ORIGINAL RESEARCH

The Colorado Heart Failure Acuity Risk Model

A Mortality Model for Waitlisted Cardiac Transplant Patients

Rachel D. Murphy, MD,[a] Sarah Y. Park, MD,[b] Larry A. Allen, MD, MHS,[b] Amrut V. Ambardekar, MD,[b] Joseph C. Cleveland, Jr, MD,[c] Michael T. Cain, MD,[c] Bruce Kaplan, MD,[d,e] Jordan R.H. Hoffman, MPH, MD,[c] John S. Malamon, PhD[d,e]

ABSTRACT

BACKGROUND Currently, there is no mathematical model used nationally to determine the medical urgency of patients on the heart transplant waitlist in the United States. While the current organ distribution system accounts for many patient factors, a truly objective model is needed to more reliably stratify patients by their medical acuity.

OBJECTIVES The aim of the study was to develop risk scores (Colorado Heart failure Acuity Risk Model [CHARM] score) to predict mortality in adults waitlisted for heart transplant.

METHODS Risk scores were based on multivariable logistic regression models with mortality endpoints at 90 days, 180 days, 1 year, and 2 years. The models included serology data and patient history variables from waitlisted patients (N = 4,176) within the Scientific Registry of Transplant Recipients database from January 1, 2017, to September 2, 2023.

RESULTS The CHARM score included serum markers (brain natriuretic peptide, creatinine, sodium, aspartate amino-transferase, albumin, total bilirubin) and clinical variables (history of cardiac surgery, prior transplant, willingness to accept an hepatitis C virus positive heart, use of extracorporeal membrane oxygenation, use of mechanical life support, implantation of a cardiac defibrillator, and ventilator support prior to transplant). Sample holdout-validation for the models yielded average area under the curves of 0.825 (90-day), 0.805 (180-day), 0.779 (1-year), and 0.766 (2-year). Risk indices for all models were 99% correlated with observed mortality rates.

CONCLUSIONS The CHARM score provides reliable calibration and prediction, offering an objective system for identifying critically ill patients on the heart transplant waitlist. The CHARM score will be useful in the era of continuous distribution to standardize organ allocation. (JACC Adv. 2025;4:101449) © 2024 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/).

From the [a]Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA; [b]Department of Medicine, Division of Cardiology, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA; [c]Department of Surgery, Division of Cardiothoracic Surgery, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA; [d]Department of Surgery, Division of Transplant Surgery, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA; and the [e]Colorado Center for Transplantation Care, Research and Education (CCTCARE), Aurora, Colorado, USA.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

The Organ Procurement and Transplantation Network (OPTN) has worked to develop a cardiac transplant allocation system that distributes donor hearts to the most critically ill patients.[1] In 2018, a new heart allocation policy revision was enacted to convert the 3-tier to a 6-tier status-based patient risk stratification model that attempted to encompass the increasing complexity of managing critical cardiac illness. However, this policy has been critiqued for its subjectivity and heterogeneity in accurately discriminating patient risk while on the waitlist. For example, a large percentage of transplant candidates are now stratified as status 1 or status 2 by exception rather than by the standard criteria, which can be modified by a physician's practice.[2] Early studies on post-transplant survival have demonstrated that the 2018 policy revision was associated with a significant reduction in post-transplant survival.[3,4] However, in 2022, Lazenby et al used the multivariable Cox proportional hazards model to demonstrate the effect of inadequate follow-up on postpolicy survival. They concluded that there is no significant difference in 1-year post-transplant survival under the new heart allocation policy.[5] A formal prognostic model that accurately stratifies waitlisted end-stage heart failure (HF) patients based on medical urgency is needed to provide objectivity, accuracy, and transparency. The need for such a model was recently highlighted by Pelzer et al, who determined the current allocation system had only a moderate ability to successfully rank transplant candidates according to medical urgency.[6]

