

Using Machine Learning to Predict Likelihood and Cause of Readmission After Hospitalization for Chronic Obstructive Pulmonary Disease Exacerbation

Matthew Bonomo^{1,*}, Michael G Hermsen^{1,*}, Samuel Kaskovich¹, Maximilian J Hemmrich¹, Juan C Rojas², Kyle A Carey³, Laura Ruth Venable⁴, Matthew M Churpek⁵, Valerie G Press^{3,6}

¹Pritzker School of Medicine, University of Chicago, Chicago, IL, USA; ²Department of Medicine, Section of Pulmonary/Critical Care, University of Chicago, Chicago, IL, USA; ³Department of Medicine, Section of General Internal Medicine, University of Chicago, Chicago, IL, USA; ⁴Department of Medicine, Section of Hospitalist Medicine, University of Chicago, Chicago, IL, USA; ⁵Department of Medicine, Division of Allergy, Pulmonary and Critical Care Medicine, University of Wisconsin-Madison, Madison, WI, USA; ⁶Department of Pediatrics, Section of Academic Pediatrics, University of Chicago, Chicago, IL, USA

*These authors contributed equally to this work

Correspondence: Valerie G Press, University of Chicago, 5841 S Maryland, MC 2007, Chicago, IL, 60637, USA, Tel +773-702-5170, Email vpress@medicine.bsd.uchicago.edu

Background: Chronic obstructive pulmonary disease (COPD) is a leading cause of hospital readmissions. Few existing tools use electronic health record (EHR) data to forecast patients' readmission risk during index hospitalizations.

Objective: We used machine learning and in-hospital data to model 90-day risk for and cause of readmission among inpatients with acute exacerbations of COPD (AE-COPD).

Design: Retrospective cohort study.

Participants: Adult patients admitted for AE-COPD at the University of Chicago Medicine between November 7, 2008 and December 31, 2018 meeting International Classification of Diseases (ICD)-9 or -10 criteria consistent with AE-COPD were included.

Methods: Random forest models were fit to predict readmission risk and respiratory-related readmission cause. Predictor variables included demographics, comorbidities, and EHR data from patients' index hospital stays. Models were derived on 70% of observations and validated on a 30% holdout set. Performance of the readmission risk model was compared to that of the HOSPITAL score.

Results: Among 3238 patients admitted for AE-COPD, 1103 patients were readmitted within 90 days. Of the readmission causes, 61% (n = 672) were respiratory-related and COPD (n = 452) was the most common. Our readmission risk model had a significantly higher area under the receiver operating characteristic curve (AUROC) (0.69 [0.66, 0.73]) compared to the HOSPITAL score (0.63 [0.59, 0.67]; $p = 0.002$). The respiratory-related readmission cause model had an AUROC of 0.73 [0.68, 0.79].

Conclusion: Our models improve on current tools by predicting 90-day readmission risk and cause at the time of discharge from index admissions for AE-COPD. These models could be used to identify patients at higher risk of readmission and direct tailored post-discharge transition of care interventions that lower readmission risk.

Keywords: chronic obstructive lung disease, COPD, readmissions, machine learning

Plain Language Summary

Chronic obstructive pulmonary disease (COPD) is a leading cause of hospital readmissions. Few resources exist that can predict a patients' risk of hospital readmission during an initial admission. We developed a model to assess risk of hospital readmission within 90 days after an initial admission for acute exacerbations of COPD. We also developed a model to predict reason for future readmission. To develop these models, we utilized previously collected medical record data for adult patients who were admitted for an acute exacerbation of COPD at the University of Chicago Medicine between 2008–2018. Variables that were included in these

models included patient demographics, co-existing conditions, and laboratory values, vitals, and other health record data from the initial hospital stay. We compared our models to another commonly-used prediction model called the HOSPITAL score. We measured accuracy of the models using a measure called area under the receiver operating characteristic curve (AUROC). Our model for predicting risk of readmission within 90-days had a higher AUROC (0.69 [0.66, 0.73]) compared to the HOSPITAL score (0.63 [0.59, 0.67]). Our model for predicting reason for readmission showed that 61% were respiratory-related causes and COPD was the most common. Our results showed that these new models improve upon existing tools for predicting 90-day readmission risk and cause of readmission for patients admitted for acute exacerbations of COPD. These models could be used to identify patients at higher risk of readmission and direct care towards the patients that need it most.

