Enhanced Prediction of 30-Day Hospital Readmission in Patients with Congestive Heart Failure Using Physiologic and Laboratory Features in Machine Learning Models: Retrospective Cohort Study

Arman Koul, Kirby Gong, Ali Sobhi Afshar, Hadi Kharrazi, Nauder Faraday

Submitted to: JMIR AI
on: April 17, 2024

Table of Contents

Enhanced Prediction of 30-Day Hospital Readmission in Patients with Congestive Heart Failure Using Physiologic and Laboratory Features in Machine Learning Models: Retrospective Cohort Study

Arman Koul[1] MS; Kirby Gong[2] MS; Ali Sobhi Afshar[2] PhD; Hadi Kharrazi[3, 4] MD, PhD; Nauder Faraday[2] MD, MPH

[1]Stanford University School of Medicine Stanford US
[2]Dept of Anesthesiology/Critical Care Medicine Johns Hopkins University School of Medicine Baltimore US
[3]Division of Biomedical Informatics and Data Science Johns Hopkins School of Medicine Baltimore US
[4]Center for Population Health IT Johns Hopkins School of Public Health Baltimore US

Corresponding Author:
Nauder Faraday MD, MPH
Dept of Anesthesiology/Critical Care Medicine
Johns Hopkins University School of Medicine
600 N Wolfe St
Baltimore
US

Abstract

Background: Models to predict hospital readmission in patients with congestive heart failure (CHF), such as those used by the Centers for Medicare & Medicaid Services (CMS), have limited accuracy. Additional features, available from electronic health records (EHR), have potential to improve accuracy of predictive models.

Objective: The goal of this study was to evaluate the ability of demographic, hospital utilization, medication, laboratory, and physiologic variables to improve prediction of 30-day hospital readmission compared to CMS's model that utilizes only age, sex, and diagnostic billing codes.

Methods: We used the Medical Information Database for Intensive Care (MIMIC-IV) to derive features from patients admitted with a diagnosis of CHF. Machine learning methods were used to develop predictive models and accuracy of models compared using area under the receiver operator characteristic curves (AUROC).

Results: 6,604 hospital CHF admissions were identified; 30-day hospital readmission occurred for 24.5% of admissions. Models that included features derived from time-series laboratory (AUROC=0.62/95% CI 0.59-0.64) and physiologic (0.63/0.61-0.65) variables performed significantly better than CMS's base model (0.54/0.51-0.57). These two models included relatively few features related to diagnostic billing codes and a high proportion of laboratory and physiologic features.

Conclusions: The addition of features derived from the EHR, particularly those derived from the laboratory and physiologic variable domains, improved predictive accuracy of 30-day hospital readmission over CMS's base model. Machine learning algorithms that use data readily available from the EHR have potential to improve accuracy of administrative risk adjustment models and reduce bias caused by reliance on diagnostic billing codes.

(JMIR Preprints 17/04/2024:59627)
DOI: https://doi.org/10.2196/preprints.59627

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Arman Koul MS [a], Kirby Gong MS [b], Hadi Kharrazi, MD, PhD [c,d], Ali Sobhi Afshar PhD[b], Nauder Faraday MD, MPH[b]

[a]Stanford University School of Medicine, Palo Alto, CA
[b]Dept of Anesthesiology/Critical Care Medicine, Johns Hopkins University School of Medicine, Johns Hopkins Hospital, 600 N Wolfe St, Baltimore, MD 21287
[c]Center for Population Health IT, Johns Hopkins School of Public Health, 615 N Wolfe St, Baltimore, MD 21205[d]Division of Biomedical Informatics and Data Science, Johns Hopkins School of Medicine

Enhanced Prediction of 30-Day Hospital Readmission in Patients with Congestive Heart Failure Using Physiologic and Laboratory Features in Machine Learning Models: Retrospective Cohort Study

Abstract

Background: Models to predict hospital readmission in patients with congestive heart failure (CHF), such as those used by the Centers for Medicare & Medicaid Services (CMS), have limited accuracy. Additional features, available from electronic health records (EHR), have potential to improve accuracy of predictive models.

Objective: The goal of this study was to evaluate the ability of demographic, hospital utilization, medication, laboratory, and physiologic variables to improve prediction of 30-day hospital readmission compared to CMS's model that utilizes only age, sex, and diagnostic billing codes.

Methods: We used the Medical Information Database for Intensive Care (MIMIC-IV) to derive features from patients admitted with a diagnosis of CHF. Machine learning methods were used to develop predictive models and accuracy of models compared using area under the receiver operator characteristic curves (AUROC).

Results: 6,604 hospital CHF admissions were identified; 30-day hospital readmission occurred for 24.5% of admissions. Models that included features derived from time-series laboratory (AUROC=0.62/95% CI 0.59-0.64) and physiologic (0.63/0.61-0.65) variables performed significantly better than CMS's base model (0.54/0.51-0.57). These two models included relatively few features related to diagnostic billing codes and a high proportion of laboratory and physiologic features.