While useful prognostic tools for patients with HF have been introduced, such as the Heart Failure Survival Score[7] and the Seattle Heart Failure Model,[8] they have not accurately predicted patient waitlist mortality.[9] The Ottawa Heart Failure Risk Scale provided a risk stratification tool designed for acute heart failure patients in emergency departments.[10] However, this score has not been applied to or validated against a waitlisted heart transplant patient population. Similarly, the Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC) risk score performed well for HF patients with preserved ejection fraction, but it has not been externally validated for reduced ejection fraction, which excludes most patients on a heart transplant waitlist.[11] More recently, the U.S.-Candidate Risk Score model highlighted the need for an accurate and objective waitlist mortality model.[12]

Predictive models that integrate patient history with serology have been successfully implemented in other organ transplant systems. The Model for End-Stage Liver Disease, including sodium, or MELD-Na, score for liver transplantation accurately predicts 90-day patient waitlist mortality and is the most significant metric in liver allocation.[13] A similar model is desperately needed in cardiac transplantation to accurately identify and prioritize the most critically ill patients. A model that assesses risk beyond 1 year can also be utilized to prioritize the timing and selection of patients who are waitlisted with the intention of getting the most critically ill patients on the waitlist first. To this end, we used the Scientific Registry of Transplant Recipients (SRTR) database to develop and validate 4 predictive mortality models for waitlisted patients with end-stage HF. Our models utilize objective serological data to determine the most urgent heart transplant candidates and stratify their relative risk of waitlist mortality at 90 days (90D), 180 days (180D), 1 year (1Y), and 2 years (2Y).

METHODS

DATA SOURCES. This study utilized retrospective data from the SRTR. The SRTR system includes data on all donors, waitlisted candidates, and transplant recipients in the United States, submitted by the members of the OPTN.[14] The Health Resources and Services Administration, U.S. Department of Health and Human Services, provides oversight to the activities of the OPTN and SRTR contractors. The data reported here have been supplied by the Hennepin Healthcare Research Institute as the contractor for SRTR. The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy of or interpretation by the SRTR or the U.S. Government.

STUDY POPULATION. The study population included adult waitlisted cardiac transplant patients with some laboratory data who registered for a single-organ heart transplant (N = 4,176) between January 1, 2017, and September 2, 2023. There were no multicenter listed patients. All study participants were retrospectively censored at 2 time points, first on January 1, 2017, and then on September 2, 2023. Patients who did not reach follow-up time were included in the model. Supplemental Figure 1 provides a participant workflow diagram with the number of participants retained for each exclusion criterion. January 1, 2017, was selected as the start date of this cohort because SRTR stopped recording

laboratory values after this time in 2018. Thus, we imputed directly on the SRTR's serology data. Because many laboratory values were missing, we utilized the Multivariate Imputation by Chain Variables algorithm[15] to impute missing laboratory values. Training and testing sets were imputed independently. Specifically, albumin, aspartate aminotransferase, bilirubin, standard B-type natriuretic peptide (BNP), and sodium were missing in approximately 80% of participants. All steps in this analysis were conducted for 4 models. Survival times for waitlisted candidates started at the date of listing and were censored at the date of death or removal from the waitlist. Patients removed from the waitlist for death or because they were "too ill" were censored as an event. Candidates who were removed due to an improvement in their health were censored as alive. All patients were censored at the date of the last administrative follow-up, September 2, 2023. This population was randomly split (50% by 50%) into 2 cohorts: a training set (N = 2,088) and a test set (N = 2,088). Table 1 provides the population characteristics of the 2 study groups. The discovery set was used for variable selection, the generation of the logistic regression equations, and the creation of a tiered-risk system. The testing set was used to independently evaluate each model's performance in predicting mortality outcomes and ranking patients' relative risk of death. All analyses presented herein were applied to the training and test sets independently.

