

Original Paper

Readmission Risk Trajectories for Patients With Heart Failure Using a Dynamic Prediction Approach: Retrospective Study

Wei Jiang¹, PhD; Sauleh Siddiqui¹, PhD; Sean Barnes², PhD; Lili A Barouch³, MD; Frederick Korley⁴, PhD, MD; Diego A Martinez⁵, PhD; Matthew Toerper⁵, BSc; Stephanie Cabral⁶, BME; Eric Hamrock^{7,8}, MBA; Scott Levin⁵, PhD

¹Department of Civil Engineering, Johns Hopkins System Institute, Johns Hopkins University, Baltimore, MD, United States

²Department of Decision, Operations & Information Technologies, Robert H Smith School of Business, University of Maryland, College Park, MD, United States

³Department of Medicine, Division of Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD, United States

⁴Department of Emergency Medicine, University of Michigan, Ann Arbor, MI, United States

⁵Department of Emergency Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

⁶Department of Epidemiology & Public Health, University of Maryland, College Park, MD, United States

⁷Innovation and Continuous Improvement Department, Howard County General Hospital, Columbia, MD, United States

⁸StoCastic, LLC, Towson, MD, United States

Corresponding Author:

Wei Jiang, PhD

Department of Civil Engineering, Johns Hopkins System Institute

Johns Hopkins University

3400 N Charles St

Baltimore, MD

United States

Phone: 1 4432404268

Email: wjiang1990@gmail.com

Abstract

Background: Patients hospitalized with heart failure suffer the highest rates of 30-day readmission among other clinically defined patient populations in the United States. Investigation into the predictability of 30-day readmissions can lead to clinical decision support tools and targeted interventions that can help care providers to improve individual patient care and reduce readmission risk.

Objective: This study aimed to develop a dynamic readmission risk prediction model that yields daily predictions for patients hospitalized with heart failure toward identifying risk trajectories over time and identifying clinical predictors associated with different patterns in readmission risk trajectories.

Methods: A two-stage predictive modeling approach combining logistic and beta regression was applied to electronic health record data accumulated daily to predict 30-day readmission for 534 hospital encounters of patients with heart failure over 2750 patient days. Unsupervised clustering was performed on predictions to uncover time-dependent trends in readmission risk over the patient's hospital stay. We used data collected between September 1, 2013, and August 31, 2015, from a community hospital in Maryland (United States) for patients with a primary diagnosis of heart failure. Patients who died during the hospital stay or were transferred to other acute care hospitals or hospice care were excluded.

Results: Readmission occurred in 107 (107/534, 20.0%) encounters. The out-of-sample area under curve for the 2-stage predictive model was 0.73 (SD 0.08). Dynamic clinical predictors capturing laboratory results and vital signs had the highest predictive value compared with demographic, administrative, medical, and procedural data included. Unsupervised clustering identified four risk trajectory groups: decreasing risk (131/534, 24.5% encounters), high risk (113/534, 21.2%), moderate risk (177/534, 33.1%), and low risk (113/534, 21.2%). The decreasing risk group demonstrated change in average probability of readmission from admission (0.69) to discharge (0.30), whereas the high risk (0.75), moderate risk (0.61), and low risk (0.39) groups maintained consistency over the hospital course. A higher level of hemoglobin, larger decrease in potassium and diastolic blood pressure from admission to discharge, and smaller number of past hospitalizations are associated with decreasing readmission risk ($P < .001$).

Conclusions: Dynamically predicting readmission and quantifying trends over patients' hospital stay illuminated differing risk trajectory groups. Identifying risk trajectory patterns and distinguishing predictors may shed new light on indicators of readmission and the isolated effects of the index hospitalization.

(*JMIR Med Inform* 2019;7(4):e14756) doi: [10.2196/14756](https://doi.org/10.2196/14756)

KEYWORDS

heart failure; patient readmission; forecasting; machine learning

Introduction

Background

Patients hospitalized with heart failure suffer the highest rates of 30-day readmission among other clinically defined patient populations in the United States [1]. National efforts to prevent avoidable hospitalizations have led to the adoption of 30-day readmission as a publicly reported performance measure linked to Medicare patient reimbursement [2]. This has motivated much investigation into the predictability of 30-day in-hospital readmissions to guide targeted interventions that could reduce risk. These investigations can lead to clinical decision support tools that can be made available to care providers to guide them in individual patient care. Clinical and administrative data available in-hospital electronic health records (EHRs) and clinical registries have been the primary data sources for these evaluations. Previous studies have applied traditional statistical- and machine learning-based methods for predicting 30-day readmission with varied success for cohorts of patients with heart failure; predictive performance measured as the area under the receiver operating characteristic curve (area under curve [AUC]) has ranged from 0.55 to 0.76 [3-8].

These past readmission prediction models were static—designed to function at a single point in time—either at the beginning of the patient's stay at admission where data are limited [3] or near discharge where all clinical and administrative data may become available for each patient [4-7]. Readmission predictions that occur at admission are useful in creating awareness of risk early in the hospitalization to trigger interventions that may be performed in-hospital and allow for anticipation of postacute care needs. However, these models tend to have lower accuracy because of the lack of availability of meaningful clinical data for prediction. Alternatively, readmission models that operate near discharge are more accurate but may be more difficult to integrate into the workflow and limit the time to perform any in-hospital-based interventions. In either case, these predictive models are static, not designed to determine how readmission risk may fluctuate as a result of clinical findings and hospital interventions. As a result, they do not lend themselves to be enhanced into dynamic decision support tools and are often not applied in general hospital settings. The motivation for this study rests on developing a tool that can eventually be used by care providers to assist in medical decision making. The study is a first step in developing an implementation toward decision support in treatment for patients with heart failure.

In this study, we dynamically predicted a 30-day readmission risk on each day of the patient's hospitalization assuming that the patient was discharged on that day. The reason to predict this counterfactual outcome is that, by definition, 30-day

readmission is readmission within 30 days after discharge. Therefore, a discharge time must be defined while predicting 30-day readmission risk. The current day is most appropriate as the discharge day for real-time prediction as care providers are most interested in learning the readmission risk if the patient is to be discharged that day. We can then keep updating this risk daily for the duration of stay. This may be able to aid physicians in their decision on when to discharge a patient.

Objective

The objective of this study was to develop a predictive model of 30-day readmission for patients with heart failure that functions in real time over the course of a patient's hospitalization. We made predictions for the risk of 30-day readmission as if the patient were discharged on each day of their hospitalization to illustrate how this risk varied over time (if at all) for each patient. We hypothesized that quantifying time-dependent trends in readmission risk had the potential to illuminate the effects of clinical measures and interventions on readmission likelihood at discharge. These predictions could potentially support physician decision making on when to safely discharge patients from the hospital based on the current level of risk and trend over time. These analyses further enabled the identification of groups of patients with heart failure having differing trajectories of readmission risk over their hospital stay.

