

Development and Implementation of a Real-Time 30-Day Readmission Predictive Model

Patrick R. Cronin, MA¹, Jeffrey L. Greenwald, MD², Gwen C. Crevensten, MD², Henry C. Chueh, MD, MS¹, Adrian H. Zai, MD, PhD¹

¹Laboratory of Computer Science and ²Department of Medicine
Massachusetts General Hospital, Boston, MA

Abstract

Hospitals are under great pressure to reduce readmissions of patients. Being able to reliably predict patients at increased risk for rehospitalization would allow for tailored interventions to be offered to them. This requires the creation of a functional predictive model specifically designed to support real-time clinical operations. A predictive model for readmissions within 30 days of discharge was developed using retrospective data from 45,924 MGH admissions between 2/1/2012 and 1/31/2013 only including factors that would be available by the day after admission. It was then validated prospectively in a real-time implementation for 3,074 MGH admissions between 10/1/2013 and 10/31/2013. The model developed retrospectively had an AUC of 0.705 with good calibration. The real-time implementation had an AUC of 0.671 although the model was overestimating readmission risk. A moderately discriminative real-time 30-day readmission predictive model can be developed and implemented in a large academic hospital.

Keywords: Predictive Modeling, 30-Day Readmissions, Real-Time, Readmission Risk

Introduction

The affordable care act of 2010 established the Hospital Readmissions Reduction Program, requiring the Centers for Medicare and Medicaid Services to reduce payments to hospitals if the cost of readmission rates at a given hospital exceeded predicted costs¹. As a result, hospitals, academic groups and independent private organizations have invested substantial resources to reduce readmissions². One area of investment is the identification of patients at high risk for rehospitalization in the 30-days after an index hospitalization in order to enable targeted resource allocation to this population.

Identification of patients at high risk for early readmission using predictive modeling has been well documented in the literature. A 2011 systematic review of 26 validated readmission prediction models concluded that most had poor predictive ability but may be useful in certain settings³. Importantly, 23 models could not be implemented early during a hospitalization (i.e. in real-time) because data used in the prediction model was not available until after discharge or not available in a structured form. Three models were identified as usable in real-time⁴⁻⁶; however, two were predicting 12-month readmission^{5,6}, and the other was developed on a small population, 1029 patients with congestive heart failure⁴. Other early readmission prediction models were developed for cardiovascular events⁷⁻¹², pancreatitis¹³, kidney transplant¹⁴, and to identify avoidable early readmissions,¹⁵ though there is no agreement in the literature on exactly how to define “avoidable.” More recently developed early readmission models have improved discrimination, and two reported AUCs of 0.75¹⁶ and 0.77¹⁷. However, both models required

data that was not available in real-time at the MGH. We did not find any published 30-day real-time readmission prediction models that could be feasibly implemented at our hospital.

As such, we desired to create a model that included elements that would be commonly available on patients admitted to the hospital at the beginning of the admission and accessible electronically in real-time so that risk identification could begin prospectively early in the hospitalization. We defined *real-time* as being able to calculate the 30-day readmission risk the day after the patient was admitted.

Methods

This project was completed at the Massachusetts General Hospital (MGH). The MGH is an academic hospital with 957 licensed beds with approximately 48,000 admissions and 95,000+ emergency visits annually. Data for this project was extracted from existing databases maintained by the MGH for operational and research purposes. We extracted 45,924 hospital admissions for patients who arrived between 2/1/2012 and 1/31/2013; of these, 5,570 (12.1%) were readmissions to MGH within 30 days of discharge. This data was merged with inpatient transactional data, emergency department historical data, billing data, laboratory tests, medication orders, and outpatient appointment history. Data was split 80:20 for developing and validating the predictive model.

Readmissions were identified by the presence of an inpatient record, with a subsequent admission date by the same patient (identified by Medical Record Number) within 30 days of the index admission discharge date. In the event that there were multiple encounters within 30 days of discharge, the readmission evaluated is the first one occurring after discharge. Any readmission can become an index admission if there is another encounter following which occurred within 30 days of the prior discharge. Please note, this only included readmissions to the MGH, and is not the same metric as the publicly reported readmission rates which include readmissions to other hospitals. Moreover, the following exclusions were applied to both the index admissions and the readmissions:

- *From the Index Admission and Readmissions:* patients discharged to rehabilitation and hospice
- *From the Index Admission Only:* discharge status of deceased, left against medical advice, transferred to another short term acute facility, discharged/transferred to a psychiatric hospital
- *From the Readmission Only:* Chemotherapy, radiation, dialysis, Obstetrics (birth/delivery)

