Machine Learning Prediction of Mortality and Hospitalization in Heart Failure with Preserved Ejection Fraction

Suveen Angraal, MD,[a,b,*] Bobak J. Mortazavi, PhD,[c,*] Aakriti Gupta, MD,[d] Rohan Khera, MD,[e]
Tariq Ahmad, MD, MPH,[f] Nihar R. Desai, MD MPH,[a,f] Daniel L. Jacoby, MD,[f] Frederick A. Masoudi, MD, MSPH,[g]
John A. Spertus, MD, MPH,[h] Harlan M. Krumholz, MD, SM[a,f,i]

ABSTRACT

OBJECTIVES This study sought to develop models for predicting mortality and heart failure (HF) hospitalization for outpatients with HF with preserved ejection fraction (HFpEF) in the TOPCAT (Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist) trial.

BACKGROUND Although risk assessment models are available for patients with HF with reduced ejection fraction, few have assessed the risks of death and hospitalization in patients with HFpEF.

METHODS The following 5 methods: logistic regression with a forward selection of variables; logistic regression with a lasso regularization for variable selection; random forest (RF); gradient descent boosting; and support vector machine, were used to train models for assessing risks of mortality and HF hospitalization through 3 years of follow-up and were validated using 5-fold cross-validation. Model discrimination and calibration were estimated using receiver-operating characteristic curves and Brier scores, respectively. The top prediction variables were assessed by using the best performing models, using the incremental improvement of each variable in 5-fold cross-validation.

RESULTS The RF was the best performing model with a mean C-statistic of 0.72 (95% confidence interval [CI]: 0.69 to 0.75) for predicting mortality (Brier score: 0.17), and 0.76 (95% CI: 0.71 to 0.81) for HF hospitalization (Brier score: 0.19). Blood urea nitrogen levels, body mass index, and Kansas City Cardiomyopathy Questionnaire (KCCQ) subscale scores were strongly associated with mortality, whereas hemoglobin level, blood urea nitrogen, time since previous HF hospitalization, and KCCQ scores were the most significant predictors of HF hospitalization.

CONCLUSIONS These models predict the risks of mortality and HF hospitalization in patients with HFpEF and emphasize the importance of health status data in determining prognosis. (Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist [TOPCAT]; NCT00094302) (J Am Coll Cardiol HF 2019;■:■–■)
© 2019 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Almost one-half of the patients presenting with heart failure (HF) have preserved ejection fraction (HFpEF), which is responsible for a high health care burden (1). Relative to HF with reduced ejection fraction (HFrEF), the proportion of adverse events related to HFpEF have been increasing, accounting for substantial morbidity, mortality, and cost (2). Hence, it is important to assess the risk factors associated with mortality and hospitalizations in patients with HFpEF while supplementing this knowledge with a risk assessment model. Although many risk assessment models are available for patients with HFrEF (3-9), risk factors associated with adverse outcomes in patients with HFpEF are not well understood. A HFpEF-specific risk model would inform clinicians and patients about their prognosis for this complex syndrome, assisting them in clinical decision making, use of disease management programs, and in discussing end-of-life preferences. Furthermore, it may help to motivate patients to adhere to treatment and help design future clinical trials in HFpEF.

Although HFpEF and HFrEF share common presenting symptoms and poor health status, they are distinct entities with different pathophysiologies (10,11). Hence, the utility of risk assessment models developed in cohorts with HFrEF or a mixture of HFrEF and HFpEF may be of limited value in those with HFpEF. The few existing models for HFpEF may have been limited by their focus on linear relationships in the assessment of risks of mortality and hospitalization (12,13). HFpEF is a complex syndrome with heterogeneous causes and manifestations, which makes it difficult to assess, diagnose, and treat (14). Moreover, the high burden of comorbidities and the unpredictable interplay between different comorbidities is likely to result in multiple HFpEF phenotypes, which makes it a challenging syndrome to study, particularly its outcomes. Hence, there may be value in assessing nonlinear and complex relationships between different patient characteristics in this complex disease, particularly because some patient characteristics, such as age, may be more variable in their association with outcomes than others, such as ejection fraction.

Advanced statistical tools and machine learning methods can improve the prediction over conventional statistical techniques through higher dimensional and possibly nonlinear effects of variables, incorporating a larger number of variables (15). Accordingly, this study used machine learning methods, in addition to conventional logistic regression, to develop and validate models for predicting mortality and hospitalization in patients with HFpEF by using data from the TOPCAT (Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist) trial. With a cohort that was developed using rigorous inclusions to ensure the presence of HFpEF, carefully curated data, and adjudicated outcomes, TOPCAT trial data were ideal to develop a risk assessment model for patients with HFpEF.

METHODS

STUDY POPULATION. TOPCAT was a multicenter, international, randomized, double-blind, placebo-controlled trial sponsored by the U.S. National Heart, Lung, and Blood Institute (NCT00094302), which randomized patients 50 years of age or older with at least 1 sign and at least 1 symptom of HF and a left ventricular ejection fraction (LVEF) of 45% or higher to receive spironolactone or placebo therapy (16). The trial enrolled patients from the United States, Canada, Brazil, Argentina, Russia, and Georgia, from 2006 through 2013, and assessed the incidence of the primary composite outcome of death from cardiovascular causes, aborted cardiac arrest, or hospitalization for HF exacerbation. For the current study, patients from Russia and Georgia were excluded due to regional discrepancies in the reported and actual use of spironolactone and improper application of the selection criteria during enrolment (17).