STATISTICAL APPROACH. Clinical variable selection and importance. For clinical applicability, we selected 13 independent patient variables (6 serology and 7 medical histories) that were readily available and clinically justified. For continuous variables, a two-way analysis of variance test was used to test the observed differences in patient characteristics and the independent variables used in this study. The chi-squared test was used to measure the significance of categorical and indicator (binary) variables. These patient predictor variables should also exhibit discriminating power and low collinearity. To measure collinearity, we calculated Pearson's cross-correlation coefficient (R) and variance inflation factor (VIF) for all independent variables. Low collinearity was defined as 2 independent variables with an R value <0.7. The VIF is used to determine the correlation between independent variables in a logistic regression model. A VIF of 1 provides no correlation, whereas values above 2.5 indicate considerable multicollinearity.[16]

Independent variable definitions. We leveraged a combination of serological predictors along with

TABLE 1 Patient Characteristics and Laboratory Values for Waitlisted Heart Transplant Patients, 2017 to 2023

	Training Data (n = 2,088)	Test Data (n = 2,088)	P Value
Age			
Mean (SD)	53.3 ± 12.6	53.1 ± 12.8	0.904
Median [Min, Max]	56.0 [18.0, 76.0]	56.0 [18.0, 73.0]	
Sex			
Female	519 (24.9%)	505 (24.2%)	0.881
Male	1,569 (75.1%)	1,583 (75.8%)	
Race			
Caucasian	1,294 (62.0%)	1,334 (63.9%)	0.994
Hispanic/Latino	184 (8.8%)	169 (8.1%)	
Black or African American	509 (24.4%)	493 (23.6%)	
Asian	76 (3.6%)	66 (3.2%)	
American Indian or Alaska Native	7 (0.3%)	8 (0.4%)	
Native Hawaiian or Pacific Islander	8 (0.4%)	8 (0.4%)	
Ethnicity			
Latino	187 (9.0%)	172 (8.2%)	0.71
Non-Latino or unknown	1901 (91.0%)	1916 (91.8%)	
Education			
High School (9-12)	761 (36.4%)	808 (38.7%)	0.941
Attended College/Technical School	568 (27.2%)	569 (27.3%)	
Associate/Bachelor's degree	440 (21.1%)	408 (19.5%)	
Post-college Graduate Degree	187 (9.0%)	176 (8.4%)	
Grade School (0-8)	75 (3.6%)	67 (3.2%)	
None	50 (2.4%)	59 (2.8%)	

patient medical history variables to construct the CHARM (Colorado Heart failure Acuity Risk Model) score. Serum values include BNP, serum creatinine (Cr), sodium, aspartate aminotransferase, albumin, and total bilirubin. The BNP marker used in these models directly tests the BNP hormone, not the surrogate protein marker known as NT-proBNP. BNP is a strong predictor of heart function.[17-20] Other variables are related to the patient's clinical course with HF, including implantable cardiac defibrillators, mechanical circulatory support, inotropic support, and ventilator support. A history of cardiac surgery and a previous transplant were also predictive of waitlist mortality. The life support variable refers to any waitlisted heart transplant recipient who received inotropic infusion right ventricular assist devices, left ventricular assist devices, including Impella devices prior to 5.5, total artificial hearts, or intra-aortic balloon pumps.[21] Extracorporeal membrane oxygenation was defined as its own specific variable outside of the term "life support." The ventilator indicator variable refers to the need for any mechanical ventilation prior to transplant.

Regression formulas for calculating patient risk scores. We set out to build a set of 4 predictive models to formally define a wide range of patient risk by developing a score known as the CHARM score,