We performed a literature review to identify predictors from published models. Then we reviewed the existing data infrastructure at the MGH to determine which variables were available and the “lag” on their availability for a real-time implementation. Two hospitalists and another physician were consulted regularly to provide feedback to identify which variables made clinical sense to be included in the model. The variables were then weighted based on their availability in real-time. We identified 40 variables, and used logistic regression to develop our predictive model. Variables that were not statistically significant or that did not meaningfully

improve the model's discrimination or calibration were removed from consideration. We proceeded iteratively with close consultation with three physicians. We purposely did not use a formal backwards elimination process because our goal was to implement a real-time model and factors other than statistical significance (i.e. resources required to access data in real-time) were essential to our predictor selection process. In addition, we aimed to maintain flexibility to perform data transformations to improve discrimination, calibration, and feasibility of implementation.

The two statistics used to measure the model's performance were the area under the ROC curve (AUC) and the calibration. We calculated the calibration by splitting the data into ten groups of lowest-to-highest risk and plotting the expected-versus-observed outcomes.

After we validated our model, we implemented it in TopCare¹⁸, our hospital's population health management system. Using the developed model, we prospectively calculated the AUC, the calibration, and the 30-day readmission risk probability for all admissions between 10/1/2013 and 10/31/2013.

Results

Based on the data available one year prior to the admission at MGH, we split the derivation dataset into the following categories:

- No History: No prior admissions and no coded Elixhauser Comorbidities¹⁹ from outpatient data
- Has Comorbidity: No prior admissions but at least 1 Elixhauser Comorbidity from outpatient data
- Has Inpatient History: Had at least 1 prior admission to the MGH

We developed three separate predictive models with this data and merged the results back together to calculate the AUC and calibration. The following variables were identified as potentially predictive but were not included because they did not meaningfully improve the final model: age, sex, race, language, insurance, presence of advanced directive, marital status, restraint use, means of arrival, count of emergency department visits, count of no-shows in the past year, count of prior admissions, testing positive for illegal drugs, warfarin, rituximab, oral hypoglycemic agents, oral antiplatelet agents, sotalol, oxycodone, fentanyl, hydrocodone, meperidine, morphine equivalent narcotic dose, sedative use, and presence of an emergency department psychiatric consult.

The following variables were included in the final model (Table 1): unplanned admission (Yes/No); admission source (i.e., physician's office, transfer from long-term care, transfer from rehabilitation or skilled nursing facility, ambulance, walk-in); had a notation of drug abuse (ICD-9 code) or homelessness (internal system flag) or had left against medical advice (identified from prior MGH discharges); count of Elixhauser comorbidities based on inpatient and outpatient ICD-9 codes in the past year; number of inpatient days at the MGH in the past year; insulin (Yes/No); antipsychotic (Yes/No); other medications(Yes/No) (furosemide \geq 40 mg, metolazone, lactulose, cyclosporine, or heparin \geq 3000 units). The list of medications was derived from a list of drugs related to adverse drug events²⁰ and supplemented with recommendations from the

Table 1: Comparison of Model Parameters in Retrospective and Prospective Validation Data sets