FIGURE 1 Analysis Flow for the Development and Evaluation of Models

TOPCAT = Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist trial.
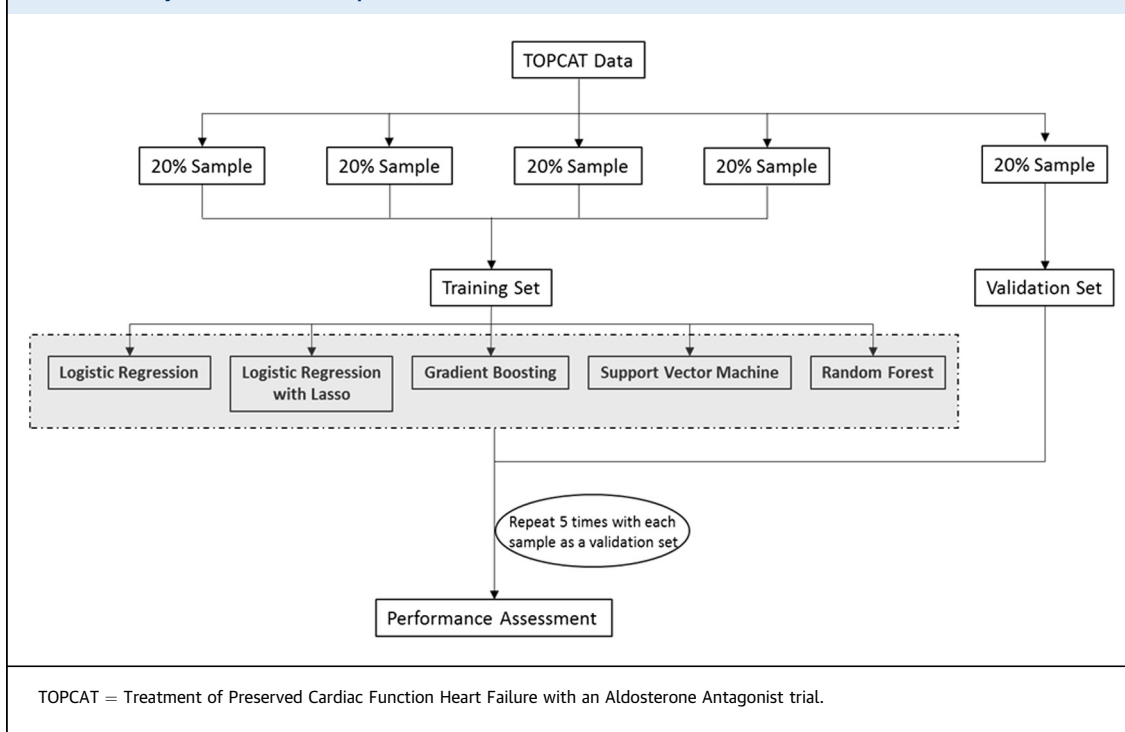
OUTCOMES. The outcomes of interest were all-cause mortality and HF hospitalization through 3 years of follow-up. All-cause mortality was defined as death due to any cause, and HF hospitalization was defined as an unexpected presentation to an acute care facility requiring overnight hospitalization with exacerbation of HF. HF as the cause of hospitalization was confirmed using a prespecified list of signs and symptoms present upon admission requiring treatment and was adjudicated by an independent endpoints committee. Kaplan-Meier survival plots were constructed to assess mortality and HF hospitalization in the study population.

CANDIDATE VARIABLES. All baseline demographic and clinical data available from patients were used in addition to laboratory data, electrocardiography data, Kansas City Cardiomyopathy Questionnaire (KCCQ) scores (physical limitation score, symptom stability score, symptom frequency score, symptom burden score, total symptom score, self-efficacy score, quality of life score, social limitation score, overall summary score, and clinical summary score) to predict the outcomes of all-cause mortality and HF hospitalization.

MODEL DEVELOPMENT. To develop derivation and validation cohorts, a stratified, 5-fold cross-validation was used (**Figure 1**). The study population was divided randomly into 5 subsets with similar event rates. To form the derivation cohort, 4 subsets (80%) were combined, and the remaining subset (20%) was reserved as the validation set. This process was repeated 5 times for each outcome, such that every subset served as the validation set, thereby accounting for variability among patients and providing risk estimates for all cases. With low missing data, mean imputation was used; 30 variables had no missing data, 44 variables had <10% data missing, and the remainder (30 variables) had >10% missing data. Any variable with >50% missing data (18 of the 30 variables) was removed. A total of 86 variables were included in the models. The list of candidate variables along with their definitions are included in Online Table 1.

Four commonly used machine learning methods and conventional logistic regression were used to train the models for assessing risk of mortality and HF hospitalization through 3 years of follow-up. The methods included logistic regression with a forward selection of variables, logistic regression with a lasso regularization for variable selection, random forest (RF), gradient descent boosting decision trees, and support vector machine (SVM). Various R packages (R software: R Core Team, Vienna, Austria) were used to conduct this analysis. The base GLM function was used for logistic regression. The glmnet package was

used for logistic regression with lasso regularization (18). randomForest package was used for the RF model (19); xgboost was used for the gradient descent boosting (20); and e1071 (libSVM) software was used for the SVM (21).

