

Development of an Automated, Real Time Surveillance Tool for Predicting Readmissions at a Community Hospital

R. Gildersleeve¹; P. Cooper¹;

¹Augusta Health, Information Technology, Fishersville, Virginia, United States

Keywords

Clinical decision support, forecasting, alerting, monitoring and surveillance, data repositories

Summary

Background: The Centers for Medicare and Medicaid Services' Readmissions Reduction Program adjusts payments to hospitals based on 30-day readmission rates for patients with acute myocardial infarction, heart failure, and pneumonia. This holds hospitals accountable for a complex phenomenon about which there is little evidence regarding effective interventions. Further study may benefit from a method for efficiently and inexpensively identifying patients at risk of readmission. Several models have been developed to assess this risk, many of which may not translate to a U.S. community hospital setting.

Objective: To develop a real-time, automated tool to stratify risk of 30-day readmission at a semi-rural community hospital.

Methods: A derivation cohort was created by extracting demographic and clinical variables from the data repository for adult discharges from calendar year 2010. Multivariate logistic regression identified variables that were significantly associated with 30-day hospital readmission. Those variables were incorporated into a formula to produce a Risk of Readmission Score (RRS). A validation cohort from 2011 assessed the predictive value of the RRS. A SQL stored procedure was created to calculate the RRS for any patient and publish its value, along with an estimate of readmission risk and other factors, to a secure intranet site.

Results: Eleven variables were significantly associated with readmission in the multivariate analysis of each cohort. The RRS had an area under the receiver operating characteristic curve (c-statistic) of 0.74 (95% CI 0.73-0.75) in the derivation cohort and 0.70 (95% CI 0.69-0.71) in the validation cohort.

Conclusion: Clinical and administrative data available in a typical community hospital database can be used to create a validated, predictive scoring system that automatically assigns a probability of 30-day readmission to hospitalized patients. This does not require manual data extraction or manipulation and uses commonly available systems. Additional study is needed to refine and confirm the findings.

Correspondence to:

P. Cooper
Augusta Health
78 Medical Drive
Fishersville, VA 22939
Email: pcooper@augustahealth.com

Appl Clin Inform 2013; 4: 153–169

DOI: 10.4338/ACI-2012-12-RA-0058

received: December 28, 2012

accepted: March 18, 2013

published: April 3, 2013

Citation: Gildersleeve R, Cooper P. Development of an automated, real time surveillance tool for predicting readmissions at a community hospital. *Appl Clin Inf* 2013; 4: 153–169

<http://dx.doi.org/10.4338/ACI-2012-12-RA-0058>

1. Background

Effective October 1, 2012, the Center for Medicaid and Medicare Services began its Readmissions Reduction Program, which adjusts payments to hospitals based on 30-day readmission rates for patients initially admitted with acute myocardial infarction, heart failure and pneumonia [1]. This holds hospitals accountable for a complex phenomenon about which there is little evidence regarding effective interventions [2, 3]. Some hospitals are making system-wide changes in education, discharge planning, medication management, and care coordination prior to, during, and after discharge to improve care and reduce readmissions [4]. Programs to reduce readmissions by improving care transitions have had mixed results [3]. Efficiently and inexpensively identifying patients at greatest risk of deterioration after hospital discharge may help focus interventions in a more effective manner. An effective tool would capture existing data from electronic documentation without manual review, be used during an index admission, and be presented in an intuitive manner to personnel who intervene with high risk patients.

Publications have evaluated readmission prediction models for decades. Kansagara et al. conducted a systematic review of twenty-six such models [5]. The predictive value, as assessed by the *c*-statistic (i.e. the area under the receiver operating characteristic curve) for these models ranged from 0.56 to 0.83. The one with the highest *c*-statistic (0.77) [6] used retrospective administrative data in a Medicare population; performance was enhanced to 0.83 by adding a questionnaire that was not typically completed until after an index hospitalization. Another model [7] using retrospective data from medical and surgical patients in Canada derived a “LACE” score that yielded a *c*-statistic of 0.68. This simple, four-variable calculation was based on data that could be gathered by the end of an index admission: length of stay of the index admission (L), acute versus planned admission (A), the Charlson Comorbidity Index (C), and the number of emergency department visits in the six months preceding the index admission (E). (The Charlson Comorbidity Index (CCI) is a validated score for predicting mortality based on ICD-9 encoded medical morbidities [8]. The LACE model modified the CCI by reweighting some of the morbidities, following Schneeweiss [9].) This LACE model has been successfully applied in a Canadian population [10] and could be adapted to real-time automation. A study of patients with heart failure at an underserved urban center in the U.S. [11] used real-time administrative and clinical data extracted from an Electronic Health Record (EHR) and yielded a *c*-statistic of 0.72. Billings has developed models to predict hospital readmission in the English healthcare system at one year [12] and thirty days [13]. Both were developed from a broad population and used real-time data, achieving *c*-statistics of 0.69 and 0.70 respectively. A study at six academic U.S. medical centers employed real-time data collection, but relied in part on an interview with the patient by a research assistant within 48 hours of admission, advised caution in applying the results to community hospitals, and had a *c*-statistic of 0.61 [14]. Not described in the literature is a model in the U.S. health system that gathers data from adult patients in a community setting; is indifferent to payer source; applies to all medical-surgical problems rather than a subset of diseases; automatically extracts data from commercially available EHR software; has favorable performance characteristics; presents risk assessments in an accessible, easy-to-use format; and can be carried out with resources typically available at community hospitals.

2. Objective

The objective was to develop a real-time, automated tool to stratify risk of 30-day readmission at a semi-rural community hospital.