Variables	Retrospective Validation				Prospective Validation				
	N(%)		Readmit (%)		N(%)		Readmit (%)		
History with MGH (past year)									
None	3,583	(38%)	221	6.2%	1,347	(42%)	75	5.6%	
Outpatient Only	2,554	(27%)	221	8.7%	693	(22%)	53	7.6%	
Prior Admission	3,210	(34%)	666	20.7%	1,134	(36%)	165	14.6%	
Admission Source									
Physician	4,345	(46%)	355	8.2%	1,431	(45%)	96	6.7%	
Transfer	1,266	(14%)	135	10.7%	450	(14%)	48	10.7%	
Other	1,942	(21%)	294	15.1%	675	(21%)	76	11.3%	
Emergency Med Svc.	1,582	(17%)	276	17.4%	565	(18%)	68	12.0%	
SNF or Rehab	212	(2%)	48	22.6%	53	(2%)	5	9.4%	
Admission Type									
Planned	2,578	(28%)	181	7.0%	821	(26%)	46	5.6%	
Unplanned	6,769	(72%)	927	13.7%	2,353	(74%)	247	10.5%	
Behavioral Issue (Past Year)									
No	9,009	(96%)	1,027	11.4%	3,025	(95%)	274	9.1%	
Yes	338	(4%)	81	24.0%	149	(5%)	19	12.8%	
Elixhauser Comorbidity Count (past year data)									
0	3,949	(42%)	272	6.9%	1,517	(48%)	86	5.7%	
1	1,290	(14%)	111	8.6%	357	(11%)	31	8.7%	
2	948	(10%)	94	9.9%	295	(9%)	27	9.2%	
3	759	(8%)	109	14.4%	222	(7%)	24	10.8%	
4	619	(7%)	119	19.2%	190	(6%)	20	10.5%	
5	483	(5%)	87	18.0%	154	(5%)	22	14.3%	
6	393	(4%)	71	18.1%	128	(4%)	25	19.5%	
7	324	(3%)	79	24.4%	86	(3%)	18	20.9%	
8+	582	(6%)	166	28.5%	225	(7%)	40	17.8%	
Inpatient Bed Days (past year)									
0	6,137	(66%)	442	7.2%	2,040	(64%)	128	6.3%	
1-5	862	(9%)	122	14.2%	395	(12%)	35	8.9%	
6-10	719	(8%)	134	18.6%	241	(8%)	32	13.3%	
10-20	721	(8%)	154	21.4%	229	(7%)	45	19.7%	
20+	908	(10%)	256	28.2%	269	(8%)	53	19.7%	
Insulin									
No	6,951	(74%)	719	10.3%	2,423	(76%)	188	7.8%	
Yes	2,396	(26%)	389	16.2%	751	(24%)	105	14.0%	
Antipsychotic									
No	7,854	(84%)	877	11.2%	2,399	(76%)	218	9.1%	
Yes	1,493	(16%)	231	15.5%	775	(24%)	75	9.7%	
Other Medication									
No	7,830	(84%)	812	10.4%	2,798	(88%)	238	8.5%	
Yes	1,517	(16%)	296	19.5%	376	(12%)	55	14.6%	

physician collaborators. Please note, all aforementioned medications were ordered during the patient's index admission.

Using the training data set of 36,462 records, the overall model had an AUC of 0.705 (95% CI 0.697 to 0.713). Odds ratios were calculated for each of the sub-models in Figure 1 and the model factors are illustrated in Table 2. The logistic coefficients were applied to the validation data set of 9,325 records and the model had an AUC of 0.714 (95% CI 0.698 to 0.730) and the calibration is in Figure 2. The model was then implemented in real-time for 3,074 admissions between 10/1/2013 and 10/31/2013. It had an AUC of 0.671 and the calibration is in Figure 3. The breakdowns of the readmission rates by each model variable for the retrospective and prospective validation is in table 1.