Methods

Setting and Data

We conducted a retrospective cohort study of patients with a primary diagnosis of heart failure between September 1, 2013, and August 31, 2015, from a community hospital in Maryland. Patients with heart failure were identified using the International Classification of Diseases—ninth revision codes: 428.x, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, and 404.93 [9,10]. Patients who died during the hospital stay or were transferred to other acute care hospitals or hospice care were excluded. The final cohort consisted of 534 encounters totaling 2750 patient days (median: 4 days; interquartile range: 4 days). Data collected as part of routine patient care were extracted from the hospital EHR system for our readmission outcome and predictor variables.

As per the Centers for Medicare and Medicaid Services definition, we defined our primary outcome 30-day readmission as all-cause readmission to the same hospital within 30 days after discharge from the index hospitalization [2].

We derived hypothesized predictors from heart failure and readmission risk factors used in previous studies as summarized in Table 1 [5,11-16]. Predictor data may be conceptually

grouped into static (unchanged over the hospital stay) or dynamic (fluctuating over the hospital stay) categories. Static predictors comprised demographics (eg, age, gender, and race), socioeconomic status (eg, insurance, marital status, and ZIP code), health care utilization (eg, discharge disposition and

number of visits in the last 6 months), and clinically categorized chief complaint [17,18]. Charlson comorbidity index at admission was computed using past medical history (ie, active problems) [12,13] and admission diagnoses.

Table 1. Descriptive summary of patient characteristics for each 30-day readmission outcome.

Static predictors	Not readmitted (n=427)	Readmitted (n=107)	P value
Age (years), mean, median (IQR ^a)	74.5, 77 (66-85)	75.2, 78 (67-87)	.37
Gender: female, n (%)	223 (52.2)	64 (59.8)	.19
Marital status, n (%)			.81
Married	168 (39.3)	44 (41.1)	
Single	68 (15.9)	19 (17.8)	
Widowed	147 (34.4)	36 (33.6)	
Race, n (%)			.96
White	258 (60.4)	65 (60.7)	
Black or African American	121 (28.3)	31 (29.0)	
Asian	36 (8.4)	9 (8.4)	
Insurance, n (%)			.46
Medicare	300 (70.2)	80 (74.8)	
Commercial	92 (21.5)	21 (19.6)	
Medicaid	13 (3.0)	4 (3.7)	
Other	22 (5.2)	2 (1.9)	
Discharge disposition, n (%)			.35
Home or self-care	276 (64.6)	65 (60.7)	
Skilled nursing facility	62 (14.5)	13 (12.1)	
Home-health care service	49 (11.5)	16 (15.0)	
Rehabilitation facility	17 (4.0)	6 (5.6)	
Short-term hospital	8 (1.9)	2 (1.9)	
Nursing facility	5 (1.2)	5 (4.7)	
Number of past visits to hospital, mean, median (IQR ^a)	1.4, 1 (0-2)	2.2, 1 (0-3)	.002
Chief complaints, n (%)			.91
Shortness of breath	282 (66.0)	64 (59.8)	
Chest pain	26 (6.1)	7 (6.5)	
Edema	17 (4.0)	6 (5.6)	
Weakness	15 (3.5)	3 (2.8)	
Lower respiratory tract infection	12 (2.8)	2 (1.9)	
Abdominal pain	9 (2.1)	3 (2.8)	
General	7 (1.6)	1 (0.9)	
Altered mental status	6 (1.4)	1 (0.9)	
Genitourinary	5 (1.2)	2 (1.9)	
Blunt trauma	5 (1.2)	3 (2.8)	
ZIP code	— ^b	—	—
Diagnoses	—	—	—
Diagnoses history	—	—	—
Dynamic predictors			
Elapsed length of stay, mean, median (IQR ^a)	4.8, 3.7 (2.2-5.8)	6.4, 4.2 (2.6-7.3)	.02
Laboratory test, mean, median (IQR^a)			
Alanine transaminase (units/L)	32.0, 20.0 (13.3-30.0)	28.0, 18.0 (12.0-28.6)	.24

Static predictors	Not readmitted (n=427)	Readmitted (n=107)	P value
Aspartate transaminase (units/L)	32.7, 24.0 (18.0-35.0)	32.3, 21.8 (16.9-30.0)	.19
Blood urea nitrogen (mg/dL)	28.5, 25.0 (18.3-35.3)	34.2, 30.8 (20.0-44.7)	.006
Creatinine (mg/dL)	1.6, 1.2 (1.0-1.6)	1.9, 1.4 (1.0-2.0)	.004
Hemoglobin (mg/dL)	11.4, 11.2 (9.9-12.7)	10.8, 10.4 (9.1-12.3)	.003
Potassium (mmol/L)	4.2, 4.1 (3.9-4.4)	4.2, 4.1 (3.9-4.4)	.83
Sodium (mmol/L)	139.0, 139.0 (137.0-141.5)	137.7, 137.8 (134.8-140.9)	.003
pro B-type natriuretic peptide (pg/mL)	8182.5, 4283.0 (2137.0-8656.0)	8683.3, 4262.0 (2013.0-9545.0)	.92
Troponin T (pg/mL)	37.1, 10.0 (10.0-30.0)	56.1, 14.2 (10.0-52.7)	.14
Vital signs			
Systolic blood pressure	—	—	—
Diastolic blood pressure	—	—	—
Temperature	—	—	—
Respiratory rate	—	—	—
Pulse	—	—	—
Peripheral capillary oxygen	—	—	—
Weight	—	—	—
Medication order	—	—	—
Procedure	—	—	—

^aIQR: interquartile range.

^bNot applicable (indicates categorical predictors with too many values or longitudinal variables with repeated measurements that cannot be easily summarized in a single table cell).

Dynamic predictors comprised variables dependent on the time of prediction and included the elapsed length of stay, vital signs (systolic blood pressure, diastolic blood pressure, temperature, respiratory rate, pulse, and peripheral capillary oxygen), and laboratory results (alanine transaminase [ALT]; aspartate transaminase [AST]; blood urea nitrogen [BUN]; creatinine; hemoglobin; potassium; sodium; pro B-type natriuretic peptide [proBNP]; and troponin T). Time series characteristics of vital signs and laboratory results were engineered as predictors in the following manner:

1. Number of measurements: n.
2. Average value
3. Standard deviation
4. Minimum value
5. Maximum value
6. Normalized index of the minimum: index of minimal value / n
7. Normalized index of the maximum: index of maximal value / n
8. Average of last 3 measurements
9. Average of first-order difference

Normalized index predictors were designed to capture the relative point during hospitalization when these extreme measurements occurred [19]. For example, if 10 temperature measurements have been observed up to a particular prediction time, and the sixth measurement is where the minimum occurred, the value of the normalized index is equal to $6/10=0.6$.

All laboratory tests were not necessarily available for each encounter in the database. For example, 70 out of the 534 encounters did not have proBNP measurement data, 55 did not have AST data, and 37 did not have troponin T data. We excluded laboratory-related predictors with more than 30% unavailable data and imputed missing values with the encounter-based mean for features with less than 30% missing data. Predictors quantifying the number of measurements (ie, lab values) were created from the hypothesis that increased measurement intensity might relate to severity of illness and likelihood of readmission [20]. This also enabled inclusion of binary indicators for missing laboratory values which might have meaning.