Introduction

In the United States, chronic obstructive pulmonary disease (COPD) affects more than 16 million adults and trails just behind heart disease and cancer as a leading cause of death.¹ Acute exacerbations of COPD (AE-COPD) are the third-leading cause of 30-day hospital readmissions and account for up to 70% of COPD-related healthcare costs.² More than one-third of patients admitted for AE-COPD are readmitted within 90 days post-hospitalization, leading to a cumulative expenditure of over \$15 billion annually.^{3,4} Given these high costs, the Hospital Readmissions Reduction Program (HRRP) began levying penalties against hospitals for early AE-COPD readmissions in 2014.^{5,6} As a result, hospitals across the country were galvanized to develop interventions in an attempt to reduce readmissions, though the effectiveness of these interventions has been challenged in part by frequent diagnostic ambiguity and International Classification of Diseases (ICD) code misclassification.⁶

Predicting the risk for and cause of readmission is difficult. Clinician judgment has been shown to be largely inaccurate, predicting readmissions correctly less than half of the time.⁷ Further, the causes of readmission for these patients are heterogeneous.⁶ AE-COPD is cited as the readmission diagnosis after an initial COPD-related hospitalization in just one-third of cases.⁶ While general risk factors for predicting readmission have been well described, they have not been useful in improving risk prediction tools for COPD patients.⁸ Accurate tools for identifying patients at high risk for readmission during their index stay and predicting their most likely cause of readmission could assist care teams with deciding which patients would benefit most from in-hospital and post-discharge interventions aimed at reducing healthcare utilization after hospitalization for AE-COPD.^{9,10}

To date, few such tools exist. Two of the most commonly used readmission risk prediction models include the HOSPITAL score for 30-day general readmission risk and the PEARL score for 90-day COPD readmission risk. Notably, both utilize predictors that are not readily available in the electronic health record (EHR). For example, the extended Medical Research Council Dyspnea (eMRCD) score used in the PEARL score and admission type (elective vs urgent or emergent) used in the HOSPITAL score make both automation and real-time use during an index admission less feasible.^{11,12} Additionally, these tools were derived using logistic regression when prior work has shown that more sophisticated machine learning algorithms can significantly increase prediction accuracy.^{13–15}

We aimed to develop machine learning models using inpatient EHR data to predict, at discharge, both the 90-day risk for and cause of readmission for patients hospitalized for AE-COPD. These two models could be used in concert at the time of discharge to alert clinicians to patients at high risk for re-hospitalization and tailor patients' transitional care interventions based on their most likely cause of readmission. These tools have the potential to decrease readmissions, improve patient quality of life, and reduce healthcare costs.

Materials and Methods

Study Population

We conducted a retrospective cohort study of patients admitted for AE-COPD at The University of Chicago Medicine (UCM). The University of Chicago Institutional Review Board approved this study and granted an exemption (IRB #17-0332). We were approved for a waiver of consent as this was a review of existing medical record data collected for clinical care purposes that did not present any more than minimal risk. Further, given the large number of records reviewed, it would have been impracticable to obtain informed consent. All adult patients admitted to UCM between November 7, 2008, and December 31, 2018, were eligible for inclusion in our study. The following inclusion criteria were used to select patients: age

≥ 40 years at admission, primary or secondary ICD-9 or ICD-10 diagnosis code indicating AE-COPD, administration of both nebulized medications and systemic steroids during admission, and discharge ≥ 90 days before the study end date. The ICD-9 and ICD-10 patient selection scheme was chosen for its high specificity (99.5%) in identifying AE-COPD as compared to gold-standard chart review.¹⁶ We followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline.¹⁷

Outcome

The outcome for the readmission risk model was all-cause readmission to UCM within 90 days of discharge from the index admission. The outcome for the readmission cause model was respiratory-related diagnosis as the cause of readmission. A readmission was considered respiratory-related if the readmission was associated with one or more specific diagnosis codes. Specifically, diagnosis codes were considered respiratory-related if they are included in Clinical Classification Software (CCS) categories 122–134 (for ICD-9 codes) or if they start with J (for ICD-10 codes), similar to prior work.⁶ In addition, there were some respiratory-related diagnosis codes that would otherwise not have been included (eg, pulmonary tuberculosis) by CCS categories 122–134 or an ICD-10 code beginning with J. For completeness, these specific diagnoses were identified by manual review of ICD-9 code descriptions and included. ICD-10 conversions of these additional codes, based on general equivalence mappings from the Centers for Medicare and Medicaid Services, were also included.