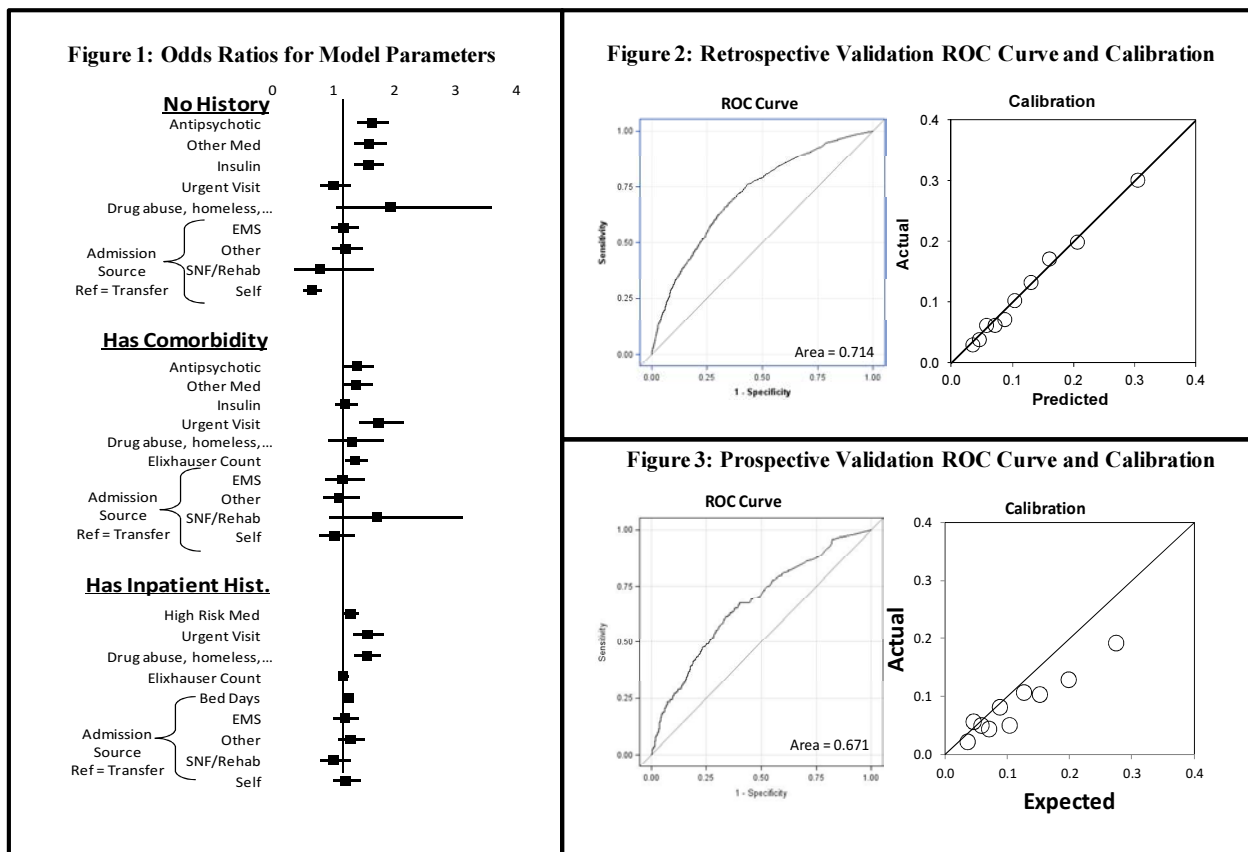


Table 2: Maximum Likelihood Estimates (MLE) of Model Parameters

Variables	No History		Outpatient Only		Has Prior Admission	
	MLE ± Error	P	MLE ± Error	P	MLE ± Error	P
Intercept	-2.904 ± 0.147	<.01	-3.189 ± 0.159	<.01	-2.944 ± 0.108	<.01
Admission Source						
Transfer (ref)	-0.091		0.145		0.110	
Emergency Med Svc.	0.227 ± 0.105	0.03	-0.024 ± 0.092	0.80	0.050 ± 0.048	0.30
SNF or Rehab	-0.193 ± 0.318	0.54	0.378 ± 0.234	0.11	-0.122 ± 0.085	0.15
Physician	-0.384 ± 0.117	<.01	-0.140 ± 0.092	0.13	0.061 ± 0.051	0.23
Other	0.259 ± 0.105	0.01	-0.070 ± 0.085	0.41	0.121 ± 0.045	<.01
Visit Type						
Urgent Visit (Yes)	-0.012 ± 0.127	0.93	0.544 ± 0.106	<.01	0.433 ± 0.080	<.01
History (Past year)						
Behavioral Issue	0.652 ± 0.318	0.04	0.245 ± 0.176	0.16	0.422 ± 0.072	<.01
Elixhauser Count			0.293 ± 0.068	<.01	0.132 ± 0.037	<.01
Inpatient Days					0.216 ± 0.017	<.01
Inpatient Medications						
Insulin	0.442 ± 0.090	<.01	0.305 ± 0.091	<.01	0.236 ± 0.053	<.01
Antipsychotic	0.478 ± 0.084	<.01	0.316 ± 0.090	<.01		
High Risk Drug	0.437 ± 0.079	<.01	0.160 ± 0.079	0.04		

Discussion

This project was a practical demonstration of developing and implementing a real-time 30-day readmission prediction model for clinical operational purposes in a large academic medical center. The model was moderately discriminative, and was successfully implemented in TopCare, our hospital’s population health management system. The infrastructure now exists at the MGH to notify providers early about patients who are at high risk for readmission within 30-days of discharge.

The AUC of the prospective real-time validation was similar to the retrospective validation. Although the model overestimated the outcome in the prospective validation, we attributed this in large part to the difference between the rates of readmission rates in the two groups (9.23% in the validation set and 12.1% in the derivation cohort). We also identified an issue where bed days in the prior year were being over-represented due to duplicates in the database, resulting in an overestimation of readmission risk.

As we developed and implemented this real-time model, we learned three important lessons essential to a successful implementation of a 30-day real-time readmission predictive model. First, risk scores had to be calculated daily for all inpatients beginning no later than one day after the patients’ admission date. The timeliness of this calculation was required because post-discharge interventions associated with reducing readmission had to be planned soon after a patient’s admission. As a result, we were forced to exclude many popular readmission predictors including length-of-stay and discharge diagnosis, as those variables are not available upon admission. Second, complexity had to be limited to readily available technical capabilities. For example, we excluded predictors identified using keyword searches because highly accurate

parsing could not be implemented easily. Third, the model had to make sense to hospitalists. Certain hospitalists had trouble believing the validity of the prior implemented model because it did not make sense clinically. We addressed this issue by inviting hospitalists to assist us with the development of the model. The success of the implementation of a real-time model required a fine balance between clinical believability, statistical requirements, availability of real-time data, and technical capabilities.