3. Methods

3.1 Context and Data Sources

The model was developed at Augusta Health, a 255-bed community hospital staffed by approximately 180 physicians and 2,300 employees. The hospital's primary service area of approximately

120,000 people is mostly agricultural and light industrial. The hospital has 60,000 annual emergency department encounters and 12,000 admissions annually, totaling 52,000 inpatient days. Service lines include most medical-surgical specialties except for neurosurgery and cardiothoracic surgery. Additionally, there are inpatient gynecologic, obstetric, pediatric, psychiatric, rehabilitative, and skilled nursing units.

During the study period, the population was served by a variety of outpatient practices (independent, employed by other health systems, and employed by the hospital) that used multiple paper and electronic records, none of which had inbound interfaces to the inpatient EHR (MEDITECH Client-Server 5.64). The MEDITECH data repository, which serves as a long-term archive of all EHR data, is a relational system that serves as the platform for data collection and analysis. Approximately 70% of admitted patients have administrative and clinical data recorded in the EMR from previous inpatient visits. During the study, the problem list did not consistently capture patients' clinical problems, but ICD-9 codes entered from previous hospitalizations provided some clinical information. Approximately 30% of admitted patients lacked ICD-9 codes because they had no previous inpatient stays. Ambulatory medication lists were updated at multiple points of care, such as preadmission testing, home health visits, the Emergency Department, and upon hospital discharge.

3.2 Study populations

A derivation cohort was created by extracting from the data repository, demographic and clinical variables for adult hospital discharges from the calendar year 2010. The entire year was selected to limit potential effects of seasonality on hospital admissions. All patients who did not meet exclusion criteria were included to avoid potential sampling bias. Patients were excluded who were admitted to the psychiatric, rehabilitative, or skilled nursing units; were less than 18 years of age; left against medical advice; or died during an index hospitalization. All repeat admissions within the 365-day time frame of the cohort were counted as readmissions as long as they followed an index admission that occurred in the preceding thirty days. Readmissions to outside hospitals were not considered as that information is not available in the data repository. Similarly, there was no mechanism to capture patients who died outside the hospital after discharge. A validation cohort was created from discharges from calendar year 2011, with the same criteria as the derivation cohort. The entire year and all eligible patients were again included to avoid seasonality and sampling bias.

3.3 Creation and Comparison of Risk Scores

The first step was development of a predictive score based on the LACE model. This required automating the calculation of the CCI (the "C" in "LACE"). This score applies variably weighted values based on whether the patient has heart failure, myocardial infarction, vascular disease, dementia, COPD, connective tissue disease, peptic ulcer disease, liver disease, diabetes, stroke, renal disease, cancer, or AIDS. To capture these comorbid conditions, we used the ICD-9 diagnoses identified by Quan et al. [15]; this is the same methodology employed by the creators of the LACE model. Similarly, the ICD-9 codes used in our model were entered by professional coders, who manually abstracted the data after hospital discharge. In automating the calculation, we added an age-adjustment described by Hall et al [16], who also provided an electronic tool for calculating the CCI and who found that adjusting the score for advancing age enhanced its predictive value. Thus, we employed a modified, age-adjusted Charlson Comorbidity Index (mCCI). Lastly, the original LACE model maps the value of the CCI to a limited number of points (e.g., a CCI of 4 or more results in a maximum comorbidity point score of 5). We categorized our mCCI based on patterns and visual breakpoints observed in the distribution of our population. Similarly, we selected cutoffs for ED visits, inpatient stays, length of stay, and ambulatory medications based on how those categorical variables were distributed among our patients.

The predictive value of this automated, modified LACE model in predicting 30-day readmission as a dichotomous outcome was assessed by the area under the receiver operating characteristic curve (the c-statistic). To see if the performance characteristics of this modified LACE model could be improved upon, we explored additional variables from previous studies included in the systematic review by Kansagara et al. [5]. We limited these items to those that are recorded as a matter of routine

in the patient's electronic record and stored in the data repository. To this list, we added two additional variables related to medication use: the numbers of ambulatory and inpatient medications. Only scheduled (i.e. non-PRN) medications were used; the count of scheduled inpatient medications was tallied two days prior to discharge.

In addition to the modified LACE elements and two medication variables, we selected six more to test for statistical significance: age, male sex, whether the patient is married, uninsured status, number of inpatient and observation hospital stays in preceding 365 days, and whether the patient lived alone. Thus, twelve candidate variables were assessed for statistical significance. Variables that were not normally distributed, such as the mCCI, were segregated into categories prior to the logistic regression analysis based on patterns and visual break-points observed in the distributions in our patient population.

A combination of dichotomous, continuous, and categorical variables identified as being statistically significant were then incorporated into a formula to yield a Risk of Readmission Score (RRS). A patient's RRS score was calculated by multiplying the variable value by its beta coefficient. However, for the categorical variables that were not normally distributed (mCCI, ED visits, inpatient stays, ambulatory medications, and length of stay), the raw value yielded a beta coefficient by virtue of its category. For these categorical variables, the beta coefficient was multiplied by 1 in the RRS calculation.

3.4 Dashboard Presentation

After developing the method to calculate the RRS, we automated the process of assigning and displaying an estimate of risk to individual patients. The presentation is limited to current inpatients, with the aim of identifying patients prior to discharge for additional interventions. The length of stay variable when applied to currently hospitalized patients is calculated based on the current stay, essentially considering it to be a potential index admission. The ED visit that led to the current hospitalization is not included in the ED count. The derivation cohort was divided into ten equal groups, stratified by increasing RRS scores. The cutoffs for each decile were then used to assign future patients to a risk group based on readmission rates in the derivation cohort, with each group having a collective percent risk of readmission. This assignment process was carried out for the validation group, and the expected versus observed rates of readmission determined. The RRS and corresponding gross risk were calculated via a scheduled job that processes a Microsoft SQL Server stored procedure. The stored procedure gathered data on all non-excluded inpatients, computed their scores, and transferred the resulting data to the hospital clinical surveillance database structure. Data were then displayed in a secure intranet environment via a dashboard developed with Microsoft Visual Studio.