Conclusions: The addition of features derived from the EHR, particularly those derived from the laboratory and physiologic variable domains, improved predictive accuracy of 30-day hospital readmission over CMS's base model. Machine learning algorithms that use data readily available from the EHR have potential to improve accuracy of administrative risk adjustment models and reduce bias caused by reliance on diagnostic billing codes.

Key Words: Machine Learning; Congestive Heart Failure; Hospital Readmission

Introduction

Readmission to the hospital after inpatient treatment for an acute illness is associated with poor patient outcomes, including increased morbidity and mortality[1]. Congestive heart failure (CHF) is the second most common cause for inpatient admission in the US, with over one million admissions in 2018 and 30-day hospital readmission rates of 17.6%-28.0%[2,3]. Each of these readmissions is estimated to cost approximately $15,732 if the CHF patient is readmitted to the same hospital and $25,879 if the patient is readmitted to a different hospital[4].

A large proportion of these 30-day hospital readmissions are preventable; thus, hospital readmission rates are proposed as a metric of quality of care. In an effort to improve care quality and reduce costs, the Center for Medicare and Medicaid Services (CMS) introduced the Hospital Readmissions Reduction Program (HRRP) in 2012, which penalizes hospitals financially for 30-day readmission rates that exceed the national average for several diagnoses including CHF.

CMS created a Hospital-Wide All-Cause Unplanned Readmission (HWR) model to predict CHF readmission and benchmark hospital performance. This model uses information from billing claims to predict 30-day readmission using CMS Hierarchical condition category (HCC) diagnosis groupings, which are derived from International Classification of Disease (ICD) codes, and simple demographic data. The current CMS model has modest accuracy, with an area under the receiver operator characteristic curve (AUROC) of 0.60[5]. However, with the widespread use of electronic health records (EHR) and rapid advances in machine learning, prior studies have shown the opportunity to improve the accuracy of administrative risk adjustment models[6].

In this study, we derive novel features from data readily available from the EHR and incorporate those features into CMS's HWR model to predict 30-day readmission among patients with CHF using machine learning approaches. We hypothesized that including novel features derived from demographic, hospital utilization, medication, laboratory, and physiologic data in the EHR, would improve the accuracy of CMS's base model for 30-day readmissions in patients with CHF.

Methods

Study Design and Data Source

We performed a retrospective data analysis using the Medical Information Mart for Intensive Care (MIMIC)-IV-2.2 database. MIMIC-IV is a publicly available dataset containing de-identified information extracted from the EHR of patients admitted to the intensive care units (ICUs) of the Beth Israel Deaconess Medical Center in Boston, MA between 2008 and 2019. The dataset

contains 523,740 unique hospital admissions representing 256,878 unique patients at the time of analysis[7]. Approval to use the MIMIC dataset was granted by the MIMIC administrators and this project was reviewed and acknowledged as exempt from human subjects research by the Johns Hopkins Institutional Review Board.

Eligibility Criteria for Inclusion in the Cohort

Subjects were eligible for inclusion based on the following criteria: 1) age ≥ 18 years; 2) hospital ICD billing code of CHF among the top 3 codes for that admission. Subjects were excluded for the following: 1) in-hospital death (i.e., not discharged alive); 2) hospital ICD billing code includes a competing penalty code as first diagnosis (exclusion criteria used by CMS for CHF readmission penalty); 3) physiologic data from the ICU stay during the admission were not available. After applying these criteria, 6,604 index hospital admissions were identified from 5,300 unique individuals (see **Supplement Figure 1**)

Primary Outcome

The primary outcome was unplanned readmission to the hospital within 30 days of the date of discharge from the index hospital admission. To determine whether an admission after the index was planned or not, we used similar logic to CMS guidelines calculating the HWR model. CMS considers certain codes indicators of a planned or elective admission, and we mirrored this logic[8].

Variables and Derived Features for Prediction Modeling

We used 6 distinct types of variables to derive features to incorporate into prediction models: 1) demographic variables; 2) diagnostic codes; 3) hospital and ICU utilization variables; 4) medication variables; 5) laboratory test variables; and 6) physiologic variables. A summary of the variable types, their individual components, and features derived from those elements is shown in **Table 1**.

The following demographic variables were extracted from MIMIC-IV for each hospital admission: age, gender, and weight on admission. Comorbidities related to each hospital admission were determined from ICD10 billing codes (i.e. codes used for insurance billing claims) present on admission. ICD10 codes were converted to HCC condition categories using CMS's published algorithm[9]. The following care utilization variables were determined: duration of the index hospital stay, duration of the ICU length of stay, and number of ICU stays during the index hospitalization.

We used the Electronic Medication Administration Record (EMAR) available in MIMIC to determine usage of several commonly prescribed medications for patients with CHF (see **Table 1**) searching for both generic and brand names. For each medication, we determined whether it was administered (dichotomous) and how much was administered during each of the following intervals: first and last 24 hours of the index hospital stay; and, first and last 24 hours of the ICU stay. For subjects who had more than one ICU stay during an index hospitalization, we used information related to the last ICU stay during that hospitalization.

