A Framework for Evaluating Healthcare Machine Learning Models: Application and Analysis Using Hospital Readmission

Presenting Author: Eilon Gabel, MD, Assistant Professor, Department of Anesthesiology and Perioperative Medicine, University of California Los Angeles

Co-Authors: Velibor V. Mišić, PhD, Assistant Professor, Kumar Rajaram, PhD, Professor, Anderson School of Management, University of California Los Angeles

Introduction: Most predictive algorithms requiring pathway implementations are evaluated using metrics focused on predictive performance, such as the *c* statistic/AUC (area under the curve). However, these metrics are of limited clinical value since they do not account for the algorithm's role within a provider's limited workflow. We propose a framework for simulating a fixed clinical resource using machine learning algorithms to predict unplanned emergency department (ED) surgical readmissions.

Methods: We extracted de-identified data corresponding to all surgical patients that were admitted at the UCLA Ronald Reagan Medical Center in 2017 and 2018 and underwent a procedure with anesthesia. The input to our simulation framework was a predictive model. We considered: lab-based (AUC: 0.85) and non-lab-based (AUC: 0.73) L1 regularized logistic regression models from Mišić et al, HOSPITAL score (AUC 0.72) and the LACE score (AUC: 0.74).[1-3]
We simulated a single provider following a weekly schedule with a limit on the number of patients that can be seen (*daily capacity)*. Secondly, the provider cannot was unable to engage a patient outside of an availability window, i.e. L1 models predict at 36 hours post-operatively while HOSPTAL and LACE require day-of-discharge data. Unseen patients were carried over to the subsequent day until the patient was discharged or seen. Lastly, as part of the cost model, an *effectiveness constant* was estimated to indicate what fraction of eventual readmissions a provider was likely to prevent.
Post-operative readmission costs were calculated using data from the Healthcare Cost and Utilization Project (HCUP).[4] Provider costs were derived from the public salaries for mid-level nurse practitioners at UCLA.

Results: Summary of predictive results for provider-seen patients is found in table 1. The difference between the lab-based and the non-lab-based model is driven by the difference in their predictive ability AUC 0.8541 vs. 0.7280. The difference between the non-lab-based model versus HOSPITAL and LACE, however, is not due to predictive ability since all achieve AUC in the range 0.71 to 0.74. The difference arises because of the provider's schedule and the patients' availability windows for the predictions.
Account of the prevented readmission savings and provider costs were calculated with similar schedules as Table 1. The break-even effectiveness constant for each of the models was calculated. In a minimal Monday-only schedule, the lab-based L1 logistic regression model leads to positive net cost savings when the effectiveness constant is 2%, where HOSPITAL and LACE needs to be 5.5% - 6%. Alternately, when increasing the provider schedule to all weekdays, the break-even effectiveness was 4.5% comparted to 6.5%-7.5%.

Discussion: We proposed a new framework to demonstrate the rich clinical and administrative value of machine learning models that is not captured by gold standard metrics for predictive performance, such as AUC. Factors such as the provider's schedule and postoperative prediction timing can have major effects on the pathway cohort size and potential cost reductions from preventing hospital readmissions.

References:
1. Mišić V V, Gabel E, Hofer I, Rajaram K, Mahajan A. Machine Learning Prediction of Postoperative Emergency Department Hospital Readmission. Anesthesiology 2020;132:968–80.
2. Donzé J, Aujesky D, Williams D, Schnipper JL. Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients. JAMA Intern Med 2013;173:632.
3. Walraven C van, Wong J, Forster AJ. LACE+ index: extension of a validated index to predict early death or urgent readmission after hospital discharge using administrative data. Open Med 2012;6:e80-90.
4. Bailey MK, Weiss AJ, Barrett ML, Jiang HJ. Statistical Brief# 248: Characteristics of 30-day all-cause hospital readmissions, 2010-2016. Rockville, MD Agency Healthc Res Qual Retrieved from www.hcup-us.ahrq.gov/reports/statbriefs/sb248-Hospital-Readmissions-2010-2016.jsp 2019.

Table 1

Schedule	Daily Capacity	Method	Patients Seen	Anticipated Readmission	Prevented with 10% Effective Constant
M	8	L1 LR (with labs)	867	260	26
M	8	L1 LR (no labs)	861	160	16
M	8	HOSPITAL (at discharge)	845	86	9
M	8	LACE (at discharge)	845	91	9
M W	8	L1 LR (with labs)	1705	423	42
M W	8	L1 LR (no labs)	1699	265	27
M W	8	HOSPITAL (at discharge)	1688	172	17
M W	8	LACE (at discharge)	1688	177	18
M T W R F	8	L1 LR (with labs)	4201	677	68
M T W R F	8	L1 LR (no labs)	4197	505	51
M T W R F	8	HOSPITAL (at discharge)	4213	433	43
M T W R F	8	LACE (at discharge)	4213	451	45