In addition to the RRS and risk grouping, selected demographic data, insurance carrier, medication counts, and mCCI were recorded. Lastly, if the patient had ICD-9 codes in their record that suggested they have diabetes mellitus, heart failure, acute myocardial infarction, pneumonia, or chronic obstructive pulmonary disease, those disease states were listed on the dashboard as well. No access to the site was granted during the study so as to avoid any possible effect on measured outcomes.

3.5 Statistical Analyses

Descriptive information for 2010 derivation and 2011 validation populations was reported as percentages for categorical elements, means with standard deviations for continuous data, and median with interquartile ranges for those variables not normally distributed. Univariate statistical analysis compared 30-day readmission and 30-day non-readmission (dichotomous categorical outcome variable) to dichotomous categorical predictor variables using Fisher's Exact Test with odds ratios and 95% confidence intervals reported. Variables with three or more categories were compared to the dichotomous outcome using a Pearson's Chi square. Univariate statistical analyses compared the dichotomous categorical outcome variable to continuous predictor variables using Student's *t*-test with mean differences and standard error reported.

For the multivariate analysis, multinomial logistic regression was employed where a single block of *a priori* selected 12 predictor variables were included at once. Post hoc comparison of the univariate and multivariate results were subsequently examined to assess the role of covariation between predictors. A Risk of Readmission Score was calculated for all non-excluded patients (including the 2011 validation sample) using the logistic regression equation result from the 2010 derivation cohort. C statistic values with 95% confidence intervals were compared for the modified LACE and RRS models. The C statistic with 95% confidence interval was also calculated for the 2011 validation cohort using the RRS as a predictor of 30-day hospital readmission. Sensitivity, specificity, and positive and negative predictive values were calculated for both populations using the mean RRS as well as a higher value arbitrarily picked as an example of a patient who would be designated by the model as being at high risk. A final set of binary logistic regression analyses were used to statistically compare the predictive value of the RRS for both cohorts. To assess goodness-of-fit, Nagelkerke R^2 and Hosmer-Lemeshow p-value statistics are reported. The statistical software used was Statistical Package for the Social Sciences (SPSS) version 20.

The study was performed in compliance with the World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects. The study was reviewed by the Augusta Health Institutional Review Board.

4. Results

The 2010 derivation cohort consisted of 8,700 patients, 14.1% of whom were readmitted within 30 days of any index admission. The 2011 validation cohort consisted of 8,189 patients, with a 14.8% readmission rate. ► Table 1 shows characteristics of demography and healthcare utilization of the 2010 and 2011 groups, with the twelve candidate predictor variables denoted with an asterisk. The patients were overwhelmingly white (92.7% in both groups), half were married (49.5%), and less than half male (39.1%). Approximately 55% in both cohorts were Medicare. The median length stay was three days in both groups. The validation cohort had higher age (65 versus 60.6 years), more comorbid conditions (median mCCI scores (seven versus six)), were on more inpatient medications (16 versus 14), and were more likely to live alone (18.1% v. 9%).

Univariate analyses were performed with each of the twelve candidate predictor variables to test for statistically significant differences between patients who were readmitted within 30 days and those who were not, as shown in ► Table 2. The only two characteristics that were not statistically significantly different were uninsured status ($p = 0.35$) and the number of ambulatory medications ($p = 0.67$). Demographically and socially, readmitted patients were older ($p < 0.0001$) and more likely to be male ($p < 0.0001$), unmarried ($p < 0.0001$), and living alone ($p = 0.001$). From a healthcare utilization standpoint, they had more ED visits ($p < 0.0001$), more admissions and observation visits ($p < 0.0001$), more unplanned (acute) admissions ($p < 0.0001$), longer lengths of stay ($p < 0.0001$), more medications on their inpatient medication list two days prior to discharge ($p < 0.0001$), and had higher mCCIs ($p < 0.0001$).

The multivariate binary logistic regression results are summarized in ► Table 3. The overall percentage of variance accounted for was 14% (Nagelkerke R^2) and the Hosmer-Lemeshow had an X^2 of 21.6, ($p = 0.006$).

Eleven of twelve candidate variables were significantly associated with 30-day readmission when all twelve predictors were simultaneously entered into the regression equation. One variable, living alone, that had been significant in the univariate analysis became non-significant in the multivariate analysis ($p = 0.80$). On the other hand, two non-significant univariate predictors became significant as part of the multivariate equation. Uninsured status predicted readmission in the multivariate analysis ($p = 0.03$). Patients with six or more ambulatory medications were significantly less likely to be readmitted ($p < 0.0001$). The other nine variables remained significant across the analyses: age, being married, being male, being acutely admitted, experiencing more ED visits and hospital stays, having longer lengths of stay, being on more inpatient medications, and having a higher mCCI.

Living alone, although a significant univariate predictor was not a significant multivariate predictor. The number of ambulatory medications was a non-significant univariate predictor but was a significant negative multivariate predictor of 30-day readmission. Both ambulatory and inpatient

medications share variance with several other variables, including age, hospital stays, uninsured status, length of stay, and the mCCI. Therefore, the unique variance, not accounted for by other predictors, may be identifying healthier patients who are appropriately medicated, as only revealed when other risk factors are held constant.