We extracted data for several common laboratory tests for each hospital admission (see **Table 1** for lab components). We also extracted data for several common physiologic variables routinely recorded by bed-side nurses (see **Table 1** for list of physiologic components). Values for each laboratory and physiologic component were identified from the start to the end of the ICU stay. More than 96% of all subjects had a least one value for each laboratory and physiologic variable (see **Supplement Table 1** for descriptive summary of data related to laboratory and physiologic variables).

Table 1. Variables and features incorporated into models to predict 30-day hospital readmission

Type of variable	Components	Number of variables examined	Number of Features Derived	Number of time periods examined	Total possible features
Demographic characteristics	Age, Sex, Weight at admission	3	1: value of the feature	N/A	3
Diagnostic codes	HCCs groupings derived from ICD codes[a]	141 HCCs	1: binary whether diagnosis code present upon admission	N/A	141
Hospital and ICU Utilization	Length of Admission, Length of all ICU stays, Length of last ICU stay, Number of ICU stays in admission, Length of non-last ICU stays	5 values	1: value of feature	N/A	5
Common Cardiac Medications [b]	Beta blockers; Calcium channel blockers: Digoxin; ACEI/ARB; Diuretics; Insulin; SGLT2 inhibitors; Aspirin; P2Y12 blockers; DOACs	28	2: binary whether medication administered, amount of medication administered	4: first and last 24 hours of ICU and hospital stays	112
Labs	Creatinine, Chloride, Sodium, Potassium, Blood urea nitrogen, HCO3, Anion gap, Glucose, Hematocrit, Platelet Count, Hemoglobin, white blood cell count	12	79: Number of different TsFresh package features[c]	6: First 12, 18, and 24 hours of ICU stay. Last 12, 18, and 24 hours of ICU stay	5,688

Physiological	Heart rate, Respiratory rate, O2 saturation pulse oximetry, GCS – eye opening, GCS – motor response, GCS – verbal response, Blood pressure - mean, Blood pressure – diastolic, Blood pressure - systolic	9	79: Number of different TsFresh package features[c]	6: First 12, 18, and 24 hours of ICU stay. Last 12, 18, and 24 hours of ICU stay	4,266

[a]*HCC codes were derived from the FY2020 CMS crosswalk that maps International Classification of Disease (ICD10) codes to a broader diagnostic grouping defined by an HCC code.*
[b]*ACEI= angiotensin converting enzyme inhibitor; ARB= angiotensin receptor blocker; SGLT2= sodium glucose transporter-2; P2Y12- platelet purinergic receptor; DOAC= direct oral anticoagulant*
GCS= Glasgow coma scale
[c]*These represent the 79 possible features that can be created from the TsFresh package. Some features have parameters that are algorithmically adjusted and optimized when calculating values. The possibility of different parameters was not included in the count of 79 possible features.*

We generated summary features from the time-series recordings for each laboratory and physiologic variable using the Python package, tsfresh[10]. Tsfresh is built on top of the open-source packages Scikit-learn[11] and Pandas[12] and allows for the summary of time-series features. The features, groupings, and descriptions of tsfresh can be found in **Supplement Table 2**. Summary features were derived for the following time intervals: first and last 12-, 16-, and 24-hour periods of a subject's ICU stay during the index hospital admission. For individuals with more than one ICU stay, we only included the summary features derived from the last ICU stay within the index hospital admission.

Statistical Analysis and Machine Learning Methods

We used logistic regression to generate models to predict the primary outcome (i.e. readmission to the hospital within 30-days of hospital discharge). We created a series of six models by starting with the features used in CMS's base model and sequentially adding features from each of the variable types as described below:

- *Model 1*: CMS base model – included age, sex, and comorbidities from HCC codes
- *Model 2*: Model 1+ healthcare utilization features
- *Model 3*: Model 2 + medication features
- *Model 4*: Model 3 + admission weight
- *Model 5*: Model 4 + lab features
- *Model 6*: Model 5 + physiologic features

Previous studies have shown the utility of weight, laboratory, and physiologic features for predicting clinical outcomes[13–15]. As we introduced additional feature categories into sequential

models, feature selection was required to reduce data dimensionality[16]. Commencing with a feature pool of more than 3,500 potential features, we applied various methodologies to narrow down the selection to the most informative features for predicting 30-day hospital readmission. Our primary method was to reduce collinearity using variance inflation factor (VIF)[17]. We continued this iterative process until the maximum value of the VIF was less than 10[18]. As a second approach to reduce data dimensionality, we used Extreme Gradient Boosting (XgBoost)[19].

Each model underwent independent training and testing; 70% of data was allocated for model training and 30% was reserved for testing model performance. We used bootstrapping[20] (n = 1000) to construct 95% confidence intervals for the AUROC values generated for each model. Results are reported as AUROC (c-statistic) with 95% CI from the test set in each predictive model. Analysis was completed using Python 3.8.3.

Results

The cohort consisted of 6,604 index hospital admissions for CHF corresponding to 5,300 unique subjects (**Supplement Fig 1**). Of the 6,604 index admissions, 1,618 (24.5%) resulted in unplanned 30-day hospital readmission. Characteristics of subjects with and without readmission after the index hospitalization are shown in **Table 2**. Subjects readmitted tended to have longer index hospital stays, were more likely to have diabetes (with or without complications), and were more likely to have renal failure. (See **Supplemental Table 3** for full list of comorbidities).