TABLE 1 Continued			
	Training Data (n = 2,088)	Test Data (n = 2,088)	P Value
Independent variables			
Albumin (g/dL)			
Mean (SD)	3.73 ± 0.570	3.74 ± 0.600	0.686
Median [Min, Max]	3.80 [0.7, 6.50]	3.80 [0.5, 7.30]	
Bilirubin (mg/dL)			
Mean (SD)	0.852 ± 1.26	0.887 ± 1.53	0.728
Median [Min, Max]	0.70 [0.1, 40.8]	0.70 [0.1, 42.0]	
BNP (pg/mL)			
Mean (SD)	1,680 ± 2,260	1,750 ± 2,320	0.607
Median [Min, Max]	797 [5.00, 10,000]	763 [8.00, 10,000]	
Cardiac surgery			
Yes	61 (2.9%)	73 (3.5%)	0.932
No	2027 (97.1%)	2015 (96.5%)	
Creatinine (mg/dL)			
Mean (SD)	1.40 ± 1.04	1.41 ± 0.902	0.994
Median [Min, Max]	1.20 [0.08, 24.0]	1.20 [0.4, 10.7]	
Previous transplant			
Yes	61 (2.9%)	73 (3.5%)	0.574
No	2027 (97.1%)	2015 (96.5%)	
SGOT (U/L)			
Mean (SD)	88.1 ± 70.2	88.9 ± 70.6	0.225
Median [Min, Max]	28.0 [0.1, 34,300]	27.0 [0.1, 1810]	
Sodium (mEq/L)			
Mean (SD)	136 ± 4.26	136 ± 4.23	0.959
Median [Min, Max]	136 [119, 153]	137 [109, 153]	
Ventilator			
Yes	41 (2.0%)	47 (2.3%)	0.811
No	2047 (98.0%)	2041 (97.7%)	
ECMO			
Yes	48 (2.3%)	67 (3.2%)	0.199
No	2040 (97.7%)	2021 (96.8%)	
Implant defibrillator			
Yes	1,550 (74.2%)	1,555 (74.5%)	0.984
No	538 (25.8%)	533 (25.5%)	
Accept HCV-positive			
Yes	801 (38.4%)	770 (36.9%)	0.612
No	1,287 (61.6%)	1,318 (63.1%)	
Life support			
Yes	947 (45.4%)	986 (47.2%)	0.48
No	1,141 (54.6%)	1,102 (52.8%)	

Values are n (%) unless otherwise indicated.
BNP = B-type natriuretic peptide; ECMO = extracorporeal membrane oxygenation; HCV = hepatitis C virus; SGOT = glutamic-oxaloacetic transaminase.

exhibited linearity, normality in the distribution of the residuals, and low variable inflation.

Model calibration and the tiered-risk index. A well-calibrated ranking system is required to accurately estimate the relative medical urgency of waitlisted heart transplant patients. This ranking system can be used to prioritize patients for transplant and for entry onto the waitlist. To maximize our ability to statistically discriminate among high- and low-risk patient subgroups, we developed a 6-tiered risk index system based on the CHARM score, which ranges from 1 to 6, where 6 is the highest risk tier. To determine the goodness-of-fit of the CHARM score as compared to the observed patient risk, we performed the Hosmer-Lemeshow test[23] on each model using 6 equally distributed patient groups. All study participants were divided into sextiles, or 6 equally distrusted patient groups. We used the Hosmer-Lemeshow test to define the boundaries of risk tiers and calibrated each regression model by maximizing the goodness-of-fit between the observed and predicted patient risk. To more accurately assess the calibration of our models, we utilized R's "CalibrationCurves" package.[24] This software package provided the calibration intercept, slope, and discrimination (c-statistic) for each model.

Validating dichotomous outcomes using logistic regression. We measured each model's performance in predicting mortality events at 4 censoring periods using sample hold-out validation; we calculated coefficients using the training set (N = 2,088) and measured the area under the curve (AUC) and other performance metrics using the independent validation set (N = 2,088). To avoid confirmation bias in selecting the training and test sets, we performed 30 random splits and recorded the AUC of each test. R's "pROC" package[25] was used to calculate the 95% CIs of each model. All patient risk scores were calculated using Equations 1-4. Coefficients were calculated using the discovery set, and predictions were independently validated using the validation set. Finally, we calculated the AUC, for all 4 models.