Predictor Selection

Predictors such as demographics, vital signs, laboratory results, medications administered, comorbidities from prior admissions as determined by the Elixhauser classification system, and other clinical data were retrieved from a secure server hosted by the University of Chicago's Clinical Research Data Warehouse. Predictors deemed of little significance based on prior literature and our clinical experience were removed.^{6,12,18–20} For vital signs and laboratory values, the last value before discharge was used. Elixhauser Index was calculated using weightings provided by the Agency for Healthcare Research and Quality (AHRQ).²¹ As performed in other similar studies, missing numeric variables and nonsense outlier values were imputed using medians calculated from our study cohort or were set to 0 in the case of binary variables.^{14,22–24}

Model Derivation and Validation

Patient data were split randomly into a derivation set (70% of encounters) and a validation set (30% of encounters). Given that decision tree algorithms are susceptible to bias from class imbalance, the derivation data were up-sampled such that there were an equal number of readmitted and non-readmitted events and an equal number of respiratory and non-respiratory readmission causes.²⁵ Random forest was selected as the prediction model of choice for this study. Random forest machine learning models generate an ensemble of hundreds of individual decision trees, whose cumulative output predicts an outcome based on averages or majority voting.²⁶ By utilizing a large number of decision trees, random forests are able to learn important variable interaction, non-linearities, and have been shown to outperform other methods in various tasks.²⁷ Ten-fold cross-validation was performed during model derivation to optimize the hyperparameter *mtry* (number of variables randomly selected to be used at every split point) for the maximal area under the receiver operating characteristic curve (AUROC). The default value of 500 trees for the hyperparameter *ntree* (number of trees to grow) was used. Two random forest models were fit in the derivation data: one for all-cause readmission risk and one for whether the cause of readmission was respiratory-related for those who were readmitted. Variable importance for the fitted models was calculated using permutation importance.²⁸

HOSPITAL Score

The predictors in the HOSPITAL score are: hemoglobin at discharge, discharge from an oncology service, sodium level at discharge, any ICD-9 coded procedure performed during admission, index admission type (elective vs urgent or emergent), number of admissions during the previous year, and length of stay.⁸ Of these, hemoglobin and sodium levels at discharge were readily available in the EHR. Patients were considered discharged from an oncology service if their

room number was located in the rooms of the hospital utilized by our oncology service. Because ICD procedure codes were not available, we considered patients to have had a procedure performed during admission if they ever went to an operating room. Since this proxy likely underestimates the proportion of patients having a procedure (eg arterial line placement would not be captured under this method), we ran a sensitivity analysis comparing this adapted HOSPITAL score to one which assumes patients had a procedure if they ever went to an operating room or were ever in the ICU. Index admissions were considered urgent or emergent if the patient was admitted from the Emergency Department. Number of admissions in the previous year was calculated using inpatient data from our Clinical Research Data Warehouse. Length of stay was calculated as the time between first and last vitals for each admission.

Statistical Analysis

Predicted probabilities from the models fit in the derivation datasets were calculated in the 30% validation sets. Model performance was assessed using the area under the receiver operating characteristic curve (AUROC) in the validation sets. Briefly, AUROC values represent a model's ability to discriminate between the two outcome classes; a value of 0.5 indicates no discriminative ability and a value of 1.0 indicates perfect discriminative ability. Comparison of the readmission risk model to the HOSPITAL score was performed using a paired Delong's test.²⁹ Calibration curves, which compare the predicted probabilities from the model to the observed outcomes, were generated for the readmission risk and cause models ([Supplementary Figures 1–3](#)). For the HOSPITAL score, calibration was assessed by plotting the percent of readmitted patients at each value of the HOSPITAL score. Model development was performed using R version 4.1.0 (The R Foundation for Statistical Computing; Vienna, Austria). All other data analysis was performed using Stata version 15.1 (StataCorps; College Station, Texas).

Results

Study Population

Of the 3238 patients who met the inclusion criteria for having an index hospital admission for AE-COPD, the majority were female (62%) and Black (90%). The mean age was 64 years, and the median length of stay was 3.47 days. Nearly one-third of patients (32%) met Elixhauser criteria for congestive heart failure. Of the patients with an index admission for AE-COPD, 1103 (34%) had an unplanned 90-day readmission. Of the 1103 readmitted patients, only 1096 had an associated readmission diagnosis in our administrative dataset, and thus only those cases with an associated diagnosis were included in the readmission cause model. Compared to the group of patients that were not readmitted, the readmitted cohort had a slightly higher proportion of Black patients (93% vs 89%; $p < 0.05$) and was slightly younger (64 years vs 65 years; $p < 0.05$).