Table 2. Characteristics of subjects by 30-day hospital readmission status

	Non-Readmissions (n=4,986)	**Readmissions (n=1,618)**
Patient Characteristics		
Male, n (%)	2,568 (51.5)	861 (53.2)
Patient Age, years, median (IQR)	72 (61-82)	71 (60-80)
Healthcare Utilization		
Length of Admission, days, median (IQR)	7.04 (4.63-11.59)	7.58 (4.87-12.88)
Length of Last ICU Stay, days, median (IQR)	1.97 (1.11-3.42)	1.99 (1.08-3.53)
Number of ICU Stays, median (IQR)[a]	1 (1-1)	1 (1-1)
Comorbidities[b]		
Specified Heart Arrhythmias, n (%)	2,868 (57.5)	913 (56.4)
Diabetes without Complication, n (%)	1,684 (33.8)	630 (38.9)
Cardio-Respiratory Failure and Shock, Including Respiratory Distress Syndromes, n (%)	1,695 (34.0)	543 (33.6)
Chronic Obstructive Pulmonary Disease, Including Bronchiectasis, n (%)	1,382 (27.7)	468 (28.9)
Unstable Angina and Other Acute Ischemic Heart Disease, n (%)	1,157 (23.2)	396 (24.5)
Diabetes with Chronic Complications, n (%)	858 (17.2)	384 (23.7)
Chronic Kidney Disease, Stage 5, n (%)	329 (6.6)	178 (11.0)

End Stage Renal Disease, n (%)	316 (6.3)	171 (10.6)
Type 1 Diabetes Mellitus, add-on to Diabetes HCCs 19-21, n (%)	297 (6.0)	151 (9.3)
Acute Myocardial Infarction	340 (6.8)	120 (7.4)

a – 602 (9.1%) CHF admissions within the examined cohort had >1 ICU stays. 74 (1.1%) of the CHF admissions within the examined cohort had >2 ICU stays

b – Top 10 most frequent comorbidities are shown in Table 1 (other comorbidities shown in Supplement Table 3.

Figure 1 shows the performance of each of the 6 models in predicting unplanned 30-day hospital readmission. The addition of healthcare utilization features improved the AUROC over CMS's base model; however, this increase was not statistically significant. Inclusion of laboratory features (AUROC 0.62/95%CI 0.59-0.64) and laboratory features plus physiologic features (AUROC 0.63/95%CI 0.61-0.65) led to significant improvements over CMS's base model (AUROC 0.54/0.51-0.57). Model 6 performed statistically significantly better than Models 1, 3, and 4.

Table 3 shows the types of variables and number of features selected in each of the 6 models. Inclusion of laboratory and physiologic features in model generation led to preferential selection of these features over comorbidity features, whose number decreased to only 3 (from 61 in Model 1). The comorbidities retained in Model 6 were related to diabetes and kidney function. In our final model (Model 6), 91% of features were selected from laboratory and physiologic variables. All models retained two or more demographic characteristics (i.e., age, sex, weight) and all included healthcare utilization features when given the opportunity to select from those features.

Table 3. Count of Features Selected by Variable Type for Predictive Models of Hospital Readmission in Patients with CHF[a]

	Variable Type (Number of Features)						
Model	**Diagnosis N = 111**	**Demographic N = 2**	**Hospital & ICU Utilization N = 5**	**Medications N = 151**	**Weight N = 1**	**Labs N = 5,688**	**Physiol al N = 4,**
Model 1 (Diagnostic Codes and Demographic Variables)	61	2	N/A[b]	N/A[b]	N/A[b]	N/A[b]	N/A
Model 2 (Model 1 + Hospital & ICU Utilization Variables)	47	2	4	N/A[b]	N/A[b]	N/A[b]	N/A

Model 3 (Model 2 + Medication Variables)	31	2	4	47	N/A[b]	N/A[b]	N/A
Model 4 (Model 3 + Weight Variable)	34	2	3	42	1	N/A[b]	N/A
Model 5 (Model 4 + Labs-Derived Variables)	3	2	3	9	1	76	N/A
Model 6 (Model 5 + Physiological Reading-Derived Variables	3	1	2	8	1	67	97

[a]*Models were created by iteratively adding features to Model 1 which was meant to emulate the CMS HWR m by only using age, sex, and diagnostic variables. The table shows each of the models and the count of feat selected using an XGBoost importance factor filter. Variable Inflation Factors were used to remove colli features.*

[b]*N/A denotes features that were not allowed to be included in the model.*

The 10 most consequential features selected in Model 6 are shown in **Figure 2**. All were laboratory or physiologic features. The lab features selected included components related to platelet count, white blood cell count, creatinine, and sodium. The physiologic features selected included components related to heart rate, motor response, and pulse oximetry saturation. The model included a total of 180 features, many derived from TsFresh, including 52 statistical summary features, 26 signal frequency analysis features, 25 signal frequency/time analysis features, and 22 trends/change analysis features (**Supplement Table 4**). A list of all features selected in Model 6 and their weights is shown on **Supplement Table 5**.