Daily Prediction Model

Our goal was to build a daily prediction model structured to yield a prediction of 30-day readmission risk each day (6 am) during the patients' stay based on data available at the time of prediction. Thus, predictions made earlier in the stay are based on less information than predictions made later. We generated predictions at the time of hospital admission, then 6 am each subsequent day, and finally at hospital discharge.

The challenge for building a daily prediction model is that the daily outcomes are unknown. Thus, for a particular day, we do not observe whether or not a patient would be readmitted if they were discharged earlier or later than their observed discharge date. To address this uncertainty, we used a 2-stage modeling approach. In the first stage, we estimated the daily risk of readmission based on knowledge of the full encounter data. In

the second stage, we used this estimated risk as the outcome for daily prediction using only the data available up to that particular prediction time. We used the first-stage predictions as a proxy for the unobservable outcomes in the second stage.

Specifically, in the first stage, we trained a logistic regression model for 30-day readmission based on all of the information available for each encounter at the time of discharge. Backward selection was used to eliminate predictors without predictive value in the first-stage model. Each observation for this training data corresponded to a patient encounter (ie, encounter-level model).

In the subsequent second stage, we trained a beta regression model to predict the daily risk of 30-day readmission based on the risk of readmission predicted (on the same day) by the first-stage model. We call this the *counterfactual readmission risk* as we are estimating this risk assuming that the patients were discharged earlier than their actual observed discharge time. Each observation in the training data for the second-stage model corresponds to a patient day that comprises the cumulative data available up to the prediction time (eg, 6 am). An example of the second-stage training data is shown in [Multimedia Appendix 1](#), which includes the same set of predictors as those used in the first stage. However, these predictors were structured differently to exclusively account for the information available at the daily prediction time. For example, the elapsed length of stay and normalized index of minimal potassium are constant for each encounter in the first-stage training data, but dynamically change for each observation in the second-stage (daily) training data. In particular, consider the situation when the minimal potassium measurement occurred on day 3 for a patient whose length of stay is 6 days. For this patient, the normalized index of minimal potassium would be 0.5 in the first-stage training data and 1 in the second-stage training data for the day 3 prediction.

Readmission risk from the first-stage model output was bounded between 0 and 1 and followed a beta distribution (Kolmogorov-Smirnov goodness-of-fit test $P > .05$) [21]. Thus, beta regression was applied in the second stage to predict daily readmission risk from the first-stage output. For the readmission

risk at discharge, the second-stage model used the estimated risk from the first-stage model as a representation of the actual risk of readmission at that point in time.

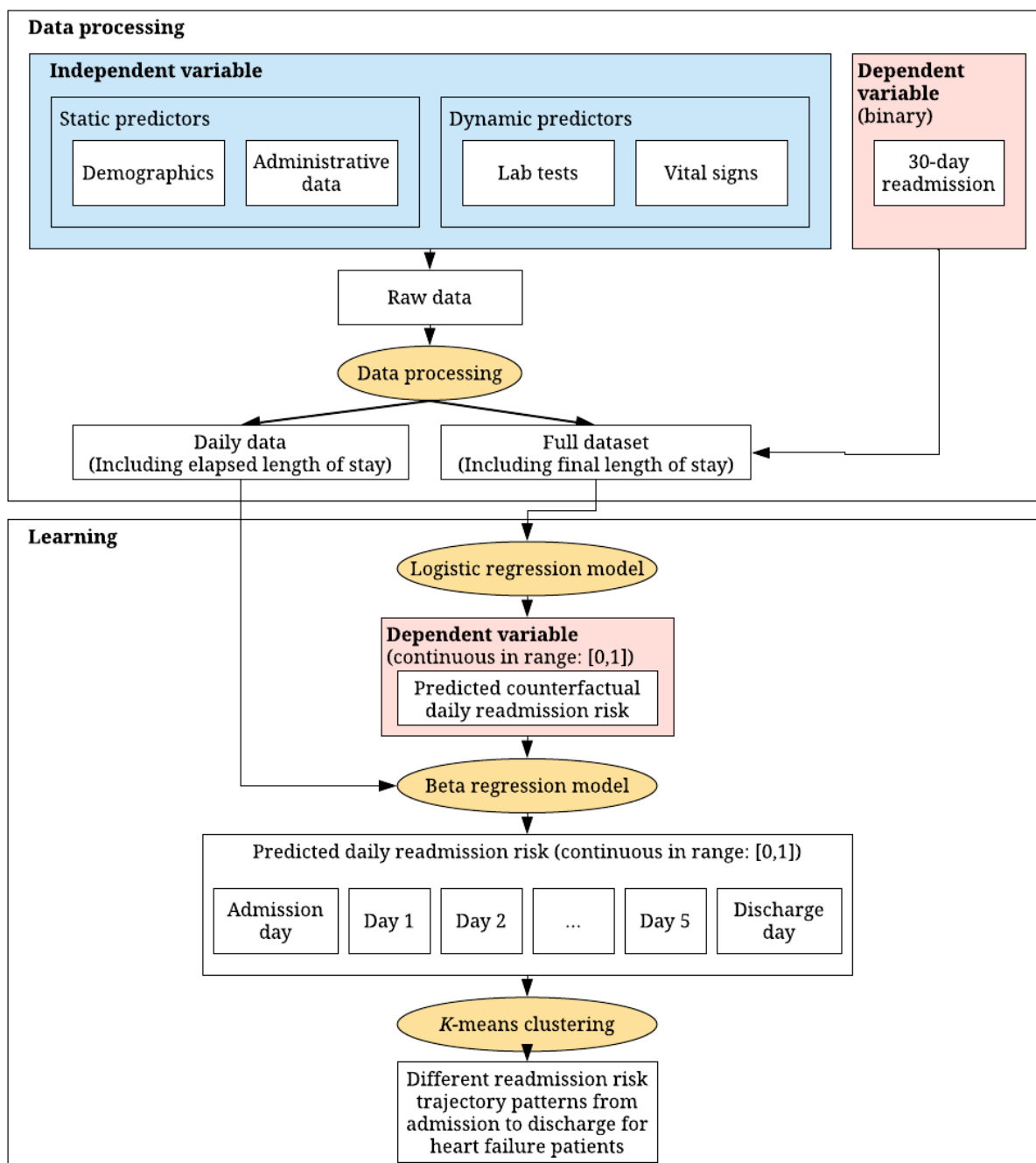
Both the first-stage and second-stage models were developed using 5-fold cross-validation (427/534, 79.9% training and 107/534, 20.0% test). Predictive performance measures of the AUC for the first-stage model (encounter-level), and pseudo R-squared value for the second-stage model (daily) were evaluated out-of-sample. R-squared represents the squared correlation coefficient between the counterfactual daily readmission risk and predicted values from the beta regression model.