The readmitted cohort also had higher median Elixhauser Index (9 vs 3; $p < 0.05$) and higher median HOSPITAL score (3 vs 2; $p < 0.05$). Approximately 61% (672 patients) of the 1096 90-day readmissions had respiratory-related readmission diagnoses ([Table 1](#)). The most common respiratory-related readmission diagnosis was COPD (452 patients; [Supplementary Figures 4](#) and [5](#)). Among readmitted patients, those with a respiratory cause of readmission had a lower median Elixhauser Index (8 vs 12; $p < 0.05$), and shorter median length of stay (3.5 vs 3.8 days; $p > 0.05$; see [Table 1](#)).

Readmission Risk Model and HOSPITAL

For the readmission risk model, the 70% derivation dataset included 2266 admissions prior to upsampling, and the 30% validation dataset included 972 admissions. The final, optimized model, with 500 trees and an *mtry* value of 12, had an AUROC of 0.69 (95% CI [0.66, 0.73]) in the validation set ([Figure 1A](#)). The HOSPITAL score had a significantly lower AUROC of 0.63 (95% CI [0.59, 0.67]) in the validation set ($p = 0.002$; [Figure 1A](#)). Our model was 22% specific at a sensitivity of 90% and 70% specific at a sensitivity of 60%. The HOSPITAL score was 14% specific at a sensitivity of 90% and 60% specific at a sensitivity of 60%. The most important variables in our readmission risk model included length of stay, age, hemoglobin at discharge, systolic blood pressure at discharge, and platelet count at discharge ([Figure 2](#)). The readmission risk model had a calibration intercept of -0.19 and a calibration slope of 1.16 with an unreliability index of 0.01 ($p < 0.001$; [Supplementary Figure 1](#)). For the HOSPITAL score, higher rates of readmission

Table I Patient Demographics by Readmission Status

Values Reported as n(%), Mean (\pm SD), or Median (IQR)	Total (n=3238)	Not Readmitted (n=2142)	Readmitted (n=1103)	Readmitted	
				Respiratory (n = 672)	Non-Respiratory (n=424)
Female	2007 (62%)	1306 (61%)	701 (64%)	427 (64%)	268 (63%)
Black	2917 (90%)	1891 (89%)	1026 (93%) [†]	635 (94%)	386 (91%) [‡]
Hispanic	37 (1%)	28 (1%)	9 (1%)	6 (1%)	3 (1%)
Age, years	64 (\pm 11)	65 (\pm 11)	64 (\pm 11)	63 (\pm 10)	65 (\pm 11) [‡]
BMI, kg/m ²	26.7 (21.9, 33.3)	26.7 (21.9, 33.4)	26.7 (21.9, 33.0)	25.9 (21.3, 31.6)	27.7 (22.7, 34.7) [‡]
Elixhauser Criteria for Renal Failure*	524 (16%)	287 (13%)	237 (21%) [†]	108 (16%)	129 (30%) [‡]
Elixhauser Criteria for CHF*	1034 (32%)	576 (27%)	458 (42%) [†]	235 (35%)	220 (52%) [‡]
AHRQ Elixhauser Index	5 (0, 17)	3 (0, 14)	9 (0, 22) [†]	8 (0, 19.3)	12 (2, 24) [‡]
HOSPITAL Score	2 (1, 4)	2 (1, 4)	3 (2, 4) [†]	3 (2, 5)	3 (2, 4)
Length of Stay, days	3.5 (2.1, 5.6)	3.3 (2.1, 5.6)	3.7 (2.2, 5.7)	3.5 (2.2, 5.1)	3.8 (2.2, 6.1)

Notes: *Criteria fulfilled from previously documented admission prior to index admission; BMI = body mass index. [†] $p < 0.05$ when compared to not readmitted cohort; [‡] $p < 0.05$ when compared to respiratory cohort.

Abbreviations: CHF, congestive heart failure; AHRQ, Agency for Healthcare Research and Quality.

were generally seen with higher HOSPITAL scores ([Supplementary Figure 2](#)). A sensitivity analysis revealed no improvement in AUC when the HOSPITAL score was calculated using location in an operating room or ICU as a proxy for having a procedure (AUC 0.62 [0.58, 0.66]; $p = 0.01$).

Readmission Cause Model

For the readmission cause model, the 70% derivation dataset included 767 readmissions prior to upsampling, and the 30% validation dataset included 329 readmissions. The final, optimized model, with 500 trees and an *mtry* of 13 had an AUROC of 0.73 (95% CI [0.68, 0.79] in the validation set ([Figure 1B](#)). The readmission cause model was 31% specific at a sensitivity of 90% and 71% specific at a sensitivity of 60%. The most important variables in readmission cause model included heart rate, blood urea nitrogen, creatinine, and body mass index ([Figure 3](#)). The readmission cause model had a calibration intercept of -0.003 and a calibration slope of 1.48 with an unreliability index of 0.01 ($p = 0.04$; [Supplementary Figure 3](#)).