Validating the 6-tiered risk index system. To validate the CHARM and our tiered-risk classification system using a univariable test, we performed the Cox proportional hazard regression method[26] to measure each tier's performance as a single survival predictor. We reported the concordance index[27] to gauge each model's overall performance with respect to survival time. The survival concordance is the fraction of randomly selected patient pairs among all pairs, where the higher-risk tier corresponds to the individual with an earlier event. The logarithm of the HRs, or log (HR), CIs, and P values were provided for

which is based on a patient's estimated likelihood of death while waitlisted. Thus, we constructed 4 logistic regression models using a generalized linear model with a logit link function at 90D, 180D, 1Y, and 2Y censoring times (Equations 1-4). All laboratory values were transformed to the logarithmic scale prior to calculation. The major model assumptions were checked for all 4 models using R's "performance" package.[22] Supplemental Figure 2 provided the quality checks for the 90-day model. All 4 models

CENTRAL ILLUSTRATION CHARM (Colorado Heart Failure Acuity Risk Model): A Mortality Model for Waitlisted Cardiac Transplant Patients
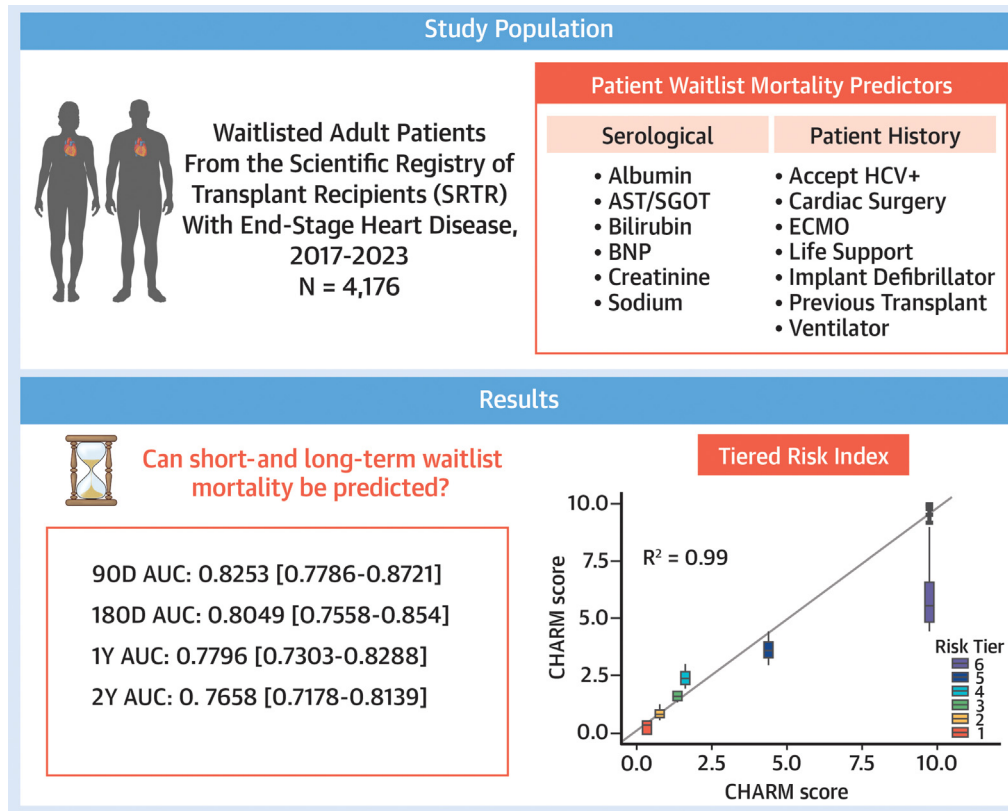
Murphy RD, et al. JACC Adv. 2025;4(1):101449.