Patient Risk Trajectory Cluster Analysis

A patient's daily 30-day readmission risk may change over the course of their hospitalization because of the evolution of their condition as therapies are administered. A total of 78.4% (107,299/136,911) of all laboratory tests, 74.0% (2173/2935) of all procedures, and 92.5% (3937/4258) of all medication orders occurred before the end of day 5 for our study cohort. Thus, we believe that the majority of variation in readmission risk occurred before this time. After training the beta regression model in the second stage, we used the trained model for predicting the daily readmission risk at admission—day 1 to day 5—and discharge. We applied unsupervised clustering using the K-means algorithm to determine readmission risk trajectories over these 7 specific milestones during a patients' stay [22]. We used this approach to learn potential trajectories (ie, time-dependent trends) in readmission risk that might naturally distinguish patient groups. Predictors differentiating these trajectory groups were then identified using Kruskal-Wallis hypothesis testing [23]. We further investigated how these discriminant predictors changed with time within each patient risk group.

Detailed steps of the entire modeling process are shown in [Figure 1](#). The detailed steps summarizing the entire modeling continuum train the predictive models, learn patient groups by trajectory, and identify predictors that differentiate these patient groups are summarized in [Multimedia Appendix 2](#).

Figure 1. Flowchart of our modeling process. The rectangular boxes represent data input and output. The yellow oval boxes represent the models. Independent and dependent variables for each model are colored in blue and red, respectively.



We conducted all of our analysis using Python (version 2.7) and R (version 3.4.1).

Results

Setting and Data

In our sample, we observed 534 patient encounters by 454 unique patients that included 164 patients who experienced multiple encounters. In total, 30-day readmission occurred in 107 (107/534, 20.0%) of these encounters. Characteristics of the patient cohort stratified by readmission outcome may be seen in [Table 1](#).

Supervised 2-Stage Prediction Model

Logistic regression with backward selection identified an optimal set of 57 predictors for our first-stage encounter-level prediction model. These optimal predictors are displayed in [Table 2](#) from a total pool listed in [Multimedia Appendix 3](#). Dynamic clinical predictors measuring laboratory results and vital signs, specifically potassium, sodium, BUN, hemoglobin, and diastolic blood pressure, demonstrated the most predictive value. Overall, the number of medication orders for digoxin, number of measurements of peripheral capillary oxygen saturation (SpO2), and number of hemodialysis performed were negatively associated with readmission risk. The number of

measurements of sodium and number of times mechanical ventilation was used were both positively associated with readmission risk. Discharge disposition was a significant administrative predictor. The patients who were discharged to a nursing facility compared with the non-nursing facility had a

much higher readmission risk (odds ratio of 10.59). Overall, the out-of-sample AUC of the first-stage encounter-level logistic regression model (Table 2) was 0.73 (SD 0.08). The R-squared value for the second-stage daily prediction beta regression model was 0.88.

Table 2. Summary of significant predictors for the logistic regression model with backward feature elimination.

Predictor	Coefficient	Odds ratio (95% CI)	P value
Potassium: normalized index of minimal potassium	-2.15	0.12 (-3.27 to -1.03)	<.001
SpO ₂ ^a : number of measurements of SpO ₂	-0.07	0.93 (-0.11 to -0.03)	.001
Hemoglobin			
First value-last value	.56	1.75 (0.20 to 0.92)	.002
Normalized index of maximal hemoglobin	1.63	5.14 (0.30 to 2.98)	.02
Digitalis glycosides: number of medication order	-0.67	0.51 (-1.11 to -0.23)	.003
Blood urea nitrogen: first value-last value	-0.04	0.96 (-0.07 to -0.01)	.004
Discharge disposition: nursing Facility	2.36	10.59 (0.62 to 4.10)	.008
Number of procedures			
Mechanical ventilation	0.08	1.08 (0.02 to 0.14)	.008
Hemodialysis	-0.41	0.67 (-0.80 to -0.01)	.04
Diastolic blood pressure			
First value-last value	-0.02	0.98 (-0.04 to -0.01)	.008
Number of measurements	0.05	1.05 (0.01 to 0.09)	.02
Average value of last 3 diastolic blood pressure	-0.05	0.96 (-0.09 to -0.004)	.03
Sodium: normalized index of minimal sodium	-1.21	0.30 (-2.25 to -0.18)	.02
Respiratory rate: normalized index of minimal respiratory rate	1.12	3.07 (0.10 to 2.15)	.03
ZIP code: 210XX	0.77	2.17 (0.05 to 1.50)	.04

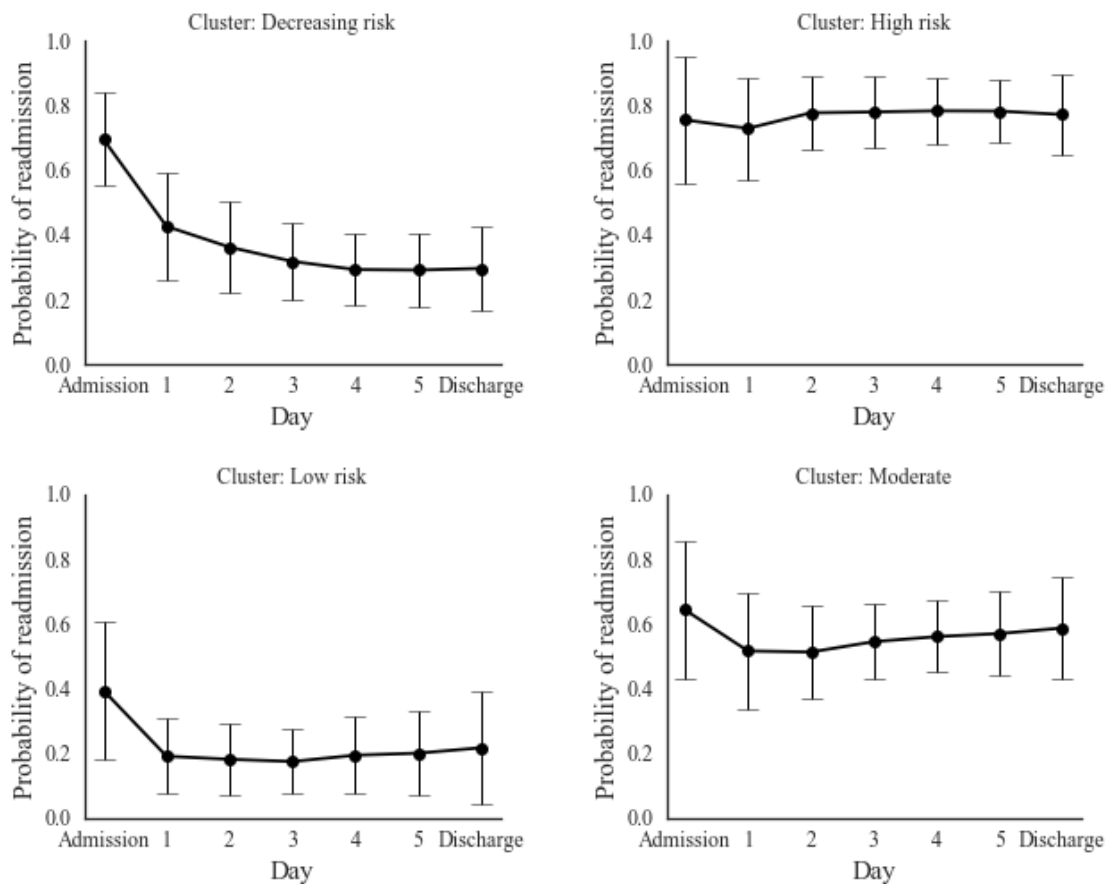
^aSpO₂: peripheral capillary oxygen saturation.