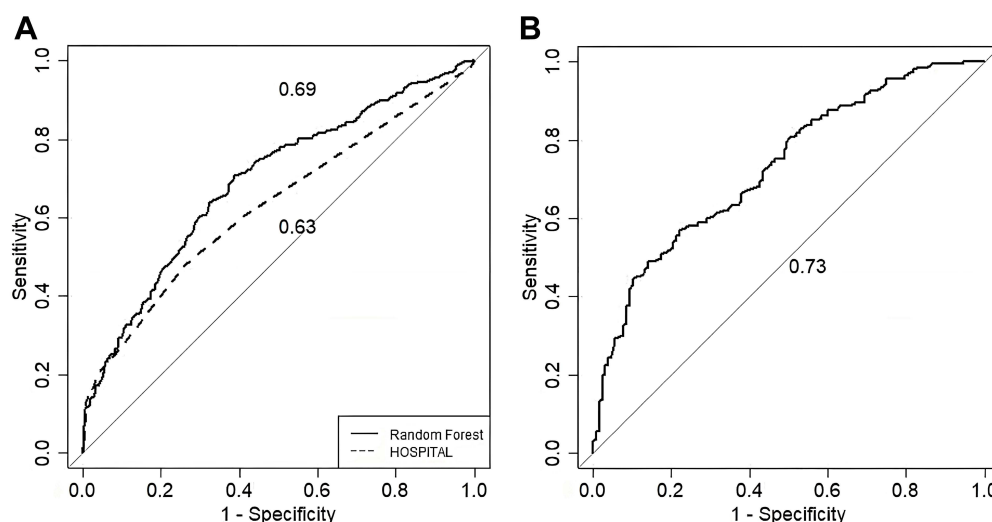


Figure 1 (A) ROC curves of 90-day readmission risk random forest model and HOSPITAL score. (B) ROC curve of 90-day readmission cause model.

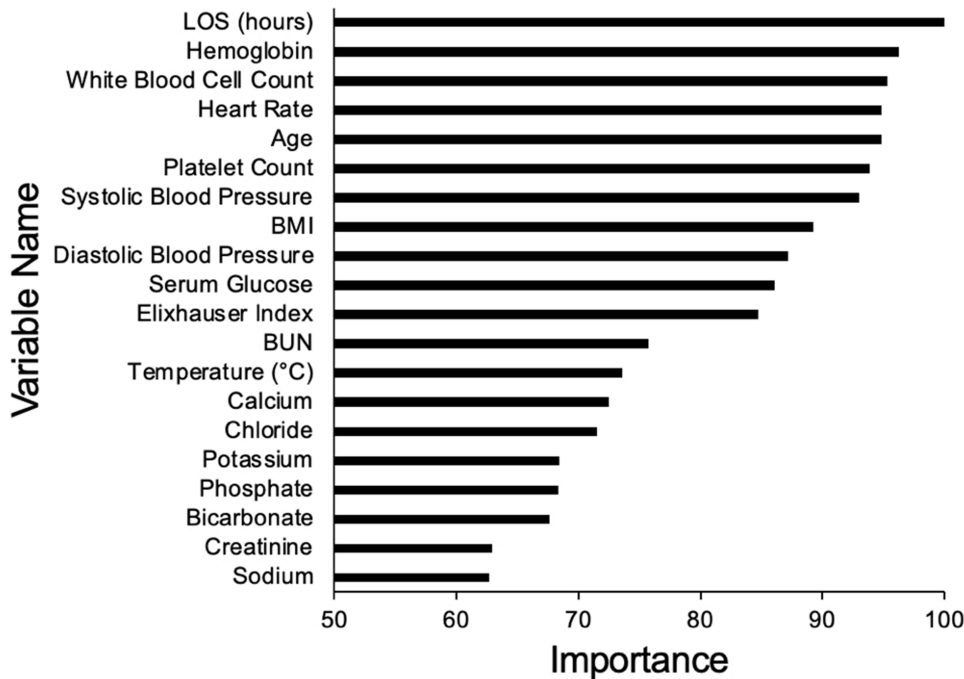


Figure 2 Important variables in 90-day readmission risk model.
Abbreviations: LOS, length of stay; BMI, body mass index; BUN, blood urea nitrogen.