Discussion

We developed and internally validated novel models to predict 30-day hospital readmission in patients with CHF. Using an ICU-based cohort from the MIMIC-IV dataset, we demonstrate that including laboratory and physiologic variables into machine learning-derived prediction models improves accuracy compared to CMS's model, which is based on age, sex, and diagnostic codes only. Our results also suggest that the laboratory and

physiologic feature spaces contain information of higher priority in predicting 30-day readmission among CHF patients than information contained in diagnostic billing codes.

Thirty-day hospital readmission occurred in 24.5% of subject-admissions in MIMIC-IV's ICU-based cohort diagnosed with CHF. This rate of readmission is within the upper range of previous reports[2,3]. Because ICU patients are more severely ill than general hospital admissions, we expected the readmission rates to be higher than for an unselected group of hospitalized patients with CHF. Using the features in CMS's HWR prediction model, we observed an AUROC of 0.54 (0.51-0.57) for predicting hospital readmission in the MIMIC-IV cohort. Lower accuracy than previously reported for CMS's model (0.60 AUROC) is likely because our cohort contained only ICU patients with CHF, whereas CMS's model was derived from all patients hospitalized with CHF. We note that CMS's HWR prediction model does not account for ICU admission status even though hospital admissions that require ICU level care are subject to the same HRRP penalty. Interestingly, inclusion of simple healthcare utilization features (e.g. durations of hospital and ICU length of stay) increased the AUROC over CMS's base model, although this improvement was not statistically significant in our study. We also note that all our predictive models include basic demographic information (i.e., age, sex, and/or admission weight) readily available from EHRs.

In our analyses, the most substantive improvements in predictive accuracy came after including features derived from laboratory and physiologic variables. Models that included features from these variable types had significantly higher AUROC compared to CMS's base model. Models 5 and 6 included many features selected from the laboratory and physiologic variable types and only 3 diagnostic comorbidities. This contrasts sharply with the CMS base model, which included 61 diagnostic groups. Our feature selection process was designed to limit collinearity in feature selection; therefore, these results suggest that information present in the diagnostic comorbidity codes overlaps with information present in the laboratory and physiologic features. All the laboratory and physiologic variables that were used to develop our predictive models are routinely recorded in the EHRs of patients admitted to an ICU in the US. Moreover, all the variables (with possible exception of GCS) are captured in the EHRs of general hospital patients with CHF without ICU admission, although the frequency of these variable recordings is lesser on the wards than in the ICU.

The most influential features derived from the laboratory variables were related to sodium, creatinine, WBC count, and platelet count. Electrolyte abnormalities are common in patients with CHF, particularly in the setting of diuretic use[21]; thus, the high importance of sodium and creatinine was not surprising. WBC count was also not surprising given that it is a marker of acute stress/illness and that infection can worsen CHF[22]. Platelet count also reflects severity of illness, particularly among hospitalized patients[23]. The most influential physiologic features were related to oxygen saturation, heart rate, and motor responses. All these variables are clinical indicators of severity of cardiopulmonary and neurologic illness.

Inclusion of common cardiac medications (Model 4) did not significantly improve prediction over CMS's base model; however, several medication-related features were included in Models 5 and 6. The most influential of these features was total insulin dose

administered in the last 24 hours of hospitalization. Given that diagnostic codes for diabetes were among the very few that were retained in our machine learning models, diabetes appears to be a particularly important risk factor for readmission in patients with CHF. Interestingly, insulin dosage was more predictive (beta coefficient =0.111, Supplement Table 4) of readmission in Model 6 than diagnosis of type 1 diabetes (beta coefficient= 0.057) and diabetes with other complications (beta coefficient= 0.029). These data suggest that the severity of diabetes, as indicated by the amount of insulin administered, contributes more strongly to the prediction of CHF readmission than diagnosis of diabetes per se.

A potential advantage of using laboratory and physiologic variables for administrative risk adjustment modeling is lesser bias compared to using diagnostic billing codes. ICD coding is subjective and prone to wide variability, leading some to question its validity[24] and potential bias in predictive models[25]. Ibrahim and colleagues reported that changes in coding practices contributed substantively to reductions in readmissions after CMS implemented the HRRP program, rather than a real improvement in quality of care[26]. In contrast, laboratory results and physiologic variables are objective measures of patient status and less prone to manipulation.