Unsupervised Clustering of Readmission Risk Trajectories

A total of 4 distinct trajectories emerged from the unsupervised clustering analysis as seen in Figure 2. The *decreasing risk* group with 131 encounters (131/534, 24.5%) was unique in demonstrating appreciable change in risk over the hospitalization from admission (.69 probability of readmission) to discharge (.30). The remaining trajectory groups maintained a consistent

readmission risk over time. This included the *high risk* group of 113 (113/534, 21.2%) encounters with average readmission risk maintained above 0.75 over their course of care. The *low risk* cluster with 113 (113/534, 21.2%) encounters were admitted with a relatively low average readmission risk of 0.39 that decreased to 0.21 at discharge. The *moderate risk* cluster had 177 (177/534, 33.1%) encounters, and its average readmission risk had a mean of 0.61.

Figure 2. Daily readmission probabilities for four clusters of patients are shown from admission to discharge. The error bars represent one standard deviation on each side of the expected readmission probabilities.

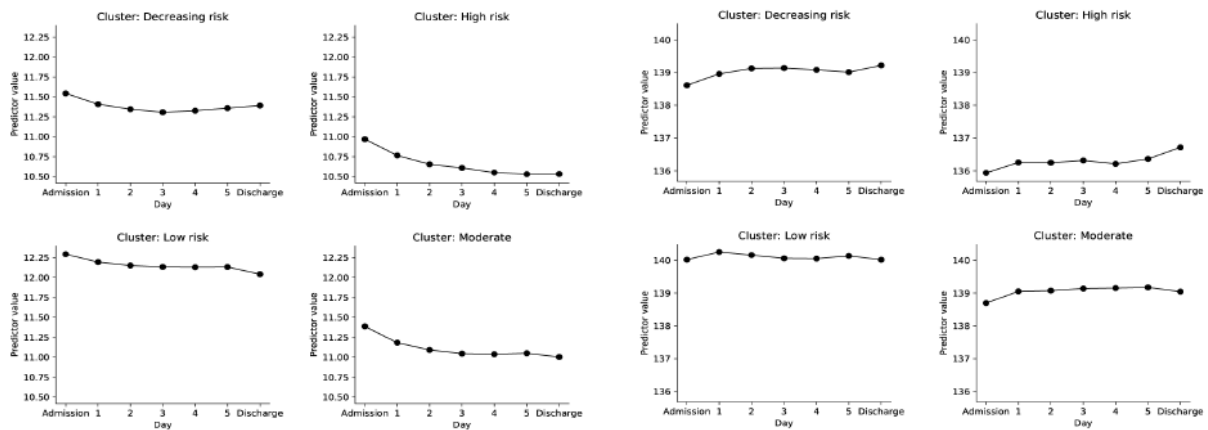


Discriminant analyses demonstrated 18 predictors with similar characteristics across the 4 risk trajectory groups (Kruskal Wallis $P < .001$) as seen in [Multimedia Appendix 4](#). The most discriminant predictors were laboratory measures (eg, potassium, hemoglobin, sodium, and ALT), diastolic blood pressure, and the number of past hospital visits. Large decreases in potassium and the minimum potassium level occurring later in the hospitalization was associated with decreasing readmission risk. Lower hemoglobin and sodium levels were associated with higher readmission risk. High variability, maximum, and average measures of diastolic blood pressure were associated with consistently low readmission risk, whereas low diastolic blood pressure measures were associated with patients whose readmission risk remained high during hospitalization. Finally,

the number of past hospital visits was associated with higher readmission risks.

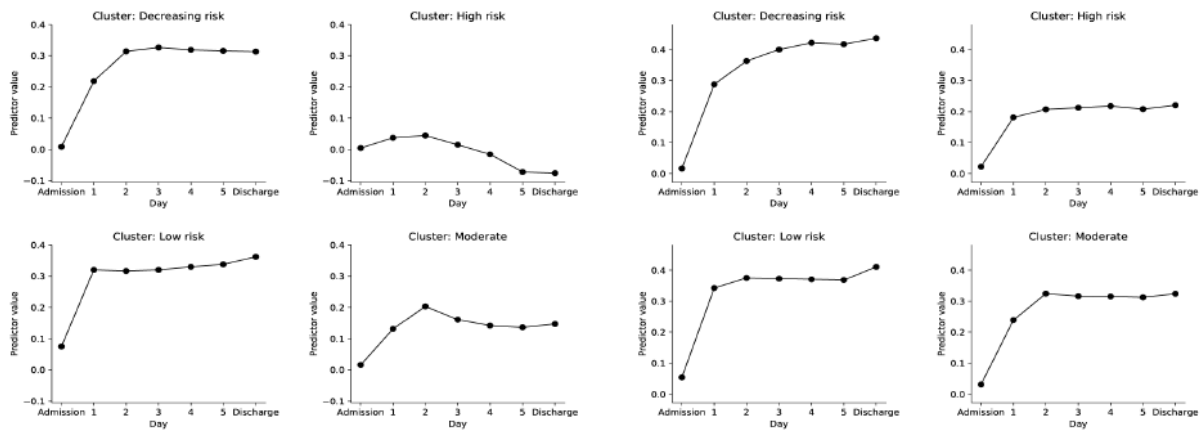
These discriminant predictors were further investigated by graphically displaying their temporal changes for each readmission trajectory group. [Figure 3](#) depicts the primary discriminant features (eg, hemoglobin, sodium, potassium, and diastolic blood pressure) for each group. Associative patterns included the following: low and decreasing levels of hemoglobin was associated with higher readmission risk, low levels of sodium was associated with higher readmission risk, large decreases in potassium and minimum potassium levels occurring close to discharge was associated with lower readmission risk, and a large decrease of diastolic blood pressure from admission was associated with lower readmission risk.

Figure 3. Change of discriminative predictors values over time from admission to discharge within each patient risk group. (a) Average value of last three hemoglobin measurements (gm/dL). (b) Average value of last three sodium measurements (mmol/L). (c) Decrease of potassium level from admission (mmol/L). (d) Normalized time of minimal potassium starting from admission. (e) Decrease of diastolic blood pressure level from admission (mmHg).



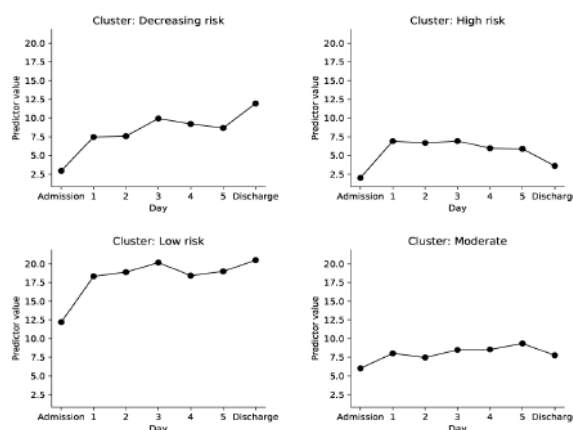
(a)

(b)



(c)

(d)



(e)

Discussion

Principal Findings

By examining the changes of readmission risk over the course of a patient's hospital stay, we discovered patient groups with distinct readmission risk trajectories. These groups may provide

additional insight about discriminant predictors that can be undetected using traditional static prediction models. A portion of our findings were consistent with previous research on readmission prediction for patients with heart failure:

- The predictive performance (AUC) for our dynamic modeling approach was 0.73 (SD 0.08), which was

relatively high but within the range (0.55-0.76 [3-8]) of performance reported in other studies that applied a static modeling approach for patients with heart failure.