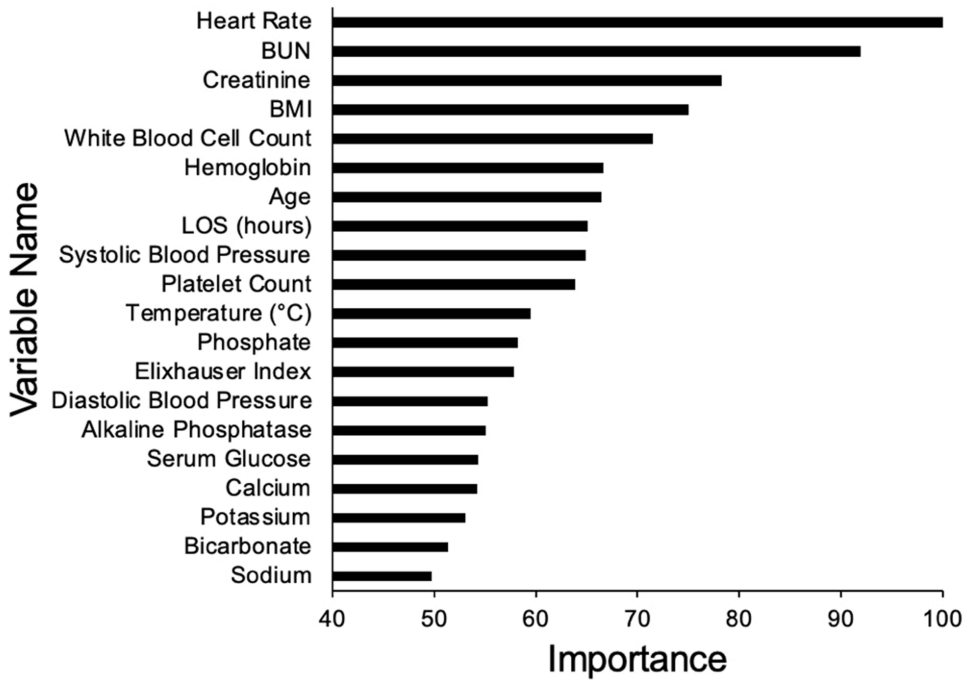


Figure 3 Important variables in 90-day readmission cause model.
Abbreviations: BUN, blood urea nitrogen; BMI, body mass index; LOS, length of stay.

Discussion

To our knowledge, we developed the first pair of machine learning models to predict both the risk for and cause of 90-day readmission after initial hospitalization for AE-COPD. Our approach demonstrated novel functionality and produced a readmission risk model that showed stronger ability to distinguish between patients who were readmitted within 90

days and patients who were not compared to the adapted HOSPITAL score. The readmission cause model also showed similarly strong predictive ability to distinguish between patients readmitted for a respiratory reason and patients admitted for other causes.

Many of the important predictors in our models included variables previously identified by others as important in forecasting readmissions. For example, important variables in our all-cause readmission risk model included length of stay, hemoglobin at discharge, sodium level at discharge, and age, all of which have been reported in other models.^{11,12,30}

A major strength of our machine learning models is that they use only predictors readily available in the EHR, which could theoretically permit use in real-time. This allows for automated prediction of both readmission risk and whether readmission is likely to be respiratory-related at the time of discharge. Other readmission risk tools, such as the COPD-specific PEARL score, include predictors like the eMRCD score which must be manually calculated at the bedside. This places an increased burden on care teams and makes use of these tools in real-time less feasible.

Other models' use of predictors not readily available in the EHR limits our ability to compare different risk tools. Though the HOSPITAL score includes some predictors that were not available in our EHR, we were able to create proxies using the variables available to us. For example, there is not a field in our EHR for index admission type (elective vs urgent or emergent). Therefore, we defined an admission as urgent or emergent if the patient was admitted through the emergency department. While this definition may miss the small percentage of urgent or emergent admissions that come directly from clinic, it likely captures the vast majority of urgent or emergent admissions. We could not compare our all-cause readmission risk model to the PEARL score, because PEARL incorporates the eMRCD score, which was not available to us and had no reasonable proxy.⁹

There are important limitations to consider. First, our models were derived using data from a single center and thus variation between patient populations may limit generalizability. Being a single-center study also limited our sample size. This limited sample size contributed to our decision to adopt the 90-day versus 30-day readmissions threshold employed by some of the previously described studies. It also influenced our decision to define readmission cause in broad categories (as either respiratory or non-respiratory). Larger, multi-site datasets will likely allow for development of improved 30-day real-time COPD-specific risk prediction and for more granular predictions of readmission cause. Importantly, patients who were readmitted to other hospitals within the 90-day timeframe would not have been detected, as this information lays outside of our database. This could be contributing, in part, to the poor calibration seen in the readmission risk model. In addition, true calibration of the HOSPITAL score could not be performed as there are no published probabilities for 90-day readmission at each level of the HOSPITAL score. Lastly, patients were selected for the study based on ICD-9-CM and ICD-10-CM coding algorithms, which are convenient to use and have been shown to have high specificity but lack sensitivity.¹⁴

Conclusion

In conclusion, this study successfully derived and validated novel machine learning models to predict both risk for and cause of 90-day readmission after an index hospitalization for AE-COPD. Our models are designed to run in parallel, glean insights from all EHR data available at the point of discharge. Clinicians may use information from our models to complement their clinical gestalt and select appropriate transitional care that reflects patients' unique risk profiles.