A Risk of Readmission Score was created for each patient by multiplying the values for each of the significant variables by the beta coefficient for each variable. Beta coefficients of the categorical variables are multiplied by one instead of a raw value. The mean RRS was 1.7 and ranged from -0.17 to 4.89. The area under the receiver operating characteristic curve (*c*-statistic) was 0.74 (95% CI 0.73-0.75) for the derivation cohort. The *c*-statistic for our modification of the LACE model (0.71, 95% CI 0.70-0.72) as applied to this population was comparable to the value of 0.68 reported by its developers. The *c*-statistic of the validation cohort was 0.70 (95% CI 0.69-0.71). The ten stratified risk groups had probabilities of readmission ranging from a low of 3% to a high of 38%, as shown in ► Table 4. ► Table 5 shows sensitivity, specificity, and positive and negative predictive values with 95% CIs for the 2010 derivation and 2011 validation cohorts using two different Risk of Readmission Scores. One is the mean RRS, and the second is a higher-risk cutoff score in the second-highest risk decile. For a clinical example of a patient with that high risk score, following the format of Billings et al. [13], see ► Table 6. Measures of test performance were similar for both populations, with somewhat lower values observed in the validation population.

A multivariate binary logistic regression equation was used to compare the predictive accuracy of the RRS for predicting readmissions in the 2010 derivation and the 2011 validation groups. The 2010 and 2011 samples were combined. Year of Sample, Risk of Readmission and an interaction term representing Year of Sample by Risk of Readmission were entered in a single regression equation with readmission as the dependent variable. As expected, the RRS ($p < 0.0001$) and Year of Sample (0.0002) were highly significant predictors. There were more 30 day readmissions in 2011 compared to 2010. The interaction term was not significant ($p = 0.36$), indicating no significant difference in the predictive value of the RRS between the two years.

► Figure 1 visually demonstrates the Hosmer-Lemeshow expected versus observed rates of readmission in the validation cohort. ► Figure 2 shows the display of current inpatients' individual scores, probability of readmission, and other clinical and demographic data in a clinical surveillance site within the hospital's intranet. Each row represents one patient. Fourteen different columns contain clinical, demographic, and predictive data. The raw score is posted in the "Total" column, and the risk of readmission is the percent value in the rightmost column. With this presentation, current inpatients can be grouped in descending order of readmission risk, by certain diagnoses, etc. by clicking on the column header. The display in ► Figure 2 is sorted by hospital unit. This functionality was chosen to allow users with different roles to drill down into certain subgroups of patients, such as those without insurance or those with certain disease states. Patients' names have been obscured in this presentation.

5. Discussion

Twelve administrative and clinical data elements were tested for association with thirty day hospital readmission. Eleven of them have been assessed in other studies and lent themselves to automated extraction from our data repository [5, 17]. The number of ambulatory medications was a novel variable also tested. Eleven of the twelve were significantly associated with thirty day readmission to the same hospital based on large derivation and validation cohorts using multivariate logistic regression. A Risk of Readmission Score was created based on these variables, with a *c*-statistic that was comparable to the predictive value of our modification of the LACE model. The *c*-statistic of the RRS for the derivation cohort was 0.74 (95% CI 0.73-0.75), which was slightly but significantly higher than the RRS model as applied to the large validation group (0.70, 95% CI 0.69-0.71). Despite trying to control for potential seasonal effects or sampling error by using full calendar years and large cohorts, the two groups had several demographic, social, and healthcare utilization characteristics, which may have contributed to the differences. Nonetheless the *c*-statistic of the model of the validation cohort compares favorably with other published models.

It has been shown that measures of healthcare utilization, medical morbidities, and demographic variables predict readmission in various populations. The findings of our study support the generalizability of other models, most of which were developed from data on urban populations, academic centers, Medicare patients, specific disease states, or outside the U.S. healthcare system. Specifically, it adds support for generalizing some of these measures to more rural communities in the U.S. It largely supports the findings of a recent, relatively small study from an academic tertiary care center of U.S. family medicine patients [17]. That study demonstrated significance for length of stay, previous hospitalizations, Emergency Department use, number of discharge medications, and common medical comorbidities. It also showed a significant protective effect of being married, which was confirmed in our multivariate analysis ($p = 0.008$). However, it showed no effect of male sex, which was unfavorably associated in our cohort ($p < 0.0001$). Living alone has been associated with readmission in an elderly population [18], but was not in our more diverse cohort ($p = 0.80$). The divergence of these factors (sex, living alone) in various studies suggests the need for additional study, or perhaps a need to derive models based on local populations.

Multivariate logistic regression identified medication use as a significant variable in this population. Increasing numbers of scheduled inpatient medications as measured two days prior to hospital discharge was associated with increased risk of readmission. A greater number of medications may be a marker of illness severity, identify patients who are heavy users of health care, or correlate with adverse effects of polypharmacy. On the other hand, we observed a protective effect of increasing numbers of ambulatory (preadmission) medications, a finding observed in both cohorts. It is conceivable that multiple ambulatory medications indicate appropriate attention to existing medical problems, medical compliance, and/or medication awareness that outweigh detrimental effects of polypharmacy. A limitation of using this variable is that preadmission medication lists are often inaccurate, particularly when the inpatient EHR does not interface with a broader prescription management database or communicate with office records. As noted, the favorable effect of ambulatory medications was not significant in univariate analysis, but became so in the multivariate logistic regression for patients on six or more medications. Regardless of the rationale or inaccuracies, the negative correlation of preadmission medications with subsequent readmission was incorporated into the model because of the multivariate statistical significance of the finding.