Cardiac surgery refers to underwent coronary artery bypass grafting, valve replacement or repair, surgery to repair a congenital disease, and left ventricular remodeling. AST = aspartate aminotransferase; AUC = area under the curve; BNP = B-type natriuretic peptide; ECMO = extracorporeal membrane oxygenation; HCV = hepatitis C virus; SGOT = glutamic-oxaloacetic transaminase.

all tiers and models. This allowed for one simple linear metric for comparing all patient tiers. All analyses were performed using the R statistical language version 4.4.3.[28] TRIPOD, or the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis, was followed throughout this study.[29]

RESULTS

POPULATION CHARACTERISTICS. The training (discovery) and test (validation) patient population characteristics are depicted in **Table 1**. The mean age of the 4,176 study participants ranged from 18 to 76 years, with a mean age of 53.2 ± 12.7 years. Of participants, 75.5% were male, 62.9% were White or Caucasian, 24% were Black or African American, 8.5% were Hispanic or Latino, 3.4% were Asian, 0.4% were

Native Hawaiian or Other Pacific Islander, and 6 (0.3%) were American Indian or Alaska Native. In summary, there were no statistically significant differences in the patient characteristics or independent variables when comparing the discovery set to the validation set.

VARIABLE COEFFICIENTS AND INFLATION FACTORS. A graphic abstract of the CHARM model and study design are provided in the **Central Illustration**. All correlation coefficients of the independent variables were below the absolute value of 0.35 (**Figure 1**). All logistic regression coefficients, P values, and VIF values are provided in Supplemental Table 1. VIF values ranged from 1 to 1.5. No significant correlation or inflation was observed among the independent variables. Also, laboratory values that were imputed provided very small coefficients. Thus, they cannot inflate or skew the model's results.

FIGURE 1 Correlation Heatmap of Independent Variables
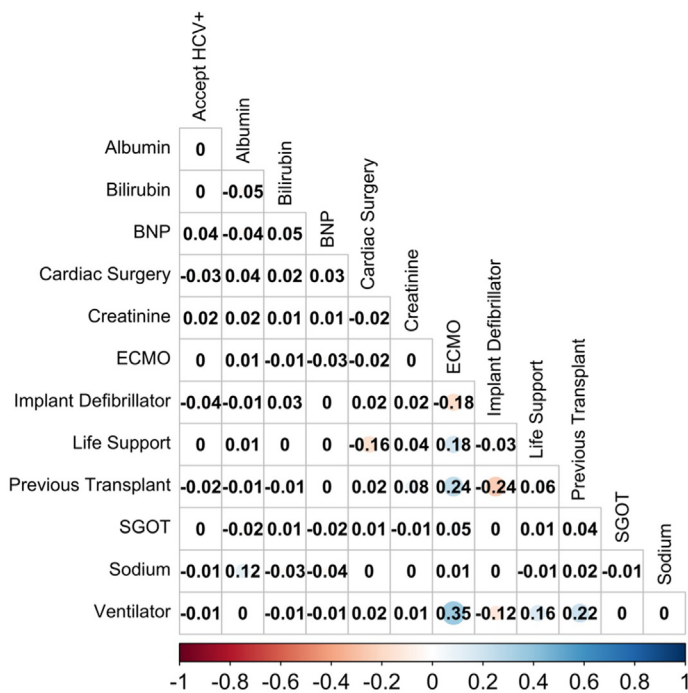
Pearson's correlation coefficients were calculated using all patients (N = 4,176) for the 13 independent variables used to construct the 4 models (90 days, 180 days, 1 year, and 2 years). This is used to provide a measure of collinearity. Blue indicates a positive correlation, and red indicates a negative correlation. The color saturation and circle area increase as the correlation coefficients increase in magnitude. No highly significant correlations were observed. BNP = B-type natriuretic peptide; ECMO = extracorporeal membrane oxygenation; SGOT = glutamic-oxaloacetic transaminase.