Logistic regression was used with forward stepwise selection to choose variables with statistically significant results in a likelihood ratio test (p < 0.05). A forward stepwise selection through this ordered list was run to identify the mean incremental C-statistic improvements for each added variable (22). In high-dimension problems, backward selection techniques may be susceptible to greater noise (23). Forward selection has strong theoretical guarantees and excellent empiric behavior (24). Logistic regression with lasso regularization builds a similar logistic regression model, but it retains the variables using more stringent thresholds, thus helping to select a parsimonious, predictive subset of variables to train the logistic regression model. Pre-built default lambda parameters in glmnet package were used through 10-fold cross-validation (18). RF is a method by which a number of decision trees (tree-like graph of decisions and their possible consequences, including event outcomes) are built from the variable set. These decision trees were used to divide patients into similar subgroups by using the most important variables. The prediction is then generated using a "voting" scheme across all decision trees. Gradient descent boosting similarly picks variables across decision trees that best help predict correct outcomes in their training sets. Both RF and gradient descent boosting internally validate their selection of variables and cases within the training set. A total of 1,000 trees were used for RF and 100 trees for boosting (to avoid overfitting), with a learning rate of 0.1 and a maximum depth of 6 for each tree. Finally, SVM is a predictive method that tries to find a separable space between 2 classes in order to generate positive or negative predictions. The SVM in this study was given the entirety of a dataset and trained with 2 different kernels (or patterns to find the separation function), which were linear and radial-based functions. These kernels identify relationships of variables within the dataset by comparing linear separation between the outcomes of interest or a Gaussian separation between the outcomes of interest.

MODEL PERFORMANCE. Receiver operating characteristic (ROC) curves were used to estimate model discrimination by calculating the C-statistic or area under the curve (AUC). The best performing model assessed by highest AUC was chosen and analyzed further. The accuracy of probability of the best performing model was assessed using the Brier score, which is defined as the mean squared difference between the observed and predicted outcome. Brier scores range from 0 to 1.00, with 0 representing the best possible calibration. The 2 primary components decomposed from Brier score are reliability and resolution, which measure how close the prediction probabilities are to the true probabilities and how much the conditional probabilities differ from the prediction average, respectively. Calibration plots were used to plot the mean risk score relative to the observed outcome rate for a given decile of predicted risk. The prediction for every patient was plotted in order of their risk to assess the prediction distribution from the model.

TOP PREDICTORS. The importance of each variable in a model was evaluated by using a variety of metrics. For logistic regression with forward selection, importance was ranked by ordering the variables that were selected. A standard statistical procedure for the ordering of variables was followed. At each iteration, 1 additional variable was added, and the maximum likelihood was measured by using a likelihood ratio test. For any variable with a p value < 0.05, the largest maximum likelihood variable was selected. This process was repeated until no variables had p significance <0.05. RF uses a mean decrease in accuracy to rank the importance of its variables. Boosting determines variable importance by the number of trees it appears in and the information gain that variable provides. The importance of each variable was evaluated in the best performing model. To avoid overfitting, model discrimination was evaluated through cross-validation, demonstrating a good fit by identifying a narrow confidence interval (25). Each training iteration can present a different order of variables; thus, variable importance was calculated from a model trained on all data. It was ensured that this model was representative of the important variables without overfitting based on the good fit demonstrated in the cross-validation. These selected variables were listed in the order of importance, and were evaluated by the incremental improvement of each variable in the 5-fold cross-validation data. This may overestimate the gain each variable provides, but the final values are similar to those obtained through cross-validation.

SENSITIVITY ANALYSES. Sensitivity analyses were performed for all patients enrolled in the trial from the Americas. The models were developed on this population for the prediction of mortality and

TABLE 1 Baseline Characteristics of Patients (N = 1,767)

Age, yrs	72 (64–79)
Women	882 (49.9)
White patients	1,384 (78.3)
LVEF, %	58 (53–64)
NYHA functional class	
I and II	1,142 (64.6)
III and IV	620 (35.1)
Current smoker	114 (6.5)
Blood pressure, mm Hg	
Systolic	129 (118–138)
Diastolic	70 (62–80)
Heart rate, beats/min	68 (61–76)
Body mass index, kg/m^2	32.9 (28.0–38.4)
Serum potassium, mmol/l	4.2 (3–6)
Estimated GFR, ml/min/1.73 m^2	61.2 (49.0–76.6)
Hemoglobin, g/dl	12.8 (11.7–14.0)
Kansas City Cardiomyopathy Questionnaire scores	
Physical limitation	58.3 (37.5–79.2)
Symptoms	62.5 (41.7–81.2)
Social limitation	62.5 (31.3–87.5)
Self-efficacy	75.0 (62.5–100.0)
Quality of life	58.3 (33.3–75.0)
Overall summary	59.4 (39.6–77.1)
Clinical summary	59.1 (41.9–77.6)

Values are median (interquartile range) or n (%).

TABLE 2 Results of the Prediction Models for Mortality and HF Hospitalization Over 3 Years' Follow-Up in Patients With HF With Preserved Ejection Fraction

	Mean AUC for Mortality	Mean AUC for HF Hospitalization
Logistic regression with a forward selection	0.66 (0.62–0.69)	0.73 (0.66–0.80)
Logistic regression with a lasso regularization	0.65 (0.61–0.70)	0.73 (0.67–0.79)
Random forest	0.72 (0.69–0.75)	0.76 (0.71–0.81)
Gradient descent boosting	0.68 (0.66–0.71)	0.73 (0.69–0.77)
Support vector machine (linear kernel)	0.66 (0.60–0.72)	0.72 (0.63–0.81)
Support vector machine (radial basis function)	0.65 (0.59–0.72)	0.72 (0.65–0.79)

AUC = area under the curve; HF = heart failure.