- Number of previous hospitalizations and decreases in hemoglobin indicating anemia have similarly been associated with increased readmission risk [5,24-34].
- Nguyen et al found that vital sign instability at discharge was associated with increased risk-adjusted 30-day mortality and readmission rates [15]. Vital sign instability means abnormal values of vital signs. For instance, higher average hemoglobin measurements near discharge demonstrating stability was linked to lower readmission risk in our prediction model, similar to the findings from the study by Nguyen et al for general readmission [15].

However, there were some differences in our model findings not present in the literature to our knowledge. First, we found that dynamic clinical predictors measuring laboratory results and vital signs had the most value in predicting readmission risk and discriminating trajectories. This was contrary to the report by Hamill et al [6] of no incremental benefit of mixing clinical data with administrative claims data for readmission risk prediction. Specifically, higher levels of hemoglobin, a larger decrease in potassium, and diastolic blood pressure from admission to discharge indicate a lower readmission risk. Potassium changes during hospitalization of patients with heart failure are iatrogenic, that is, decreased by using intravenous diuretics and increased by repletion with oral or intravenous potassium supplementation. Therefore, it is not an intrinsic factor that affects heart failure risk. However, it can still serve as a surrogate indicator of heart failure readmission risk. A larger decrease in diastolic blood pressure probably reflects the condition of patients with diastolic heart failure who have increased diastolic blood pressure at admission [35]. Second, being different from past readmission prediction models that are static and only predict at discharge time [4-7], our model dynamically predicts readmission risk and facilitates informed interventions throughout patients' stay. Finally, using our dynamical prediction model, we identified 4 different readmission risk trajectory patterns over patients' stay. To our knowledge, our study is the first study that investigated the readmission risk trajectories over patient's stay for patients with heart failure.

Overall, these findings support assessments that may be useful in better predicting readmission risk over the course of a patient's stay. They include the following: (1) recognizing high readmission risks associated with patients who have been hospitalized frequently in the past, (2) abnormally low hemoglobin that is either consistent throughout the stay or drops near discharge is associated with higher risk of readmission, (3) abnormally low sodium is associated with high readmission risk, (4) abnormally high diastolic blood pressure late in the hospital stay or no considerable decreases (<4 mmHg) from admission to discharge may indicate higher readmission risk, (5) abnormally high potassium levels near discharge may indicate high risk of readmission, and (6) a minimal potassium occurring closer to discharge is associated with decreasing risk of readmission.

Limitations

There were several limitations in our study data and methods that should be considered when interpreting results. First, clustering of time-dependent trends in readmission used up to 7 time points (admission, 5 days, and discharge) for each encounter. This approach was advantageous in enabling comparisons of readmission risk profiles across patients but limited in its ability to compare short-stay patients (eg, <5 days) with longer-stay patients (>5 days). Ideally, we would perform clustering analysis controlling for the entire duration of each patient's stay. However, we believe that our 7-time point approach yielded results that would not qualitatively change as readmission risk trends vary less toward the end of the hospitalization, particularly for longer-stay patients. Second, owing to the limited sample size, we treated each encounter in our first-stage model as independent. When the same patient has multiple encounters, as seen in our data, this assumption may be contravened. This becomes particularly important to avoid bias in out-of-sample performance evaluation (ie, training and test sets). However, we determined that treating each patient encounter independently was analogous to how this predictive tool could be run in real time. We also further mitigated risks of bias by performing 5-fold cross-validation. Third, we used mean imputation to impute missing data in features. The disadvantage of mean imputation is that it reduces the variance of the imputed features and, as a result, the discriminability of that feature. A more comprehensive method to impute missing data may improve the prediction accuracy. Finally, evaluating predictive performance of our second-stage predictive model presented challenges methodologically. We were unable to observe the counterfactual readmission outcomes (readmission outcomes if the patient were discharged earlier than their actual discharge time). Using the last day's outcomes to estimate the predictive performance of the second-stage model is an incomplete estimate, as we ignore the predictive performance on all days before the day of discharge. Therefore, we used a pseudo R-squared value to estimate the predictive performance of the second-stage model.

Broad Impact

We applied supervised machine learning using a novel 2-stage approach to construct a 30-day readmission prediction model that yields daily predictions for patients with heart failure over time. We estimated the unknown counterfactual daily readmission risk using an encounter-level logistic regression model (first stage). A beta regression model using predictors engineered from EHR data accumulated to the prediction time (ie, each day) was fit to form the daily readmission risk prediction model (second stage). Investigating dynamic trends in readmission risk over time via unsupervised clustering, we uncovered for distinct groups based on readmission risk trajectory: decreasing, high, moderate, and low. The clinical features distinguishing these groups (eg, hemoglobin, sodium, potassium, and diastolic blood pressure) may shed light on new indicators of readmission for patients with heart failure and the isolated effects of the index hospitalization. Our ultimate goal is to use the important features and our model to develop a clinical tool for dynamically predicting readmission for patients with heart failure and target patients with high readmission risk.

The clinical tool is aimed to show readmission risk for each patient and also the corresponding factors that are associated with high or low readmission risk. This study provides a first step in the development of such a tool by proposing a dynamic prediction model for readmission risk of patients with heart failure.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Example of the daily data (patient-day data) used in the second stage model.

[\[PDF File \(Adobe PDF File\)42 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Algorithm 1. Steps for the entire modeling process.

[\[PDF File \(Adobe PDF File\)107 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Relevant categories of features (57 total) selected using logistic regression and backward feature elimination.

[\[PDF File \(Adobe PDF File\)23 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Summary of discriminative predictors for each patient group.