Limitations: The model was designed with data from a single institution from data for a single year, and may not be replicable at other institutions with different available data sources. In addition, we did not try to determine if a patient was readmitted at a non-MGH institution and are therefore underestimating total readmissions.

Conclusion

A moderately discriminative real-time 30-day readmission predictive model can be successfully implemented in a large academic hospital using existing data. Developers of real-time clinical predictive models need to consider more than the discrimination when developing models. An implementable real-time model balances clinical priorities, statistical requirements, availability of real-time data, and technical requirements.

References

1. Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. *JAMA : the journal of the American Medical Association*. Jan 23 2013;309(4):342-343.
2. Evans M. Healthcare's 'moneyball'. *Predictive modeling being tested in data-driven effort to strike out hospital readmissions*. 2011. <http://www.modernhealthcare.com/article/20111010/MAGAZINE/111009989#>. Accessed 3/12/2014.
3. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA : the journal of the American Medical Association*. Oct 19 2011;306(15):1688-1698.
4. Amarasingham R, Moore BJ, Tabak YP, et al. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. *Medical care*. Nov 2010;48(11):981-988.
5. Billings J, Dixon J, Mijanovich T, Wennberg D. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. *BMJ*. Aug 12 2006;333(7563):327.
6. Billings J, Mijanovich T. Improving the management of care for high-cost Medicaid patients. *Health Aff (Millwood)*. Nov-Dec 2007;26(6):1643-1654.
7. Wasfy JH, Rosenfield K, Zelevinsky K, et al. A prediction model to identify patients at high risk for 30-day readmission after percutaneous coronary intervention. *Circulation. Cardiovascular quality and outcomes*. Jul 2013;6(4):429-435.
8. Brown JR, Conley SM, Niles NW, 2nd. Predicting Readmission or Death After Acute ST-Elevation Myocardial Infarction. *Clinical cardiology*. Oct 2013;36(10):570-575.
9. Watson AJ, O'Rourke J, Jethwani K, et al. Linking electronic health record-extracted psychosocial data in real-time to risk of readmission for heart failure. *Psychosomatics*. Jul-Aug 2011;52(4):319-327.

10. Wallmann R, Llorca J, Gomez-Acebo I, Ortega AC, Roldan FR, Dierssen-Sotos T. Prediction of 30-day cardiac-related-emergency-readmissions using simple administrative hospital data. *International journal of cardiology*. Apr 5 2013;164(2):193-200.
11. Hammill BG, Curtis LH, Fonarow GC, et al. Incremental value of clinical data beyond claims data in predicting 30-day outcomes after heart failure hospitalization. *Circulation. Cardiovascular quality and outcomes*. Jan 1 2011;4(1):60-67.
12. Au AG, McAlister FA, Bakal JA, Ezekowitz J, Kaul P, van Walraven C. Predicting the risk of unplanned readmission or death within 30 days of discharge after a heart failure hospitalization. *American heart journal*. Sep 2012;164(3):365-372.
13. Whitlock TL, Tignor A, Webster EM, et al. A scoring system to predict readmission of patients with acute pancreatitis to the hospital within thirty days of discharge. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*. Feb 2011;9(2):175-180; quiz e118.
14. McAdams-DeMarco MA, Law A, Salter ML, et al. Frailty and early hospital readmission after kidney transplantation. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons*. Aug 2013;13(8):2091-2095.
15. Donze J, Aujesky D, Williams D, Schnipper JL. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. *JAMA internal medicine*. Apr 22 2013;173(8):632-638.
16. He D, Mathews SC, Kalloo AN, Hutfless S. Mining high-dimensional administrative claims data to predict early hospital readmissions. *Journal of the American Medical Informatics Association : JAMIA*. Sep 27 2013.
17. van Walraven C, 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 medicine : a peer-reviewed, independent, open-access journal*. 2012;6(3):e80-90.
18. TopCare Powered by Blender Patient Population Management Software. Fort Lauderdale, FL, FL: SRG Tech, Inc.: SRG Technology; 2013.
19. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Medical care*. Jan 1998;36(1):8-27.
20. Budnitz DS, Lovegrove MC, Shehab N, Richards CL. Emergency hospitalizations for adverse drug events in older Americans. *The New England journal of medicine*. Nov 24 2011;365(21):2002-2012.

Acknowledgements

The authors thank Steven Wong, for his invaluable assistance with collecting data for this study, and Tina Rong, for her critical role with implementation of the pilot study.