LOGISTIC REGRESSION FORMULAS. Equation 1: The 90-day CHARM score.

$$90D\ CHARM\ Score = -2.46 + 0.02(Albumin) \\ + 0.03(Bilirubin) + 0.01(BNP) \\ + 0.21(Creatinine) + 0.01(AST) \\ - 0.02(SODIUM) + 0.15(CARDIAC\_SURGERY) \\ - 0.01(IMPLANT\_DEFIB) + 1.12(LIFE\_SUPPORT) \\ - 0.06(VENTILATOR) + 0.65(PREV\_TX) \\ - 0.53(ACPT\_HCV\_POS) + 1.65(ECMO)$$

Equation 2: The 180-day CHARM score.

$$180D\ CHARM\ Score = -1.07 - 0.04(Albumin) \\ + 0.02(Bilirubin) + 0.01(BNP) + 0.25(Creatinine) \\ + 0.01(AST) - 0.02(SODIUM) \\ + 0.19(CARDIAC\_SURGERY) \\ - 0.1(IMPLANT\_DEFIB) + 0.93(LIFE\_SUPPORT) \\ - 0.26(VENTILATOR) + 0.65(PREV\_TX) \\ - 0.44(ACPT\_HCV\_POS) + 1.66(ECMO)$$

Equation 3: The 1-year CHARM score.

$$1Y\ CHARM\ Score = -2.3 + 0.07(Albumin) \\ + 0.01(Bilirubin) + 0.01(BNP) + 0.22(Creatinine) \\ + 0.01(AST) - 0.01(SODIUM) \\ + 0.2(CARDIAC\_SURGERY) \\ - 0.01(IMPLANT\_DEFIB) + 0.75(LIFE\_SUPPORT) \\ - 0.34(VENTILATOR) + 0.88(PREV\_TX) \\ - 0.56(ACPT\_HCV\_POS) + 1.56(ECMO)$$

Equation 4: The 2-year CHARM score..

$$2Y\ CHARM\ Score = -1.97 + 0.1(Albumin) \\ + 0.01(Bilirubin) + 0.01(BNP) + 0.2(Creatinine) \\ + 0.01(AST) - 0.02(SODIUM) \\ + 0.18(CARDIAC\_SURGERY) \\ + 0.02(IMPLANT\_DEFIB) + 0.63(LIFE\_SUPPORT) \\ - 0.39(VENTILATOR) + 0.89(PREV\_TX) \\ - 0.47(ACPT\_HCV\_POS) + 1.56(ECMO)$$

MODEL CALIBRATION AND THE TIERED-RISK INDEX. The predicted mortality rates are presented as a function of the observed mortality rates in **Figure 2** for each risk tier. For the 90-day model, Risk Index (RI) 1 had an observed mortality rate of 0.43%, RI 2 had a mortality rate of 0.08%, RI 3 had a mortality rate of 1.44%, RI 4 had a mortality rate of 1.72%, RI 5 had a mortality rate of 4.45%, and RI 6 had a mortality rate of 9.77%. Using the Hosmer-Lemeshow method, the goodness-of-fit observed between the observed and predicted patient risk was greater than 0.99 for all 4 models (Supplemental Figure 3). Interestingly, observed and predicted patient risk was exponentially distributed, with few waitlisted patients having a relatively higher risk of experiencing waitlist mortality. For example, in the 90D model, approximately two-thirds of patients have less than a 2.5% chance of mortality. All 4 models produced a sensitivity above 99%. Supplemental Table 2 provides the contingency tables (observed vs predicted outcomes) for all 4 models. Using "CalibrationCurves," we obtained calibration curve slopes (mean/SD) of 1.02 (0.80-1.24), 1.00 (0.80-1.21), 1.00 (0.79-1.21), and 1.03 (0.81-1.25) for the 90D, 180D, 1Y, and 2Y models, respectively. The intercepts (mean/SD) were −0.09 (−0.36 to 0.17), −0.02 (−0.26 to 0.22), −0.01 (−0.23 to 0.22), −0.05 (−0.27 to 0.17), respectively. The c-statistics (mean/SD) were 0.81 (0.75-0.86), 0.79 (0.73-0.84), 0.77 (0.72-0.82), and 0.76 (0.71-0.81), respectively.