This study has several limitations. Our predictive models were derived entirely from individuals admitted to the ICU. Patients admitted to the ICU represent a select population of the sickest hospitalized patients, thus our results may not generalize to all patients with CHF. At a minimum, our data suggest that more accurate risk modeling is possible for this subset of patients with CHF. Feature extraction from laboratory and physiologic variables was limited to the first and last 24 hours of the ICU stay (and subintervals of those periods). We chose to focus on these time intervals because of their clinical importance as exemplified by historical use in ICU risk scoring systems like APACHE II[27] and because the average length of ICU stay was ~48 hours. Our methods may have missed important information that could have been extracted by examining the entire duration of the ICU stay. Furthermore, we only used physiologic data that was recorded by bedside nurses in the EHR. We did not use the high frequency physiologic data available on a subset of these patients in MIMIC. These high frequency data may provide additional information useful for risk prediction that is worthy of future investigations[28]. Finally, we used the tsfresh package to extract summary features from the laboratory and physiologic data. This shotgun approach yielded a final model (Model 6) that included a large number of features mathematically derived from TsFresh. While the statistical summary features (e.g. mean, median, standard deviation) may be intuitive to interpret, some of the other features (e.g. signal frequency analysis) are difficult to conceptualize. We acknowledge that other mathematical methods to summarize data may be superior and clinicians may be resistant to deploying a model that is challenging to interpret clinically.

In summary, this study demonstrates that including novel features, derived from medication, laboratory, and physiologic data in the EHR, improves prediction of 30-day hospital readmission in patients with CHF. Our study provides proof of concept that data that is routinely available from the EHR (and less subject to bias than billing codes) can be used to improve the accuracy of administrative risk adjustment models.

References

1. McIlvennan CK, Eapen ZJ, Allen LA. Hospital Readmissions Reduction Program. *Circulation*. 2015;131(20):1796-1803. doi:10.1161/CIRCULATIONAHA.114.010270

2. Overview of Clinical Conditions With Frequent and Costly Hospital Readmissions by Payer, 2018 #278. Accessed October 13, 2023. https://hcup-us.ahrq.gov/reports/statbriefs/sb278-Conditions-Frequent-Readmissions-By-Payer-2018.jsp

3. Ziaeian B, Fonarow GC. The Prevention of Hospital Readmissions in Heart Failure. *Prog Cardiovasc Dis*. 2016;58(4):379-385. doi:10.1016/j.pcad.2015.09.004

4. Urbich M, Globe G, Pantiri K, et al. A Systematic Review of Medical Costs Associated with Heart Failure in the USA (2014–2020). *PharmacoEconomics*. 2020;38(11):1219-1236. doi:10.1007/s40273-020-00952-0

5. Keenan PS, Normand SLT, Lin Z, et al. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. *Circ Cardiovasc Qual Outcomes*. 2008;1(1):29-37. doi:10.1161/CIRCOUTCOMES.108.802686

6. Weissman GE, Joynt Maddox KE. Guiding Risk Adjustment Models Toward Machine Learning Methods. *JAMA*. 2023;330(9):807-808. doi:10.1001/jama.2023.12920

7. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023;10(1):1. doi:10.1038/s41597-022-01899-x

8. Measure Information Form. Published online 2016.

9. 2022 Model Software/ICD-10 Mappings | CMS. Accessed October 14, 2023. https://www.cms.gov/medicarehealth-plansmedicareadvtgspecratestatsrisk-adjustors/2022-model-softwareicd-10-mappings

10. Christ M, Braun N, Neuffer J, Kempa-Liehr AW. Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing*. 2018;307:72-77. doi:10.1016/j.neucom.2018.03.067

11. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res*. 2011;12(85):2825-2830.

12. McKinney W. Data Structures for Statistical Computing in Python. In: ; 2010:56-61. doi:10.25080/Majora-92bf1922-00a

13. Nagy R, Gede N, Ocskay K, et al. Association of Body Mass Index With Clinical Outcomes in Patients With Cystic Fibrosis: A Systematic Review and Meta-analysis. *JAMA Netw*

Open. 2022;5(3):e220740. doi:10.1001/jamanetworkopen.2022.0740

14. Blanco N, Leekha S, Magder L, et al. Admission Laboratory Values Accurately Predict In-hospital Mortality: a Retrospective Cohort Study. *J Gen Intern Med*. 2020;35(3):719-723. doi:10.1007/s11606-019-05282-2

15. Toloui A, Madani Neishaboori A, Rafiei Alavi SN, et al. The Value of Physiological Scoring Criteria in Predicting the In-Hospital Mortality of Acute Patients; a Systematic Review and Meta-Analysis. *Arch Acad Emerg Med*. 2021;9(1):e60. doi:10.22037/aaem.v9i1.1274

16. Velliangiri S, Alagumuthukrishnan S, Thankumar joseph SI. A Review of Dimensionality Reduction Techniques for Efficient Computation. *Procedia Comput Sci*. 2019;165:104-111. doi:10.1016/j.procs.2020.01.079

17. Kutner MH, ed. *Applied Linear Statistical Models*. 5. ed. McGraw-Hill Irwin; 2005.

18. Marcoulides KM, Raykov T. Evaluation of Variance Inflation Factors in Regression Models Using Latent Variable Modeling Methods. *Educ Psychol Meas*. 2019;79(5):874-882. doi:10.1177/0013164418817803

19. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. Association for Computing Machinery; 2016:785-794. doi:10.1145/2939672.2939785

20. Tibshirani BE RJ. *An Introduction to the Bootstrap*. Chapman and Hall/CRC; 1994. doi:10.1201/9780429246593

21. Mullens W, Damman K, Harjola VP, et al. The use of diuretics in heart failure with congestion — a position statement from the Heart Failure Association of the European Society of Cardiology. *Eur J Heart Fail*. 2019;21(2):137-155. doi:10.1002/ejhf.1369