The mCCI was used as a composite variable to assess medical morbidities, in addition to the medication variables described above. However, it relied on ICD-9 encoded diagnoses, all of which were entered by professional coders after previous hospital stays. Because the data repository has no coded medical diagnoses for patients who have not received care in the hospital system, those patients will have an inappropriately low mCCI and their predicted risk of readmission would be inappropriately low. This is a weakness of the model. Furthermore, ICD-9 diagnoses abstracted by professional coders have been shown to be of limited sensitivity and positive predictive value [19]. A SNOMED-encoded problem list is increasingly used by clinicians in the hospital, but is not yet adequately populated to provide timely and accurate clinical information. Accessing an actively managed problem list that is available during a given hospitalization could improve the model.

The model presented here does not improve upon the overall predictive ability of some of the published models, although it compares favorably with most. However, it does show that the necessary elements for creating a predictive algorithm can be readily collected from a commercial EHR data repository and synthesized into an automated calculation with the level of expertise and software available at a community hospital.

In addition to deriving a risk model based on local data, we created a graphical interface that could be used by approved personnel. Discharge planners and case managers in particular might use this to target patients for focused evaluation and follow-up care. The ability to sort patients by disease states, insurance type, and unit location could allow specialized administrative personnel to identify populations of interest, such as for a subspecialty continuity clinic or free clinic. Assigning an RRS to current inpatients brings up methodological questions about variables tied to an index admission (such as length of stay) or that would change on a daily basis (such as inpatient medication counts). The model as developed and presented here could be applied to current inpatients on their day of discharge, but its accuracy would need to be reassessed if used earlier in the stay or even on admission. Use of the model by physicians remains an unexplored topic.

5.1 Limitations

Limitations discussed above include reliance on administrative codes to calculate the clinical morbidity score and incomplete or inaccurate ambulatory medication lists. The finding that ambulatory medications may protect against readmission is an unstudied and potentially counterintuitive finding. Data used to derive and validate the model did not include hospitalizations at other facilities, thus probably underestimating the risk of readmission; this is only partially mitigated by the fact the hospital is responsible for the great majority of hospital care in its catchment area. In addition, out-of-hospital deaths after discharge were not included, further underestimating clinically significant post-discharge events. Furthermore, our use of a validation cohort of the same magnitude as the derivation group is not typical, and there was no attempt to identify potentially preventable readmissions. Pediatric, psychiatric, and rehabilitative admissions were not studied.

6. Conclusion

This automated, real-time forecasting tool was derived from readily available data in a community hospital population and created using EHR and data processing applications in widespread use across the U.S. This study supports the generalizability of several risk factors for readmission in a semi-rural adult population (health care utilization, medication use, comorbidities), but also suggests that modeling based on local data may be necessary for certain factors (sex, living alone). The model includes multiple disease states and in fact applies to all types of admissions outside of pediatrics, psychiatry, and rehabilitation. Its predictive ability compares favorably with other published models. It results in a dashboard that allows designated users to obtain information of interest with minimal interaction with the user interface, and do so at or even prior to discharge. Weaknesses include incomplete information about clinical diagnoses, medications, deaths, and readmissions to other facilities. Ongoing study is needed to externally validate readmission risk prediction in a community setting, particularly the influence of ambulatory medications.

Clinical Relevance Statement

Community hospitals can develop a tool that predicts readmission among their populations using readily available software and a commercial EHR. This can be achieved automatically, without any manual data collection or manipulation. The information can be presented to end-users in an intuitive format that may assist hospitals in directing scarce resources to at-risk patients.

Conflicts of Interest

The authors declare that they have no conflicts of interest in this research.

Acknowledgements

We thank Dr. Fred Castello for his guidance and acknowledge Relana Pinkerton of the University of Virginia for assistance with the statistical evaluation of the data.