Abbreviations

COPD, chronic obstructive pulmonary disease; EHR, electronic health record; AE-COPD, acute exacerbations of COPD; ICD, International Classification of Diseases; AUROC, area under the receiver operating characteristic curve; HRRP, Hospital Readmissions Reduction Program; eMRCD, extended Medical Research Council Dyspnea; UCM, University of Chicago Medicine; TRIPOD, Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis; CCS, Clinical Classification Software; AHRQ, Agency for Healthcare Research and Quality.

Data Availability Statement

Since these data are highly granular and contains potentially sensitive patient information, public sharing of the data would breach the University of Chicago's IRB protocol requirements. The data used in this study was accessed in a way that was compliant with relevant data protection and privacy regulations per the University of Chicago's IRB (IRB #17-0332). Interested researchers may contact Mary Akel, University of Chicago, via email at makel@medicine.bsd.uchicago.edu for data access requests.

Acknowledgments

We thank the University of Chicago Center for Research Informatics for supporting this project through data curation and Tom Sutton for acting as our Honest Broker. We also thank Mary Akel for administrative assistance.

Funding

This work was supported by funding from the Pritzker School of Medicine Summer Research Program (M Bonomo, M Hermesen, and S Kaskovich) and by a Student Hospitalist Scholar Grant from the Society of Hospital Medicine (M Hemmrich). It was also supported by University of Chicago Medicine Data Science Pilot Funding.

Disclosure

VG Press reports general research support from National Heart Lung and Blood Association (R01HL146644), the Agency for Healthcare Research and Quality (R01HS027804), and the American Lung Association. VG Press also discloses consultant fees from Vizient Inc. and Humana. MM Churpek reports general support from the Department of Defense (W81XWH-21-1-0009), the NIH NIDDK (R01DK126933), NHLBI (R01HL157262) and NIGMS (R01GM123193). MM Churpek also discloses a patent pending (ARCD. P0535US.P2) for risk stratification algorithms for hospitalized patients. None of the other authors have any conflicts of interest in this work.

References

- Centers for Disease Control and Prevention (CDC). FastStats: Chronic Obstructive Pulmonary Disease (COPD) Includes: Chronic Bronchitis and Emphysema; 2022. Available from: <https://www.cdc.gov/nchs/fastats/copd.htm>. Accessed February 24, 2022.
- Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl J Med*. 2009;360(14):1418–1428. doi:10.1056/NEJMsa0803563
- Halpern MT, Stanford RH, Borker R. The burden of COPD in the U.S.A.: results from the confronting COPD survey. *Respir Med*. 2003;97:S81–S89. doi:10.1016/S0954-6111(03)80028-8
- Roberts C, Lowe D, Bucknall C, Ryland I, Kelly Y, Pearson M. Clinical audit indicators of outcome following admission to hospital with acute exacerbation of chronic obstructive pulmonary disease. *Thorax*. 2002;57(2):137–141. doi:10.1136/thorax.57.2.137
- CMS. Hospital Readmissions Reduction Program (HRRP). Available from: <https://www.cms.gov/medicare/medicare-fee-for-service-payment/acuteinpatientpps/readmissions-reduction-program>. Accessed February 24, 2022.
- Shah T, Churpek MM, Coca Perrillon M, Konetzka RT. Understanding why patients with COPD get readmitted. *Chest*. 2015;147(5):1219–1226. doi:10.1378/chest.14-2181
- Allaudeen N, Schnipper JL, Orav EJ, Wachter RM, Vidyarthi AR. Inability of providers to predict unplanned readmissions. *J Gen Intern Med*. 2011;26(7):771–776. doi:10.1007/s11606-011-1663-3
- Press VG. Is it time to move on from identifying risk factors for 30-day chronic obstructive pulmonary disease readmission? A call for risk prediction tools. *Ann Am Thorac Soc*. 2018;15(7):801–803. doi:10.1513/AnnalsATS.201804-246ED
- Press VG, Myers LC, Feemster LC. Preventing COPD readmissions under the hospital readmissions reduction program: how far have we come? *Chest*. 2021;159(3):996–1006. doi:10.1016/j.chest.2020.10.008
- Press VG, Au DH, Bourbeau J, et al. Reducing chronic obstructive pulmonary disease hospital readmissions. An official American thoracic society workshop report. *Ann Am Thorac Soc*. 2019;16(2):161–170. doi:10.1513/AnnalsATS.201811-755WS
- Donzé JD, Williams MV, Robinson EJ, et al. International Validity of the HOSPITAL score to predict 30-day potentially avoidable hospital readmissions. *JAMA Intern Med*. 2016;176(4):496–502. doi:10.1001/jamainternmed.2015.8462
- Echevarria C, Steer J, Heslop-Marshall K, et al. The PEARL score predicts 90-day readmission or death after hospitalisation for acute exacerbation of COPD. *Thorax*. 2017;72(8):686–693. doi:10.1136/thoraxjnl-2016-209298
- Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Crit Care Med*. 2016;44(2):368–374. doi:10.1097/CCM.0000000000001571
- Churpek MM, Yuen TC, Winslow C, et al. Multicenter Development and Validation of a Risk Stratification Tool for Ward Patients. *Am J Respir Crit Care Med*. 2014;190(6):649–655. doi:10.1164/rccm.201406-1022OC
- Churpek MM, Yuen TC, Park SY, Gibbons R, Edelson DP. Using electronic health record data to develop and validate a prediction model for adverse outcomes on the wards. *Crit Care Med*. 2014;42(4):841–848. doi:10.1097/CCM.0000000000000038