Kaplan-Meier survival plots for mortality and HF hospitalization are shown in Online Figure 1. A gradual decline in survival was observed for both mortality and HF hospitalization over 3 years. Mortality survival rate was 0.92 at 1 year, and 0.82 at 2 years follow-up. Furthermore, the HF hospitalization survival rate was 0.78 at 1 year and 0.69 at 2 years' follow-up.

MACHINE LEARNING FOR PREDICTION OF OUTCOMES. The results of the 5 methods are shown in Table 2. Of the 5 methods, RF performed the best, with the highest overall C-statistic. The RF models, over 3 years' follow-up, achieved a mean C-statistic of 0.72 (95% confidence interval [CI]: 0.69 to 0.75) for mortality and 0.76 (95% CI: 0.71 to 0.81) for HF hospitalization. This was in contrast to the logistic regression model; C-statistics of 0.66 (95% CI: 0.62 to 0.69) and 0.73 (95% CI: 0.66 to 0.80) for mortality and HF hospitalization, respectively.

PREDICTOR VARIABLES. The variables were ranked in order of importance by RF with the forward, stepwise C-statistic when added in the rank order to the model for mortality and HF hospitalization (Table 3, Online Table 2). The blood urea nitrogen (BUN) level and body mass index, along with variables pertaining to the KCCQ were top predictors of mortality over 3 years. Furthermore, alkaline phosphatase level and age were highly predictive for mortality. For HF hospitalization over 3 years, top predictors included hemoglobin level, BUN level, KCCQ variables, and time since previous HF hospitalization. Other predictor variables for HF hospitalization were glomerular filtration rate and blood glucose levels.

CALIBRATION OF THE MODELS AND PREDICTION DISTRIBUTIONS. The final models were well calibrated. For predicting mortality over 3 years' follow-up, the mean Brier score of the model was 0.17, with a reliability of 0.001 and a resolution of 0.019. For predicting HF hospitalization over 3 years' follow-up, the mean Brier score was 0.19; reliability was 0.004,

hospitalization, followed for the entirety of the study period. Variable importance was calculated, and the incremental improvement of each variable was evaluated in the 5-fold cross-validation. Furthermore, a 1-year prediction of mortality and HF hospitalization was assessed to find out how the models' performance varied over different shorter follow-up times. All analyses were conducted using R version 3.4.0.

RESULTS

PATIENT CHARACTERISTICS. A total of 1,767 patients with HFpEF were included from 4 countries (United States, Canada, Argentina, and Brazil); 1,088 of these patients were followed for at least 3 years or died within 3 years. The baseline characteristics of the patients are shown in Table 1. The study population included 49.9% women; 78.3% were white. The median LVEF was 58%, with more than one-third of the study population presenting with New York Heart Association functional class symptoms of III or IV. A total of 387 patients (22%) died during the trial follow-up, and 400 (23%) were hospitalized for HF. At 3 years' follow-up, a total of 268 patients had died (24.6% of the patients enrolled for at least 3 years), and 343 were hospitalized for exacerbation of HF (31.5% of the patients enrolled for at least 3 years).

TABLE 3 Top Predictors of Mortality and Heart Failure Hospitalization for Patients With Heart Failure With Preserved Ejection Fraction

| Risk of Mortality	AUC Increment (95% Confidence Interval)	Risk of Heart Failure Hospitalization	AUC Increment (95% Confidence Interval)
Blood urea nitrogen level, mg/dl	0.63 (0.60-0.66)	Hemoglobin level, g/dl	0.59 (0.54-0.64)
KCCQ clinical summary score	0.62 (0.59-0.65)	Blood urea nitrogen level, mg/dl	0.60 (0.54-0.66)
KCCQ overall summary score	0.62 (0.57-0.67)	Time since previous hospitalization for heart failure, yrs	0.64 (0.60-0.68)
KCCQ social limitation score	0.61 (0.57-0.65)	KCCQ physical limitation score	0.67 (0.63-0.70)
Body mass index, kg/m²	0.65 (0.61-0.70)	Glomerular filtration rate, ml/min	0.68 (0.64-0.72)
KCCQ physical limitation score	0.67 (0.62-0.73)	KCCQ clinical summary score	0.70 (0.66-0.74)
KCCQ symptom frequency score	0.66 (0.61-0.71)	KCCQ overall summary score	0.70 (0.66-0.73)
KCCQ total symptom score	0.67 (0.62-0.72)	KCCQ social limitation score	0.70 (0.67-0.73)
KCCQ quality of life score	0.67 (0.61-0.74)	Glucose level, mg/dl	0.71 (0.67-0.75)
Alkaline phosphatase level, U/l	0.69 (0.64-0.74)	Urine microalbumin/creatinine ratio, mg/g	0.72 (0.68-0.76)
KCCQ symptom burden score	0.69 (0.63-0.75)	KCCQ total symptom score	0.72 (0.69-0.76)
Age on entering the study	0.70 (0.66-0.74)	White blood cell count, k/μl	0.73 (0.70-0.77)
Is subject of Hispanic, Latino, or Spanish origin?	0.70 (0.67-0.74)	KCCQ symptom frequency score	0.74 (0.70-0.77)
Glucose level, mg/dl	0.69 (0.65-0.73)	Previous hospitalization for heart failure	0.74 (0.70-0.77)
Potassium level, mmol/l	0.69 (0.66-0.72)	Patient is black	0.74 (0.71-0.77)
White blood cell count, k/μl	0.70 (0.67-0.73)	Patient is white	0.74 (0.71-0.77)
Total bilirubin level, mg/dl	0.70 (0.67-0.73)	NYHA functional class	0.74 (0.71-0.77)
Alanine aminotransferase level, U/l	0.70 (0.68-0.72)	Albumin level, g/dl	0.74 (0.72-0.77)
Heart rate, beats/min	0.70 (0.68-0.72)	KCCQ quality of life score	0.74 (0.72-0.77)
Diastolic blood pressure, mm Hg	0.71 (0.71-0.72)	Orthopnea present at screening	0.74 (0.72-0.76)