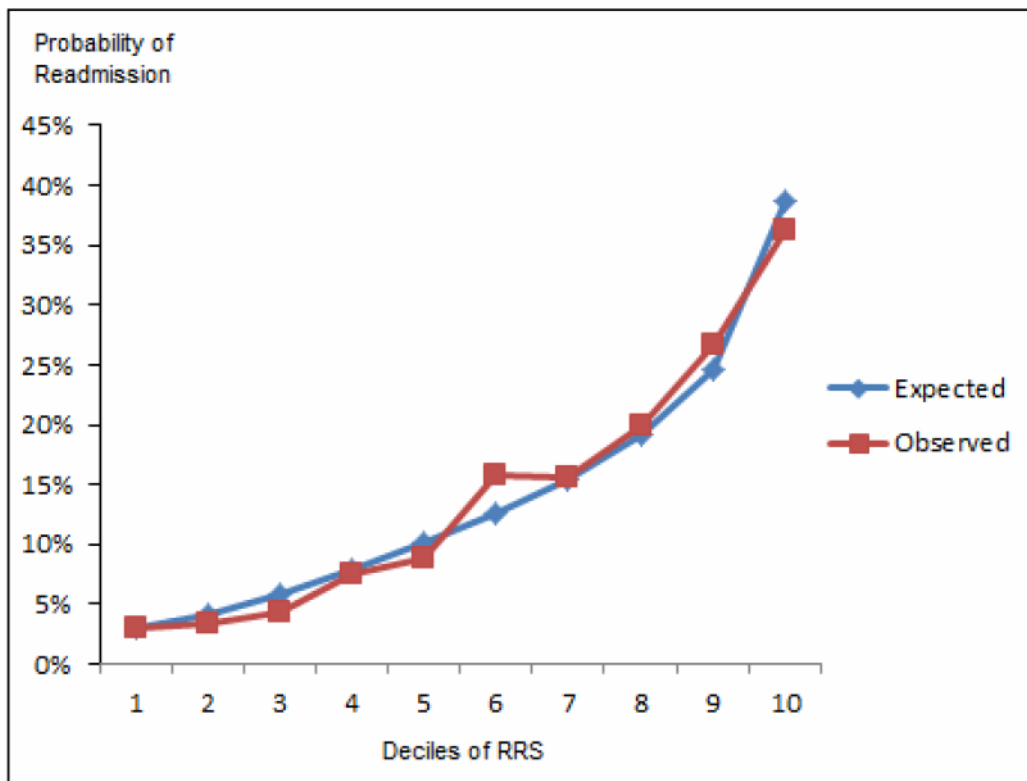


Fig. 1 Expected vs Observed

Census Influenza c.Difficile Restraints Falls Readmission															
Show 10 entries Print Search:															
Account Number	Name	Unit	ER Adm	Age	M/F	ER Cnt	IN Cnt	FC	LOS	In Med	Out Med	CCI	Area	Total	RRS
[Redacted]	[Redacted]	[Redacted]	Y	75	M	7	0	MC	4	23	0	8		3.545	38%
[Redacted]	[Redacted]	[Redacted]	Y	88	F	6	1	MC	2	29	15	11	COPD DM HF	3.376	38%
[Redacted]	[Redacted]	[Redacted]	Y	49	M	9	2	SP	0	6	1	5		3.233	38%
[Redacted]	[Redacted]	[Redacted]	Y	59	M	6	1	MC	2	18	2	7	HF	3.183	38%
[Redacted]	[Redacted]	[Redacted]	Y	78	F	3	0	MC	2	36	10	9		3.036	38%
[Redacted]	[Redacted]	[Redacted]		56	M	3	0	SH	3	17	9	6		2.012	17%
[Redacted]	[Redacted]	[Redacted]		86	F	2	1	MC	2	22	0	10	DM	1.962	17%
[Redacted]	[Redacted]	[Redacted]	Y	57	F	2	5	MD	3	23	43	6		1.959	17%
[Redacted]	[Redacted]	[Redacted]	Y	69	M	0	1	MC	5	12	25	7		1.923	13%
[Redacted]	[Redacted]	[Redacted]	Y	59	M	2	0	MC	4	27	9	6		1.913	13%

Showing 31 to 40 of 92 entries

Fig. 2 Readmission Dashboard

Characteristic	Derivation Cohort n = 8,700		Validation Cohort n = 8,189	
	n	%	n	%
Readmitted	1230	14.1	1211	14.8
Demographic/social				
Male sex*	3404	39.1	3204	39.1
Age*	Mean	Stdev	Mean	Stddev
	60.6	20.6	65	20.7
Age group	n	%	n	%
• 18 – 35 Years	1489	17.1	1392	17.0
• 36 – 44 Years	517	5.9	433	5.3
• 45 – 64 Years	2416	27.8	2200	26.9
• 65 + Years	4278	49.2	4164	50.8
Married*	4304	49.5	4057	49.5
Lives alone*	783	9.0	1487	18.1
Race/Ethnicity				
• Asian	10	0.1	14	0.2
• African American	519	6	507	6.2
• Hispanic	48	0.6	43	0.5
• American Indian/Alaskan Native	4	0.0	3	0.0
• Unknown	57	0.7	34	0.4
• White	8062	92.7	7588	92.7
Healthcare utilization				
Payers				
• Commercial	2241	25.8	2089	25.5
• Medicare	4811	55.3	4676	57.1
• Medicaid	952	10.9	922	11.3
• Uninsured*	644	7.4	467	5.7
• Other	52	0.6	35	0.4
• Acute admission*	5734	66.0	5466	66.8
ED Visits within one year*	Median	IQR	Median	IQR
	0	0–2	0	0–1
	n	%	n	%
• 0	4476	51.5	5760	70.3
• 1–2	2611	30.0	1877	22.9
• 3–5	990	11.4	450	5.5
• ≥6	623	7.2	102	1.2

Table 1
Descriptive characteristics of study cohorts

Table 1 Continued

Characteristic	Derivation Cohort n = 8,700		Validation Cohort n = 8,189	
	Inpatient Visits within one year*	Median	IQR	Median
	0	0–1	0	0–1
	n	%	n	%
• 0	4935	56.7	4598	56.1
• 1–2	2641	30.4	2538	31.0
• 3–5	886	10.2	832	10.2
• ≥6	238	2.7	221	2.7
Length of stay*	median	IQR	median	IQR
	3	2–4	3	2–4
	n	%	n	%
• 0–1	1641	18.9	1418	17.3
• 2–3	4383	50.4	4178	51.0
• 4–8	2236	25.7	2067	25.2
• ≥9	440	5.1	526	6.4
Inpatient medications*	Mean	Stdev	Mean	Stddev
	14.0	5.9	16.1	6.5
Ambulatory medications*	Median	IQR	Median	IQR
	2	0–6	2	0–5
	n	%	n	%
• 0–1	3661	42.1	3656	44.6
• 2–5	2497	28.7	2543	31.1
• 6–14	2035	23.4	1686	20.6
• ≥15	507	5.8	304	3.7
Modified Charlson Comorbidity Index*	Median	IQR	Median	IQR
	6	0–8	7	0–9
	n	%	n	%
• 0–3	3405	39.1	3064	37.4
• 4–8	3453	39.7	2231	27.2
• ≥9	1842	21.2	2894	35.3

*Candidate predictor variables. Variables that are not normally distributed, such as the mCCI, are grouped according to patterns and visual break-points observed in the distributions in our patient population.

