hospitalizations in 8 out of 10 patients (C Statistic: 0.79). When integrated into the clinical workflow, such tools may offer the ability to focus resources on patients at highest risk for unplanned admission.