[\[PDF File \(Adobe PDF File\)192 KB-Multimedia Appendix 4\]](#)

References

1. Ambrosy AP, Fonarow GC, Butler J, Chioncel O, Greene SJ, Vaduganathan M, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. *J Am Coll Cardiol* 2014 Apr 1;63(12):1123-1133 [FREE Full text] [doi: [10.1016/j.jacc.2013.11.053](https://doi.org/10.1016/j.jacc.2013.11.053)] [Medline: [24491689](https://pubmed.ncbi.nlm.nih.gov/24491689/)]
2. Centers for Medicare and Medicaid Services. Hospital Readmissions Reduction Program (HRRP)URL:<https://www.cms.gov/medicare/medicare-fee-for-service-payment/acuteinpatientpps/readmissions-reduction-program.html> [accessed 2019-07-12]
3. Futoma J, Morris J, Lucas J. A comparison of models for predicting early hospital readmissions. *J Biomed Inform* 2015 Aug;56:229-238 [FREE Full text] [doi: [10.1016/j.jbi.2015.05.016](https://doi.org/10.1016/j.jbi.2015.05.016)] [Medline: [26044081](https://pubmed.ncbi.nlm.nih.gov/26044081/)]
4. Cholleti S, Post A, Gao J, Lin X, Bornstein W, Cantrell D, et al. Leveraging derived data elements in data analytic models for understanding and predicting hospital readmissions. *AMIA Annu Symp Proc* 2012;2012:103-111 [FREE Full text] [Medline: [23304278](https://pubmed.ncbi.nlm.nih.gov/23304278/)]
5. Lemke KW, Weiner JP, Clark JM. Development and validation of a model for predicting inpatient hospitalization. *Med Care* 2012 Feb;50(2):131-139. [doi: [10.1097/MLR.0b013e3182353ceb](https://doi.org/10.1097/MLR.0b013e3182353ceb)] [Medline: [22002640](https://pubmed.ncbi.nlm.nih.gov/22002640/)]
6. Hammill BG, Curtis LH, Fonarow GC, Heidenreich PA, Yancy CW, Peterson ED, et al. Incremental value of clinical data beyond claims data in predicting 30-day outcomes after heart failure hospitalization. *Circ Cardiovasc Qual Outcomes* 2011 Jan 1;4(1):60-67. [doi: [10.1161/CIRCOUTCOMES.110.954693](https://doi.org/10.1161/CIRCOUTCOMES.110.954693)] [Medline: [21139093](https://pubmed.ncbi.nlm.nih.gov/21139093/)]
7. Meadem N, Verbiest N, Zolfaghar K, Agarwal J, Chin SC, Roy SB. CiteSeerX. Exploring Preprocessing Techniques for Prediction of Risk of Readmission for Congestive Heart Failure PatientsURL:<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.677.1282> [accessed 2019-07-12]
8. Tang YD, Katz SD. Anemia in chronic heart failure: prevalence, etiology, clinical correlates, and treatment options. *Circulation* 2006 May 23;113(20):2454-2461. [doi: [10.1161/CIRCULATIONAHA.105.583666](https://doi.org/10.1161/CIRCULATIONAHA.105.583666)] [Medline: [16717164](https://pubmed.ncbi.nlm.nih.gov/16717164/)]
9. Centers for Disease Control and Prevention. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)URL:<https://www.cdc.gov/nchs/icd/icd9cm.htm> [accessed 2019-07-12]
10. Saczynski JS, Andrade SE, Harrold LR, Tjia J, Cutrona SL, Dodd KS, et al. A systematic review of validated methods for identifying heart failure using administrative data. *Pharmacoepidemiol Drug Saf* 2012 Jan;21(Suppl 1):129-140 [FREE Full text] [doi: [10.1002/pds.2313](https://doi.org/10.1002/pds.2313)] [Medline: [22262599](https://pubmed.ncbi.nlm.nih.gov/22262599/)]
11. Huynh QL, Saito M, Blizzard CL, Eskandari M, Johnson B, Adabi G, MARATHON Investigators. Roles of nonclinical and clinical data in prediction of 30-day rehospitalization or death among heart failure patients. *J Card Fail* 2015 May;21(5):374-381. [doi: [10.1016/j.cardfail.2015.02.002](https://doi.org/10.1016/j.cardfail.2015.02.002)] [Medline: [25724302](https://pubmed.ncbi.nlm.nih.gov/25724302/)]
12. Flint KM, Allen LA, Pham M, Heidenreich PA. B-type natriuretic peptide predicts 30-day readmission for heart failure but not readmission for other causes. *J Am Heart Assoc* 2014 Jun 10;3(3):e000806 [FREE Full text] [doi: [10.1161/JAHA.114.000806](https://doi.org/10.1161/JAHA.114.000806)] [Medline: [24922626](https://pubmed.ncbi.nlm.nih.gov/24922626/)]