Table 2 Univariate analysis of variables assessed for association with 30-day readmission for the 2010 derivation cohort (n = 8,700). ****

	Not Readmitted		Readmitted		Statistics		
	n	%	n	%	OR	95% CI**	P-value
Acute Admission (i.e., not scheduled)** 66% (5734/8700)	4746	83	988	17	2.3	2.0-2.7	<0.0001
Age***							
Mean = 60.6 (Std = 20.6)	Mean	Std	Mean	Std	Mean Dif.	Std Error***	P-value
	59.7	20.9	66.1	17.0	-6.4	0.63	<0.0001
Male Sex** 39% (3404/8700)	n	%	n	%	OR	95% CI**	P-value
	2828	83	576	16.9	1.4	1.3-1.6	<0.0001
Married** 50% (4304/8700)	n	%	n	%	OR	95% CI**	P-value
	3777	88	527	12	0.73	0.65-0.83	<0.0001
ED Visits within one year*	n	%	n	%	Chi-Square*		P-value
● 0	4024	53.9	452	36.7	236.0		<0.0001
● 1-2	2232	29.9	379	30.8			
● 3-5	780	10.4	210	17.1			
● ≥6	434	5.8	189	15.4			
Inpatient Visits within one year*	n	%	n	%	Chi-Square*		P-value
● 0	4451	59.6	484	39.3	312.4		<0.0001
● 1-2	2223	29.8	418	34.0			
● 3-5	649	8.7	237	19.3			
● ≥6	147	2.0	91	7.4			
Uninsured** 7%(645/8700)	n	%	n	%	OR	95% CI**	P-value
	546	85	99	15	1.2	0.89-1.4	0.35

Table 2 Continued

	Not Readmitted		Readmitted		Statistics		
	n	%	n	%	OR	95% CI**	P-value
Lives alone** 9% (783/8700)	641	18	641	82	1.4	1.1-1.7	0.001
Length of Stay*	n	%	n	%	Chi-square*		P-value
● 0-1	1442	19.3	199	16.2	127.5		<0.0001
● 2-3	3893	52.1	490	39.8			
● 4-8	1804	24.1	432	35.1			
● ≥9	331	4.5	109	8.9			
Medications on the Patient's Inpatient Medication List Two Days Prior to Discharge*** 14.0 (5.9), 0 – 55	Mean	Std.	Mean	Std.	Mean Dif.	Std error***	P-value
	13.6	5.8	16.2	6.3	-2.5	0.18	<0.0001
Medications on the Patient's Ambulatory Medication List*	n	%	n	%	Chi-square*		P-value
● 0-1	3150	42.2	511	41.5	1.581		0.67
● 2-5	2138	28.6	359	29.2			
● 6-14	1755	23.5	280	22.8			
● ≥15	427	5.7	80	6.5			
Modified Charlson Comorbidity Index*	n	%	n	%	Chi-square*		P-value
● 0-3	3203	42.9	202	16.4	311.4		<0.0001
● 4-8	2795	37.4	658	53.5			
● ≥9	1472	19.7	370	30.1			

*** Statistical comparisons based on readmission for categorical variables were assessed using Fisher's Exact Test 2x2 with Odd Ratios with 95 percent confidence interval and Pearson's Chi Square for all other categorical predictors. Continuous variables were reported with a mean difference with standard error of the difference and assessed using a Student's t-test.

Variable	Beta Coefficient	Std. Error	P-value
Acute Admission (i.e., not scheduled)	0.26	0.063	<0.0001
Age	0.007	0.002	<0.0001
Male Sex	0.28	0.048	<0.0001
Married	-0.13	0.051	0.008
ED Visits within one year			
• 0	--		
• 1–2	-0.25	0.053	<0.0001
• 3–5	0.82	0.073	<0.0001
• ≥6	1.16	0.113	<0.0001
Inpatient Visits within one year			
• 0	--		
• 1–2	0.39	0.055	0.39
• 3–5	0.82	0.073	<0.0001
• ≥6	1.16	0.113	<0.0001
Uninsured	0.22	0.10	0.03
Lives alone	0.017	0.067	0.80
Length of Stay			
• 0–1	--		
• 2–3	-0.16	0.069	0.019
• 4–8	0.23	0.072	0.002
• ≥9	0.39	0.099	<0.0001
Medications on the Patient's Inpatient Medication List Two Days Prior to Discharge	0.03	0.004	<0.0001
Medications on the Patient's Ambulatory Medication List			
• 0–1	--		
• 2–5	-0.10	0.056	0.087
• 6–14	-0.23	0.061	<0.0001
• ≥15	-0.48	0.105	<0.0001
Modified Charlson Comorbidity Index			
• 0–3	--		
• 4–8	0.68	0.075	<0.0001
• ≥9	0.72	0.096	<0.0001

Table 3

Multivariate binary logistic regression results using the 2010 derivation cohort (n = 8,700) and with all variables maintained in the model. A patient's RRS score is calculated by multiplying the variable value by its beta coefficient. However, for the categorical variables that are not normally distributed (mCCI, ED visits, inpatient stays, ambulatory medications, and length of stay), the raw value identifies a beta coefficient by virtue of its category. For these categorical variables, the beta coefficient is multiplied by 1 in the RRS calculation.

Table 4 Risk of Readmission Score grouped into ten deciles and assigned to probability of 30-day readmission.