KCCQ = Kansas City Cardiomyopathy Questionnaire.

and resolution was 0.044. A Brier score closer to 0 (and similar reliability and resolution) provided measurements of better calibration. Online Figure 2 shows the calibration plots for the models. The prediction distribution plots of the models with patients sorted in the order of risk show positive clustering of patients who either died or were hospitalized with HF exacerbation (Figure 2), asserting that the models accurately stratified patients at risk of mortality and hospitalization.

RESULTS OF THE SENSITIVITY ANALYSIS. For the risk of mortality and hospitalization over the entire study period, the C-statistics for mortality and HF hospitalization were 0.70 (95% CI: 0.67 to 0.73) and 0.69 (95% CI: 0.67 to 0.71), respectively, through the RF method. Furthermore, similar variables were as strongly associated with mortality and HF hospitalization as those for 3 years' risk of mortality and HF hospitalization but with different rank order of importance (Online Table 3). These models were well calibrated, with a mean Brier score of 0.16 for mortality, 0.002 for reliability, and 0.016 for resolution; and a mean Brier score of 0.16 for HF hospitalization, 0.0008 for reliability, and 0.014 for resolution (Online Figure 2). The assessment of risks of mortality and HF hospitalizations at 1 year follow-up were comparable in the 5 models (Online Table 4).

DISCUSSION

Accurately predicting prognosis is fundamental to patient-centered care, both in selecting treatment strategies and informing patients as a foundation for shared decision making. Using data from an outpatient cohort of individuals with HFpEF followed in the TOPCAT trial, this study explored 5 alternative statistical methods to build risk models for mortality and hospitalization in patients with HFpEF. Ultimately it was found that an RF model best stratified patients' risks with good internal validation and excellent calibration (Central Illustration). Furthermore, clinical characteristics of patients' risk were also identified that may be underappreciated in clinical practice. In particular, patients' health status, as quantified by the KCCQ scores, was among the strongest predictors of both mortality and HF hospitalization. These models form the foundation for future testing of the clinical utility of more accurate risk stratification of patients' care and outcomes.