16. Stein BD, Bautista A, Schumock GT, et al. The validity of international classification of diseases, ninth revision, clinical modification diagnosis codes for identifying patients hospitalized for COPD exacerbations. *Chest*. 2012;141(1):87–93. doi:10.1378/chest.11-0024
17. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *Ann Intern Med*. 2015;162(1):55–63. doi:10.7326/M14-0697
18. 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;173(8):632–638. doi:10.1001/jamainternmed.2013.3023
19. Rinne ST, Graves MC, Bastian LA, et al. Association between length of stay and readmission for COPD. *Am J Manag Care*. 2017;23(8):e253–e258.
20. Garcia-Aymerich J. Risk factors of readmission to hospital for a COPD exacerbation: a prospective study. *Thorax*. 2003;58(2):100–105. doi:10.1136/thorax.58.2.100
21. Moore BJ, White S, Washington R, Coenen N, Elixhauser A. Identifying increased risk of readmission and in-hospital mortality using hospital administrative data: the AHRQ elixhauser comorbidity index. *Med Care*. 2017;55(7):698–705. doi:10.1097/MLR.0000000000000735
22. Hackmann G, Chen M, Chipara O, et al. Toward a two-tier clinical warning system for hospitalized patients. *AMIA Annu Symp Proc*. 2011;2011:511–519.
23. van den Boogaard M, Pickkers P, Slooter AJC, et al. Development and validation of PRE-DELIRIC (PREdiction of DELIRium in ICu patients) delirium prediction model for intensive care patients: observational multicentre study. *BMJ*. 2012;344:e420. doi:10.1136/bmj.e420
24. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system: risk prediction of hospital mortality for critically III hospitalized adults. *Chest*. 1991;100(6):1619–1636. doi:10.1378/chest.100.6.1619
25. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic Minority Over-sampling Technique. *J Artif Intell Res*. 2002;16:321–357. doi:10.1613/jair.953
26. Qi Y. Random Forest for Bioinformatics. In: Zhang C, Ma Y, editors. *Ensemble Machine Learning: Methods and Applications*. US: Springer; 2012:307–323. doi:10.1007/978-1-4419-9326-7_11
27. Fernandez-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res*. 2014;15(1):3133–3181.
28. Altmann A, Tološi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics*. 2010;26(10):1340–1347. doi:10.1093/bioinformatics/btq134
29. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837–845. doi:10.2307/2531595
30. Min X, Yu B, Wang F. Predictive modeling of the hospital readmission risk from patients' claims data using machine learning: a case study on COPD. *Sci Rep*. 2019;9:2362. doi:10.1038/s41598-019-39071-y

International Journal of Chronic Obstructive Pulmonary Disease

Dovepress

Publish your work in this journal

The International Journal of COPD is an international, peer-reviewed journal of therapeutics and pharmacology focusing on concise rapid reporting of clinical studies and reviews in COPD. Special focus is given to the pathophysiological processes underlying the disease, intervention programs, patient focused education, and self management protocols. This journal is indexed on PubMed Central, MedLine and CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-chronic-obstructive-pulmonary-disease-journal>