Decile	RRS	Not Readmitted	Readmitted	% Probability
1	-0.173 – 0.481	846	24	3%
2	0.481 – 0.78	832	38	4%
3	0.78 – 1.111	829	41	5%
4	1.111 – 1.444	815	55	6%
5	1.445 – 1.684	769	101	12%
6	1.684 – 1.933	759	111	13%
7	1.933 – 2.188	726	144	17%
8	2.188 – 2.495	696	174	20%
9	2.496 – 3.033	657	213	24%
10	3.033 – 4.891	541	329	38%

Table 5 Risk of readmission score cutoffs, sensitivity, specificity, positive and negative predictive values with 95% CI for 2010 derivation and 2011 validation populations.

	2010 Derivation		2011 Validation	
	Mean Cutoff (1.7)	High Score (2.5)	Mean Cutoff*	High Score**
Sensitivity	74.9 (72.4–77.3)	30.9 (28.3–33.6)	79.2 (76.8–81.4)	30.9 (28.3–33.6)
Specificity	54.4 (53.2–55.5)	54.4 (53.2–55.6)	55.4 (54.3–56.5)	88.4 (87.8–89.3)
Neg, Predictive Value	92.6 (91.8–93.4)	81.9 (80.8–83.0)	94.2 (93.4–94.9)	88.1 (87.3–88.8)
Positive Predictive Value	22.2 (20.9–23.5)	10.5 (9.5–11.6)	22.6 (21.4–23.9)	32.0 (29.3–34.7)

*Mean cutoff was derived from the 2010 Derivation Population. ** High Score was derived from the 2010 Derivation Population.

Table 6 Calculation of the RRS is for an insured single 51 year-old female who lives with a friend. The patient has congestive heart failure, prior myocardial infarction, moderate to severe liver disease and diabetes, who reports five medications at home, has had four ED visits in the past year, and is being discharged today after a five day hospitalization for an acute problem. She has had no prior hospitalizations this year, and she is now on nine medications.

Variable	Patient's score for variable	Beta Coefficient	Variable score x beta coefficient
Acute Admission (i.e., not scheduled)	1	0.26	0.26
Age	51	0.007	0.357
Male Sex	0	0.28	0
Married	0	-0.13	0
ED Visits within one year (4)	1	0.82	0.82
Inpatient Visits within one year	0	0.39	0
Uninsured	0	.22	0
Length of Stay (5)	1	.23	.23
Medications on the Patient's Inpatient Medication List Two Days Prior to Discharge	9	0.03	0.27
Medications on the Patient's Ambulatory Medication List (5)	1	-0.1	-0.1
Modified Charlson Comorbidity Index (Modified and Age Adjusted) (8)	1	0.68	.68
		RRS	2.517

References

1. Medicare Program; Hospital inpatient prospective payment systems for acute care hospitals and the long-term care hospital prospective payment system and FY 2012 rates. 42 C.F.R. Pt. 412.150 - 412.154 (2011).
2. McCarthy D, Johnson MB, Audet A-M. Recasting readmissions by placing the hospital role in community context. *JAMA* 2013; 309(4): 351-352.
3. Brock J, Mitchell J, Irby K, Stevens B, Archibald T, Goroski A, Lynn J. Association between quality improvement for care transitions in communities and rehospitalizations among medicare beneficiaries. *JAMA* 2013; 309(4): 381-389.
4. Hansen L, Young R, Hinami K, Leung A, Williams M. Interventions to reduce 30-day rehospitalization: a systematic review. *Ann Int Med* 2011; 155: 520-528.
5. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: A systematic review. *JAMA* 2011; 306(15): 1688-1698.
6. Coleman EA, Min SJ, Chomiak A, et al. Posthospital care transitions: patterns, complications, and risk identification. *Health Serv Res* 2004; 39(5): 1449-1465.
7. van Walraven C, Dhalla IA, Bell C, Etchells E, Stiell IG, Zarnke K, Austin PC, Forster AJ. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *CMAJ* 2010; 182(6): 551-557.
8. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; 40(5): 373-383.
9. Schneeweiss S, Wang PS, Avorn J, et al. Improved comorbidity adjustment for predicting mortality in Medicare populations. *Health Serv Res* 2003; 38: 1103-1120.
10. Grunier A, Dhalla IA, van Walraven C, et al. Unplanned readmissions after hospital discharge among patients identified as being at high risk for readmission using a validated predictive algorithm. *Open Medicine* 2011; 5(2): E104.
11. 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. *Med Care* 2010; 48(11): 981-988.
12. 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* 2006; 333(7563): 327.
13. Billings J, Blunt I, Steventon A, Georghiou T, Lewis G, Bardsley M. Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). *BMJ Open* 2012; 00:e001667.doi:10.1136/bmjopen-2012-001667.
14. Hasan O, Meltzer DO, Shaykevich SA, Bell CM, Kaboli PJ, Auerbach AD, Wetterneck TB, Arora VM, Zhang J, Schnipper JL. Hospital readmission in general medicine patients: a prediction model. *J Gen Intern Med* 2009; 25(3): 211-219.
15. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care* 2005; 43(11).
16. Hall W, Ramachandran R, Narayan S, Jani A, Vijayajumar S. An electronic application for rapidly calculating Charlson comorbidity score. *BMC Cancer* 2004; 4: 94.1
17. Garrison G, Mansukhani M, Bohn B. Predictors of thirty-day readmission among hospitalized family medicine patients. *J Am Board Fam Med* 2013; 26: 71-77.
18. Arbaje A, Wolff J, Yu Q, Powe N, Anderson G, Boulton C. Postdischarge environmental and socioeconomic factors and the likelihood of early hospital readmission among community-dwelling Medicare beneficiaries. *Gerontologist* 2008; 48: 495-504.
19. Prins H, Hasman A. Appropriateness of ICD-9 coded diagnostic inpatient hospital discharge data for medical practice assessment: a systematic review. *Methods Inf Med* 2013; S2: 3-17.