Although published reports are abundant with prediction models to assess risk in patients with HF, the present study extends this knowledge in several important ways. First, models have been generated that are specific to individuals with HFpEF. One of the most widely used risk prediction models, the
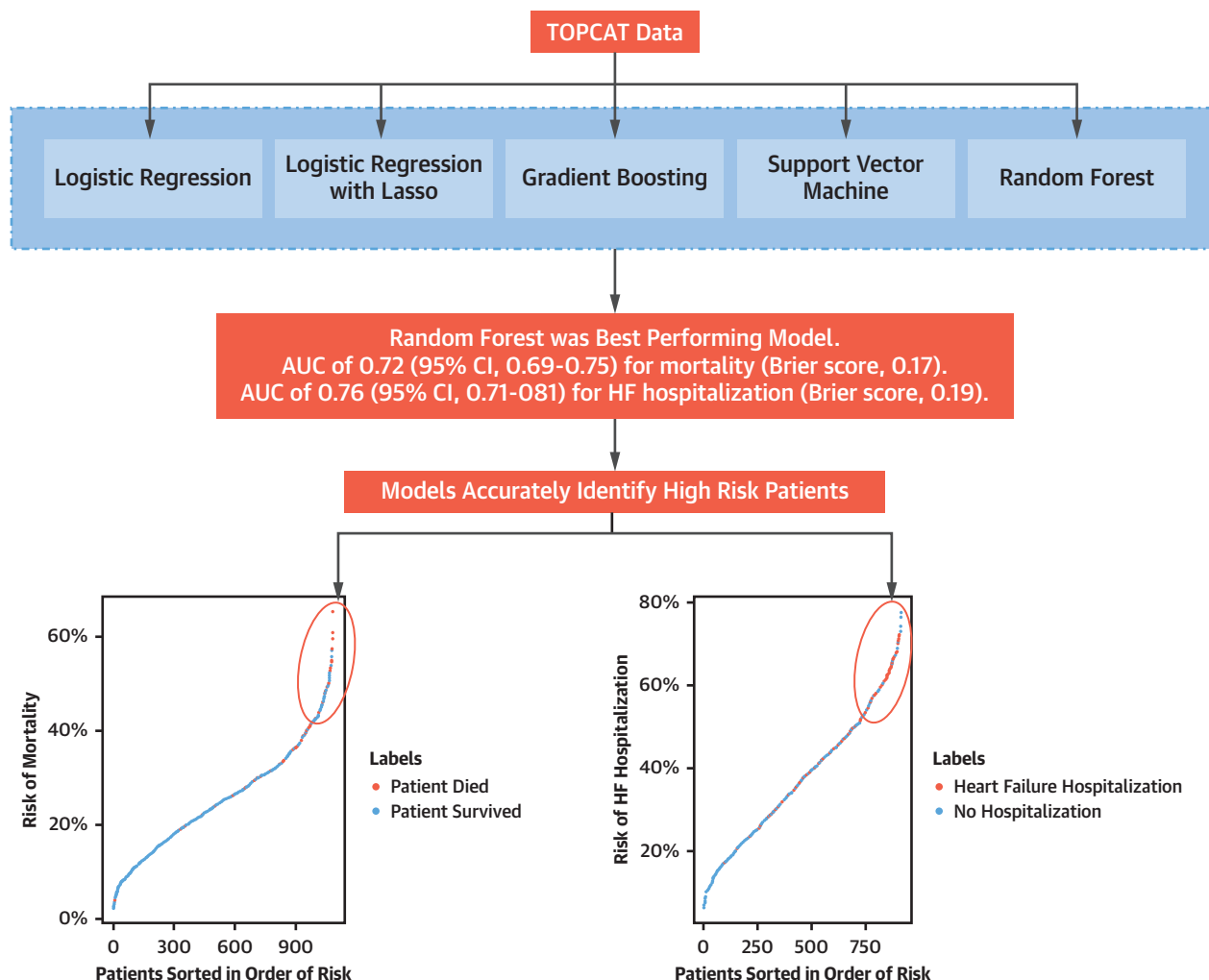
FIGURE 2 Prediction Distributions of Patients With HFpEF
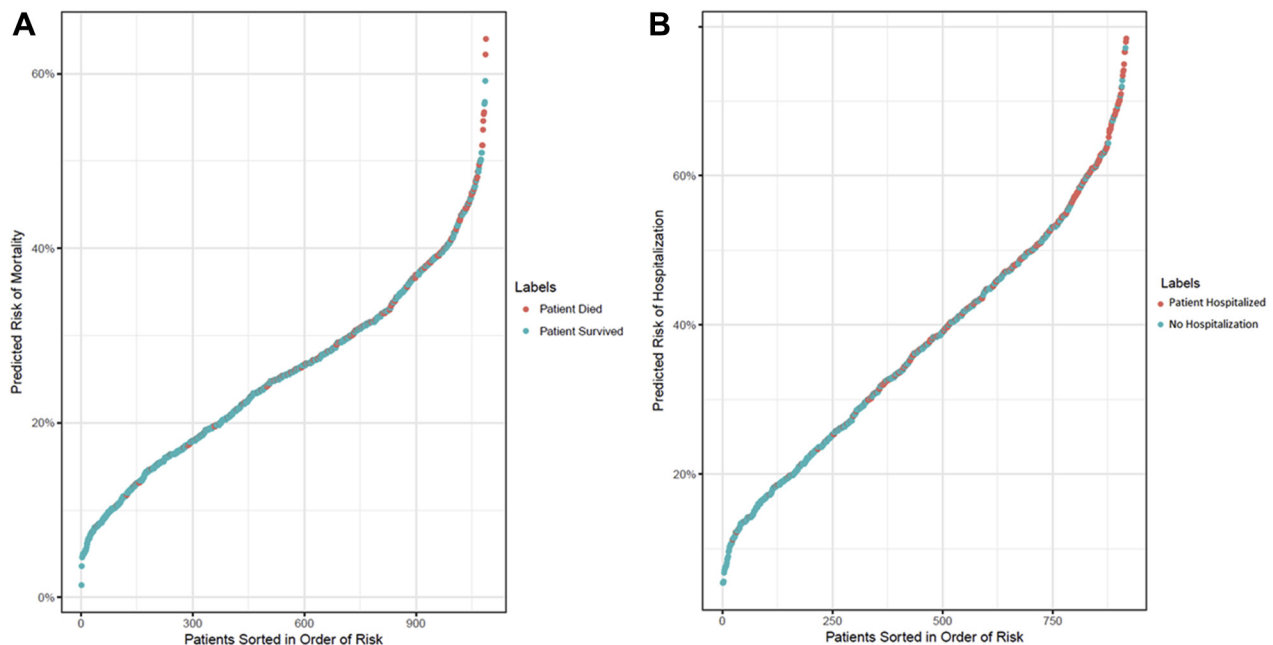
Prediction distributions of patients with HFpEF according to **(A)** the risk of mortality and **(B)** the risk of heart failure hospitalization. HFpEF = heart failure with preserved ejection fraction.

Seattle Heart Failure Model, achieved a C-statistic of 0.73 for mortality (9). However, the derivation cohort consisted of patients with LVEF <30%, and did not include health status, which in this study was informative. Second, advanced machine learning methods were used to assess risk in this complex syndrome, which previously have been shown to improve the prediction of adverse events in patients with HF (15). The previously available HFpEF risk models, I-PRE-SERVE (Irbesartan in Heart Failure with Preserved Ejection Fraction) and MAGGIC (Meta-Analysis Global Group in Chronic Heart Failure) score, used Cox proportional hazard models to assess the risk of adverse outcomes (12,13). Third, the present study used carefully curated data from a clinical trial with adjudicated outcomes to develop the risk prediction models, allowing the use of a comprehensive set of variables to account for complex interactions for more accurate prediction of mortality and HF hospitalization. Neither the I-PRESERVE nor the MAGGIC model included important clinical information which may be pertinent to risk assessment in HFpEF. Moreover, the I-PRESERVE model did not assess the risk of HF hospitalization independently. Other studies have used observational datasets

predominantly consisting patients with HFrEF (3-8). Given that observational data may be more applicable to routine clinical care, replicating the RF model in a broader spectrum of patients would further support its clinical utility. Fourth, the present study identifies important predictors of mortality and hospitalization in patients with HFpEF, such as health status and quality of life, which are not routinely collected but may be critical in risk assessment.