22. Bales AC, Sorrentino MJ. Causes of congestive heart failure. Prompt diagnosis may affect prognosis. *Postgrad Med*. 1997;101(1):44-49, 54-56. doi:10.3810/pgm.1997.01.141

23. McMahon CM, Cuker A. Hospital-acquired thrombocytopenia. *Hosp Pract 1995*. 2014;42(4):142-152. doi:10.3810/hp.2014.10.1151

24. Surján G. Questions on validity of International Classification of Diseases-coded diagnoses. *Int J Med Inf*. 1999;54(2):77-95. doi:10.1016/S1386-5056(98)00171-3

25. Wang HE, Landers M, Adams R, et al. A bias evaluation checklist for predictive models and its pilot application for 30-day hospital readmission models. *J Am Med Inform Assoc*. 2022;29(8):1323-1333. doi:10.1093/jamia/ocac065

26. Ibrahim AM, Dimick JB, Sinha SS, Hollingsworth JM, Nuliyalu U, Ryan AM. Association of Coded Severity With Readmission Reduction After the Hospital Readmissions

Reduction Program. *JAMA Intern Med*. 2018;178(2):290-292. doi:10.1001/jamainternmed.2017.6148

27. Lee H, Lim CW, Hong HP, Ju JW, Jeon YT, Hwang, JW, Park HP. Efficacy of the APACHE II score at ICU discharge in predicting post-ICU mortality and ICU readmission in critically ill surgical patients. *Anaesth Intensive Care.* 2015;43(2):175-86

28. Afshar AS, Li Y, Chen Z, et al. An exploratory data quality analysis of time series physiologic signals using a large-scale intensive care unit database. *JAMIA Open*. 2021;4(3):ooab057. doi:10.1093/jamiaopen/ooab057

Figure Legends

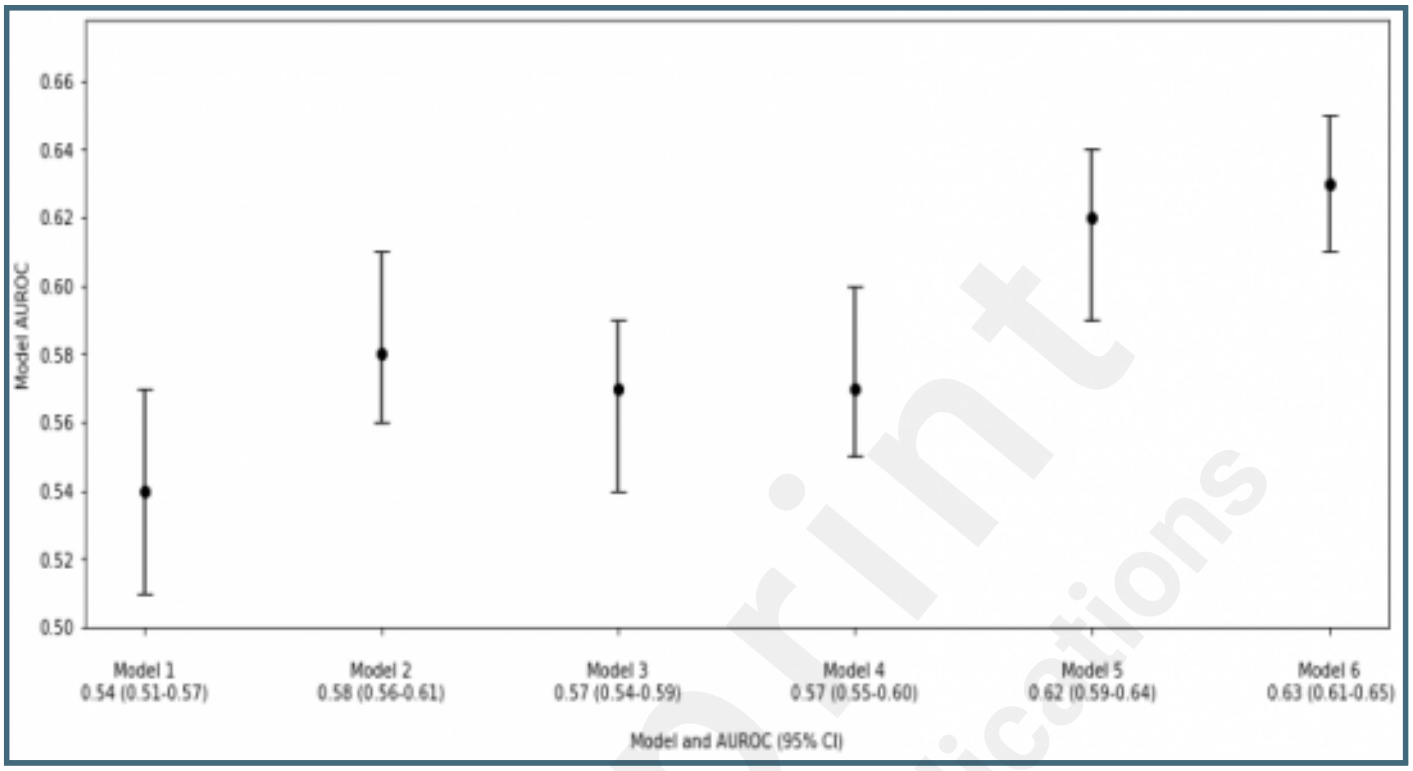
Figure 1. Performance of models (AUROC c-statistic with 95% confidence intervals) *to predict 30-day* hospital readmission in patients with CHF