13. Januzzi JL, van Kimmenade R, Lainchbury J, Bayes-Genis A, Ordonez-Llanos J, Santalo-Bel M, et al. NT-proBNP testing for diagnosis and short-term prognosis in acute destabilized heart failure: an international pooled analysis of 1256 patients: the international collaborative of NT-proBNP study. *Eur Heart J* 2006 Feb;27(3):330-337. [doi: [10.1093/eurheartj/ehi631](https://doi.org/10.1093/eurheartj/ehi631)] [Medline: [16293638](https://pubmed.ncbi.nlm.nih.gov/16293638/)]
14. Amarasingham R, Moore BJ, Tabak YP, Drazner MH, Clark CA, Zhang S, 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 Nov;48(11):981-988. [doi: [10.1097/MLR.0b013e3181ef60d9](https://doi.org/10.1097/MLR.0b013e3181ef60d9)] [Medline: [20940649](https://pubmed.ncbi.nlm.nih.gov/20940649/)]
15. Nguyen OK, Makam AN, Clark C, Zhang S, Xie B, Velasco F, et al. Vital signs are still vital: instability on discharge and the risk of post-discharge adverse outcomes. *J Gen Intern Med* 2017 Jan;32(1):42-48 [FREE Full text] [doi: [10.1007/s11606-016-3826-8](https://doi.org/10.1007/s11606-016-3826-8)] [Medline: [27503438](https://pubmed.ncbi.nlm.nih.gov/27503438/)]
16. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, et al. Risk prediction models for hospital readmission: a systematic review. *J Am Med Assoc* 2011 Oct 19;306(15):1688-1698 [FREE Full text] [doi: [10.1001/jama.2011.1515](https://doi.org/10.1001/jama.2011.1515)] [Medline: [22009101](https://pubmed.ncbi.nlm.nih.gov/22009101/)]
17. Dugas AF, Kirsch TD, Toerper M, Korley F, Yenokyan G, France D, et al. An electronic emergency triage system to improve patient distribution by critical outcomes. *J Emerg Med* 2016 Jun;50(6):910-918. [doi: [10.1016/j.jemermed.2016.02.026](https://doi.org/10.1016/j.jemermed.2016.02.026)] [Medline: [27133736](https://pubmed.ncbi.nlm.nih.gov/27133736/)]
18. Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, et al. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. *Ann Emerg Med* 2018 May;71(5):565-74.e2. [doi: [10.1016/j.annemergmed.2017.08.005](https://doi.org/10.1016/j.annemergmed.2017.08.005)] [Medline: [28888332](https://pubmed.ncbi.nlm.nih.gov/28888332/)]
19. Wiens J, Gutttag JV, Horvitz E. Patient Risk Stratification for Hospital-Associated C. Diff as a Time-Series Classification Task. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems*. 2012 Presented at: NIPS'12; December 3-6, 2012; Lake Tahoe, Nevada p. 467-475.
20. Levin SR, Harley ET, Fackler JC, Lehmann CU, Custer JW, France D, et al. Real-time forecasting of pediatric intensive care unit length of stay using computerized provider orders. *Crit Care Med* 2012 Nov;40(11):3058-3064. [doi: [10.1097/CCM.0b013e31825bc399](https://doi.org/10.1097/CCM.0b013e31825bc399)] [Medline: [22824935](https://pubmed.ncbi.nlm.nih.gov/22824935/)]
21. Cribari-Neto F, Zeileis A. Beta regression in R. *J Stat Soft* 2010;34(2):1-24. [doi: [10.18637/jss.v034.i02](https://doi.org/10.18637/jss.v034.i02)]
22. Friedman J, Hastie T, Tibshirani R. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Volume 1. Berlin: Springer; 2001.
23. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc* 1952 Dec;47(260):583-621. [doi: [10.1080/01621459.1952.10483441](https://doi.org/10.1080/01621459.1952.10483441)]
24. Krumholz HM, Chen YT, Wang Y, Vaccarino V, Radford MJ, Horwitz RI. Predictors of readmission among elderly survivors of admission with heart failure. *Am Heart J* 2000 Jan;139(1 Pt 1):72-77. [doi: [10.1016/s0002-8703\(00\)90311-9](https://doi.org/10.1016/s0002-8703(00)90311-9)] [Medline: [10618565](https://pubmed.ncbi.nlm.nih.gov/10618565/)]
25. Donzé J, Aujesky D, Williams D, Schnipper JL. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. *JAMA Intern Med* 2013 Apr 22;173(8):632-638. [doi: [10.1001/jamainternmed.2013.3023](https://doi.org/10.1001/jamainternmed.2013.3023)] [Medline: [23529115](https://pubmed.ncbi.nlm.nih.gov/23529115/)]
26. Hasan O, Meltzer DO, Shaykevich SA, Bell CM, Kaboli PJ, Auerbach AD, et al. Hospital readmission in general medicine patients: a prediction model. *J Gen Intern Med* 2010 Mar;25(3):211-219 [FREE Full text] [doi: [10.1007/s11606-009-1196-1](https://doi.org/10.1007/s11606-009-1196-1)] [Medline: [20013068](https://pubmed.ncbi.nlm.nih.gov/20013068/)]
27. 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. *Br Med J* 2006 Aug 12;333(7563):327 [FREE Full text] [doi: [10.1136/bmj.38870.657917.AE](https://doi.org/10.1136/bmj.38870.657917.AE)] [Medline: [16815882](https://pubmed.ncbi.nlm.nih.gov/16815882/)]
28. Al-Ahmad A, Rand WM, Manjunath G, Konstam MA, Salem DN, Levey AS, et al. Reduced kidney function and anemia as risk factors for mortality in patients with left ventricular dysfunction. *J Am Coll Cardiol* 2001 Oct;38(4):955-962 [FREE Full text] [doi: [10.1016/s0735-1097\(01\)01470-x](https://doi.org/10.1016/s0735-1097(01)01470-x)] [Medline: [11583864](https://pubmed.ncbi.nlm.nih.gov/11583864/)]
29. Anand IS, Kuskowski MA, Rector TS, Florea VG, Glazer RD, Hester A, et al. Anemia and change in hemoglobin over time related to mortality and morbidity in patients with chronic heart failure: results from Val-HeFT. *Circulation* 2005 Aug 23;112(8):1121-1127. [doi: [10.1161/CIRCULATIONAHA.104.512988](https://doi.org/10.1161/CIRCULATIONAHA.104.512988)] [Medline: [16103233](https://pubmed.ncbi.nlm.nih.gov/16103233/)]
30. Androne AS, Katz SD, Lund L, LaManca J, Hudaihed A, Hryniewicz K, et al. Hemodilution is common in patients with advanced heart failure. *Circulation* 2003 Jan 21;107(2):226-229. [doi: [10.1161/01.cir.0000052623.16194.80](https://doi.org/10.1161/01.cir.0000052623.16194.80)] [Medline: [12538419](https://pubmed.ncbi.nlm.nih.gov/12538419/)]
31. Brucks S, Little WC, Chao T, Rideman RL, Upadhyya B, Wesley-Farrington D, et al. Relation of anemia to diastolic heart failure and the effect on outcome. *Am J Cardiol* 2004 Apr 15;93(8):1055-1057. [doi: [10.1016/j.amjcard.2003.12.062](https://doi.org/10.1016/j.amjcard.2003.12.062)] [Medline: [15081458](https://pubmed.ncbi.nlm.nih.gov/15081458/)]
32. Ezekowitz JA, McAlister FA, Armstrong PW. Anemia is common in heart failure and is associated with poor outcomes: insights from a cohort of 12 065 patients with new-onset heart failure. *Circulation* 2003 Jan 21;107(2):223-225. [doi: [10.1161/01.cir.0000052622.51963.fc](https://doi.org/10.1161/01.cir.0000052622.51963.fc)] [Medline: [12538418](https://pubmed.ncbi.nlm.nih.gov/12538418/)]

33. Felker GM, Gattis WA, Leimberger JD, Adams KF, Cuffe MS, Gheorghide M, et al. Usefulness of anemia as a predictor of death and rehospitalization in patients with decompensated heart failure. *Am J Cardiol* 2003 Sep 1;92(5):625-628. [doi: [10.1016/s0002-9149\(03\)00740-9](https://doi.org/10.1016/s0002-9149(03)00740-9)] [Medline: [12943893](https://pubmed.ncbi.nlm.nih.gov/12943893/)]
34. Sharma R, Francis DP, Pitt B, Poole-Wilson PA, Coats AJ, Anker SD. Haemoglobin predicts survival in patients with chronic heart failure: a substudy of the ELITE II trial. *Eur Heart J* 2004 Jun;25(12):1021-1028. [doi: [10.1016/j.ehj.2004.04.023](https://doi.org/10.1016/j.ehj.2004.04.023)] [Medline: [15191772](https://pubmed.ncbi.nlm.nih.gov/15191772/)]
35. Gutierrez C, Blanchard DG. Diastolic heart failure: challenges of diagnosis and treatment. *Am Fam Physician* 2004 Jun 1;69(11):2609-2616 [[FREE Full text](#)] [Medline: [15202695](https://pubmed.ncbi.nlm.nih.gov/15202695/)]

Abbreviations

ALT: alanine transaminase
AST: aspartate transaminase
AUC: area under curve
BUN: blood urea nitrogen
EHR: electronic health record
proBNP: pro B-type natriuretic peptide

Edited by G Eysenbach; submitted 18.05.19; peer-reviewed by E van der Velde, M Nomali; comments to author 26.06.19; revised version received 14.07.19; accepted 19.07.19; published 16.09.19

Please cite as:

Jiang W, Siddiqui S, Barnes S, Barouch LA, Korley F, Martinez DA, Toerper M, Cabral S, Hamrock E, Levin S
Readmission Risk Trajectories for Patients With Heart Failure Using a Dynamic Prediction Approach: Retrospective Study
JMIR Med Inform 2019;7(4):e14756

URL: <https://medinform.jmir.org/2019/4/e14756>

doi: [10.2196/14756](https://doi.org/10.2196/14756)

PMID:

©Wei Jiang, Sauleh Siddiqui, Sean Barnes, Lili A Barouch, Frederick Korley, Diego A Martinez, Matthew Toerper, Stephanie Cabral, Eric Hamrock, Scott Levin. Originally published in *JMIR Medical Informatics* (<http://medinform.jmir.org>), 05.10.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Medical Informatics*, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.