Health status and quality of life variables were highly predictive of mortality and HF hospitalization in patients with HFpEF. Although studies have shown that KCCQ scores provide prognostic information for mortality and hospitalizations in patients with HFrEF (26), the present study shows an association between health status variables and outcomes in HFpEF patients, confirming the insights provided from 2 prior studies (27,28) while supplementing these data with a full prediction model. Moreover, several composites of the KCCQ were enabled to enter the model, including the total symptom, clinical and overall summary scores, as well as the individual domains of symptom frequency, symptom burden, physical and social limitation, and quality of life. The fact that many of these combined and individual domains

CENTRAL ILLUSTRATION Prediction Models to Assess the Risks of Mortality and Heart Failure Hospitalization in Patients With Heart Failure With Preserved Ejection Fraction

TOPCAT Data

| Logistic Regression | Logistic Regression with Lasso | Gradient Boosting | Support Vector Machine | Random Forest |

Random Forest was Best Performing Model.
AUC of 0.72 (95% CI, 0.69-0.75) for mortality (Brier score, 0.17).
AUC of 0.76 (95% CI, 0.71-081) for HF hospitalization (Brier score, 0.19).

Models Accurately Identify High Risk Patients

Angraal, S. et al. J Am Coll Cardiol HF. 2019;■(■):■-■.

These models accurately identify high risk patients who died or were hospitalized for heart failure exacerbation. AUC = area under the curve; HF = heart failure; TOPCAT = Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist trial.

were among the strongest predictors of outcomes underscores the importance of these patient-reported variables. Despite the evidence associating health status with other important clinical outcomes, these data are not collected in the routine care of patients with HF. Our study emphasizes the advantage of collecting these data beyond the inherent value of quantifying patients' symptoms and quality of life.

Other predictors of mortality and hospitalization from the present study differ from the predictors reported in previous studies (29). BUN level, body mass index, and health status were predictive of death,

whereas hemoglobin level, BUN, time since previous HF hospitalization, and health status were predictive of HF hospitalization. However, in a recent systematic review of 117 HF prediction models, using predominantly patients with HFrEF, investigators found that BUN, sodium levels, systolic blood pressure, and age had the highest prediction values for mortality, whereas BUN levels, sodium levels, and race had the highest prediction values for HF hospitalization (29). Although some variables, such as age, that predict survival may have been influenced by the inclusion criteria for TOPCAT or the availability of specific data

elements such as the KCCQ scores, these findings may also reflect the fact that HFpEF has different risk factors for mortality and hospitalization than its counterpart, HFrEF.

STUDY LIMITATIONS. First, the models use baseline patient characteristics without follow-up data. Although a dynamic model incorporating the time-varying values of baseline data may be superior, the present models predict mortality and HF hospitalization by using clinical data that can be acquired with reasonable accuracy and used at a point in time to predict long-term prognosis. Second, the study population used in the development of models was obtained from a clinical trial, and results may not be applicable to a broader population or to those with additional comorbidities. Although additional work in less selected populations would be important, the RF models can be readily updated with new data. Third, the present study used prediction analysis which did not include time-to-event analysis. Although a method that investigates these advanced machine learning techniques along with time-to-event analysis would be superior, many such techniques are not currently suited to execute such analysis. However, the present authors have provided survival plots for the readers to understand the survival patterns in this study population. Fourth, there may remain additional data (e.g., imaging data, novel biomarkers, atherosclerotic burden, and environment factors) that could further improve prediction. Future work could explore the addition of such variables in further improving the RF models proposed here.

CONCLUSIONS

Using advanced machine learning techniques and easily obtainable patient characteristics, the present models predict the risk of mortality and HF hospitalizations in patients with HFpEF. Furthermore, the results emphasize the fact that the quality of life and health status data that are not routinely collected in a clinical encounter have a profound impact on outcomes in patients with HFpEF. These models may be used by clinicians as a decision-making tool to estimate the prognosis of patients suffering from HFpEF.

ADDRESS FOR CORRESPONDENCE: Dr. Harlan M. Krumholz, 1 Church Street, Suite 200, New Haven, Connecticut 06510. E-mail: harlan.krumholz@yale.edu. Twitter: @hmkyale.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: These models may be used by clinicians as decision-making tools to estimate the prognosis of patients with HFpEF, allowing for a more efficient use of treatment strategies, shared decision making, and identification of high-risk patients for more intensive treatment.

TRANSLATIONAL OUTLOOK: Although significant advances have been made in the management of HF, few tools exist to assess risks of mortality and hospitalization in patients with HFpEF. These models form the foundation for future testing of the clinical utility of more accurate risk stratification.

REFERENCES

1. Borlaug BA, Paulus WJ. Heart failure with preserved ejection fraction: pathophysiology, diagnosis, and treatment. Eur Heart J 2011;32:670–9.

2. Redfield MM, Jacobsen SJ, Burnett JC Jr., Mahoney DW, Bailey KR, Rodeheffer RJ. Burden of systolic and diastolic ventricular dysfunction in the community: appreciating the scope of the heart failure epidemic. JAMA 2003;289:194–202.

3. O'Connor CM, Hasselblad V, Mehta RH, et al. Triage after hospitalization with advanced heart failure: the ESCAPE (Evaluation Study of Congestive Heart Failure and Pulmonary Artery Catheterization Effectiveness) risk model and discharge score. J Am Coll Cardiol 2010;55:872–8.