Figure 2. Logistic regression weights of the top 10 features in model 6 to predict 30-day readmission.
[1] *Fourier coefficient n=0 of the one-dimensional discrete fourier transform for real input by fast fourier transform algorithm.*
[2] *Mean of absolute change of time series values created by a corridor with a lower quantile of 0.2 and an upper quantile of 0.6.*
[3] *Variance of the absolute fourier transform spectrum.*
[4] *Sum over the absolute value of consecutive changes in the series.*
[5] *Fourier coefficient n=1 of the one-dimensional discrete fourier transform for real input by fast fourier transform algorithm.*
[6] *Values of the Rickler wavelet with coefficient 9 and width 5.*
[7] *CID of the series which is a proxy for time series complexity.*
[8] *Values of the Rickler wavelet with coefficient 0 and width 2.*
[9] *C3 measurement of non-linearity in the time series with a lag operator of 1.*
[10] *Sum over the absolute value of consecutive changes in the series.*

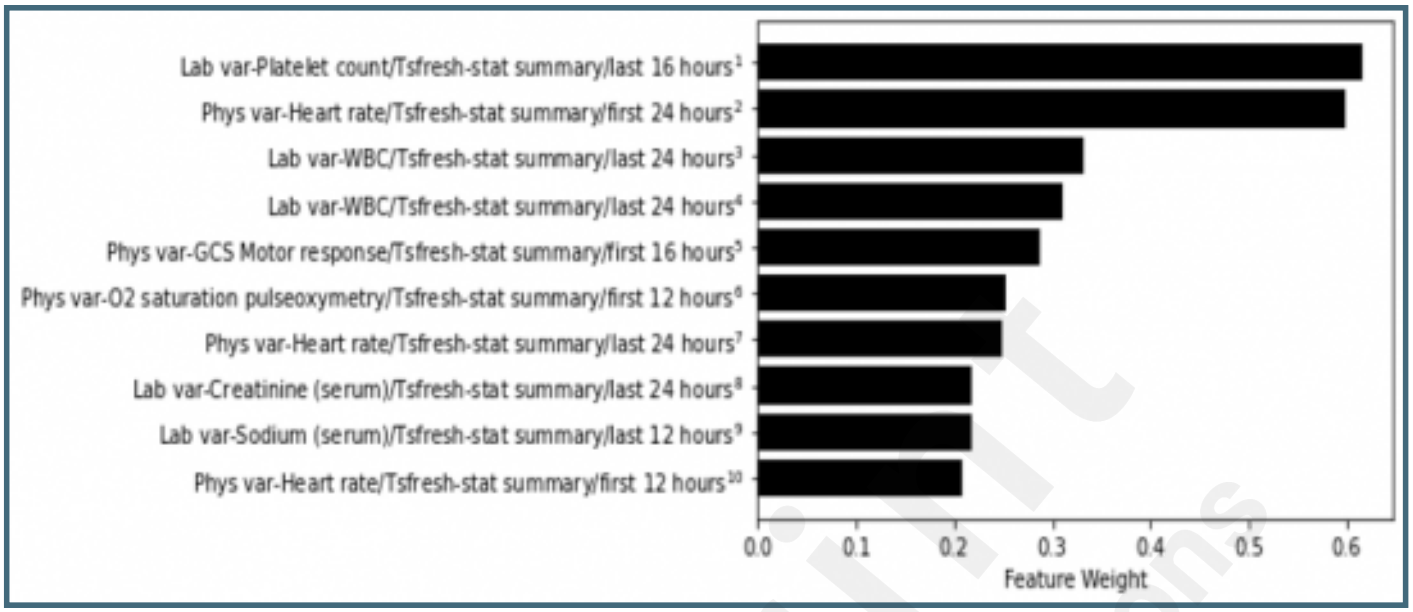
Supplementary Files

Figures

Performance of models (AUROC c-statistic with 95% confidence intervals) to predict 30-day hospital readmission in patients with CHF.



Logistic regression weights of the top 10 features in model 6 to predict 30-day readmission.



Multimedia Appendixes

Untitled.
URL: http://asset.jmir.pub/assets/077f2ceb64223ff67b1e22463e0215ea.docx

Feature weights for Model 6.
URL: http://asset.jmir.pub/assets/025898c3bc5827d315a12d868b98c2f4.xlsx

CONSORT (or other) checklists

CREMLS Checklist for Project.
URL: http://asset.jmir.pub/assets/31d0bc5a55a6fbf7b84793f699be3fa7.pdf