4. Subramanian D, Subramanian V, Deswal A, Mann D. New predictive models of heart failure mortality using time-series measurements and ensemble models. Circ Heart Fail 2011;4:456–62.

5. Amarasingham R, Moore BJ, Tabak YP, et al. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. Med Care 2010;48:981–8.

6. Postmus D, van Veldhuisen DJ, Jaarsma T, et al. The COACH risk engine: a multistate model for predicting survival and hospitalization in patients with heart failure. Eur J Heart Fail 2012;14:168–75.

7. Kearney MT, Nolan J, Lee AJ, et al. A prognostic index to predict long-term mortality in patients with mild to moderate chronic heart failure stabilised on angiotensin converting enzyme inhibitors. Eur J Heart Fail 2003;5:489–97.

8. Peterson PN, Rumsfeld JS, Liang L, et al. A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association get with the guidelines program. Circ Cardiovasc Qual Outcomes 2010;3:25–32.

9. Levy WC, Mozaffarian D, Linker DT, et al. The Seattle heart failure model: prediction of survival in heart failure. Circulation 2006;113:1424–33.

10. Lee DS, Gona P, Vasan RS, et al. Relation of disease pathogenesis and risk factors to heart failure with preserved or reduced ejection fraction: insights from the framingham heart study of the national heart, lung, and blood institute. Circulation 2009;119:3070–7.

11. Hogg K, Swedberg K, McMurray J. Heart failure with preserved left ventricular systolic function; epidemiology, clinical characteristics, and prognosis. J Am Coll Cardiol 2004;43:317–27.

12. Rich Jonathan D, Burns J, Freed Benjamin H, et al. Meta-analysis global group in chronic (maggic) heart failure risk score: validation of a simple tool for the prediction of morbidity and mortality in heart failure with preserved ejection fraction. J Am Heart Assoc 2018;7:e009594.

13. Komajda M, Carson PE, Hetzel S, et al. Factors associated with outcome in heart failure with preserved ejection fraction: findings from the Irbesartan in Heart Failure with Preserved Ejection Fraction Study (I-PRESERVE). Circ Heart Fail 2011;4:27–35.

14. Oktay AA, Rich JD, Shah SJ. The emerging epidemic of heart failure with preserved ejection fraction. Curr Heart Fail Rep 2013;10:401-10.

15. Mortazavi BJ, Downing NS, Bucholz EM, et al. Analysis of machine learning techniques for heart failure readmissions. Circ Cardiovasc Qual Outcomes 2016;9:629-40.

16. Pitt B, Pfeffer MA, Assmann SF, et al. Spironolactone for heart failure with preserved ejection fraction. N Engl J Med 2014;370:1383-92.

17. de Denus S, O'Meara E, Desai AS, et al. Spironolactone metabolites in TOPCAT— new insights into regional variation. N Engl J Med 2017;376: 1690-2.

18. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33:1-22.

19. Liaw A, Wiener M. Classification and Regression by randomForest. R News 2010. Available at: https://www.researchgate.net/profile/Andy_Liaw/publication/228451484_Classification_and_Regression_by_RandomForest/links/53fb24cc0cf2 0a45497047ab/Classification-and-Regression-by-RandomForest.pdf. Accessed July 18, 2019.

20. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd Association for Computing Machinery (ACM) Special Interest Group Knowledge Discovery and Data Mining (SIGKDD) International Conference on Knowledge Discovery and Data Mining Papers presented at: Proceedings of ACM SIGKDD; August 13-17, 2016; San Francisco, CA. In: Krishnapuram B, editor. ACM 2016;785:94.

21. Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2011;2:27.

22. Hosmer DW, Wang C-Y, Lin IC, Lemeshow S. A computer program for stepwise logistic regression using maximum likelihood estimation. Comput Programs Biomed 1978;8:121-34.

23. Zhang T. On the consistency of feature selection using greedy least squares regression. Journal of Machine Learning Research. 2009. Available at http://www.jmlr.org/papers/volume10/zhang09a/zhang09a.pdf. Accessed July 18, 2019.

24. Friedman J, Hastie T, Tibshirani R. The Elements Of Statistical Learning: Series in Statistics. Berlin; Springer; 2001;219:259.

25. Wasserman L. All of statistics: A Concise Course In Statistical Inference. Berlin; Springer Science & Business Media;2013;362: 264.

26. Heidenreich PA, Spertus JA, Jones PG, et al. Health status identifies heart failure outpatients at risk for hospitalization or death. J Am Coll Cardiol 2006;47:752-6.

27. Joseph SM, Novak E, Arnold SV, et al. Comparable performance of the Kansas City Cardiomyopathy Questionnaire in patients with heart failure with preserved and reduced ejection fraction. Circ Heart Fail 2013;6:1139-46.

28. Pokharel Y, Khariton Y, Tang Y, et al. Association of serial Kansas City cardiomyopathy questionnaire assessments with death and hospitalization in patients with heart failure with preserved and reduced ejection fraction: a secondary analysis of 2 randomized clinical trials. JAMA Cardiol 2017;2:1315-21.

29. Ouwerkerk W, Voors AA, Zwinderman AH. Factors influencing the predictive power of models for predicting mortality and/or heart failure hospitalization in patients with heart failure. J Am Coll Cardiol HF 2014;2:429-36.

KEY WORDS health status, HFpEF, KCCQ, risk

APPENDIX For supplemental figures and tables, please see the online version of this paper.