LOGISTIC REGRESSION FOR PREDICTING SHORT-TERM MORTALITY OUTCOMES. Using sample hold-out validation, the calculated AUCs (mean/SD) were 0.825 (0.7786-0.8721) (90-day), 0.805 (0.7558-0.854)
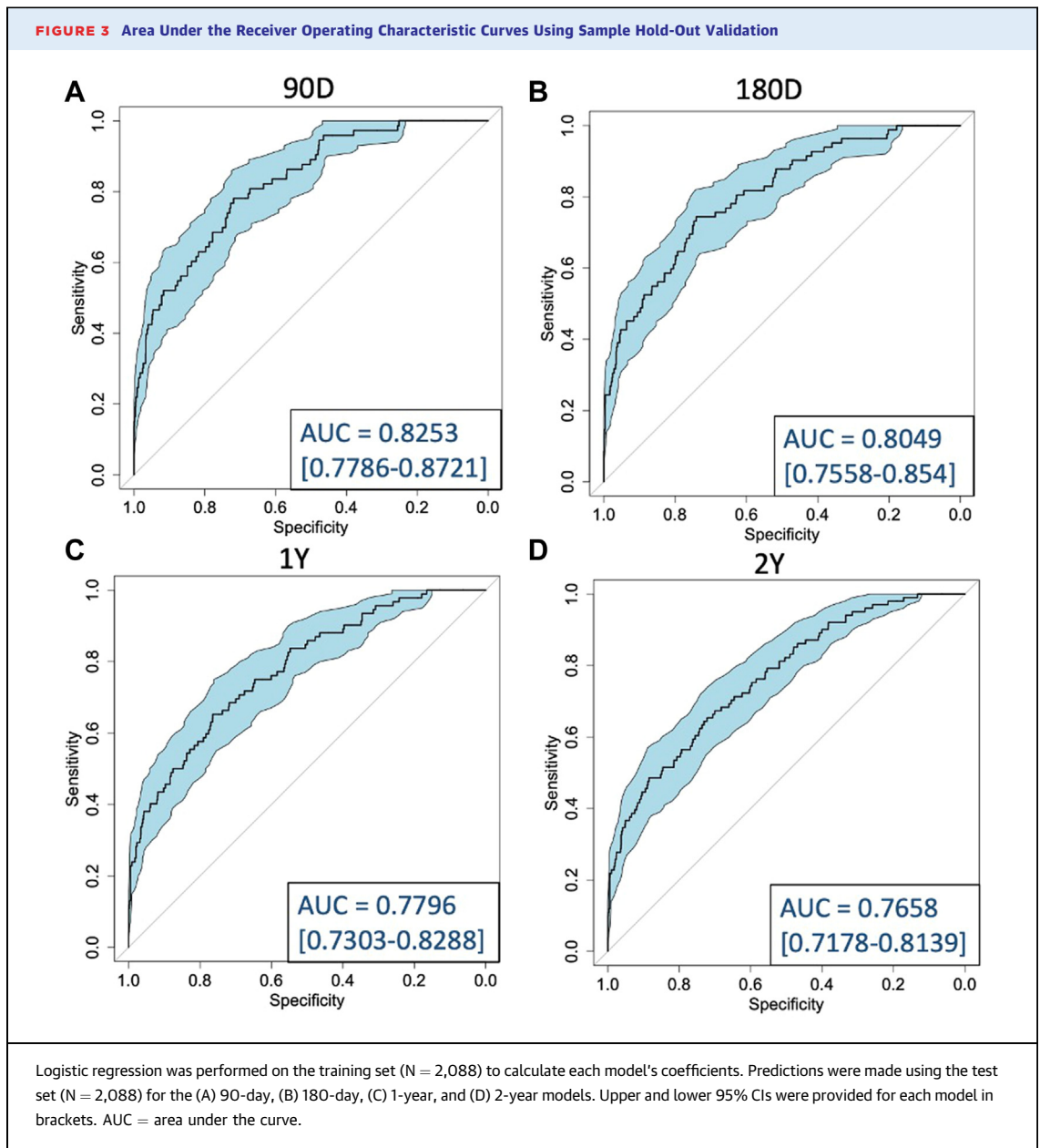
FIGURE 2 Tiered-Risk Index System
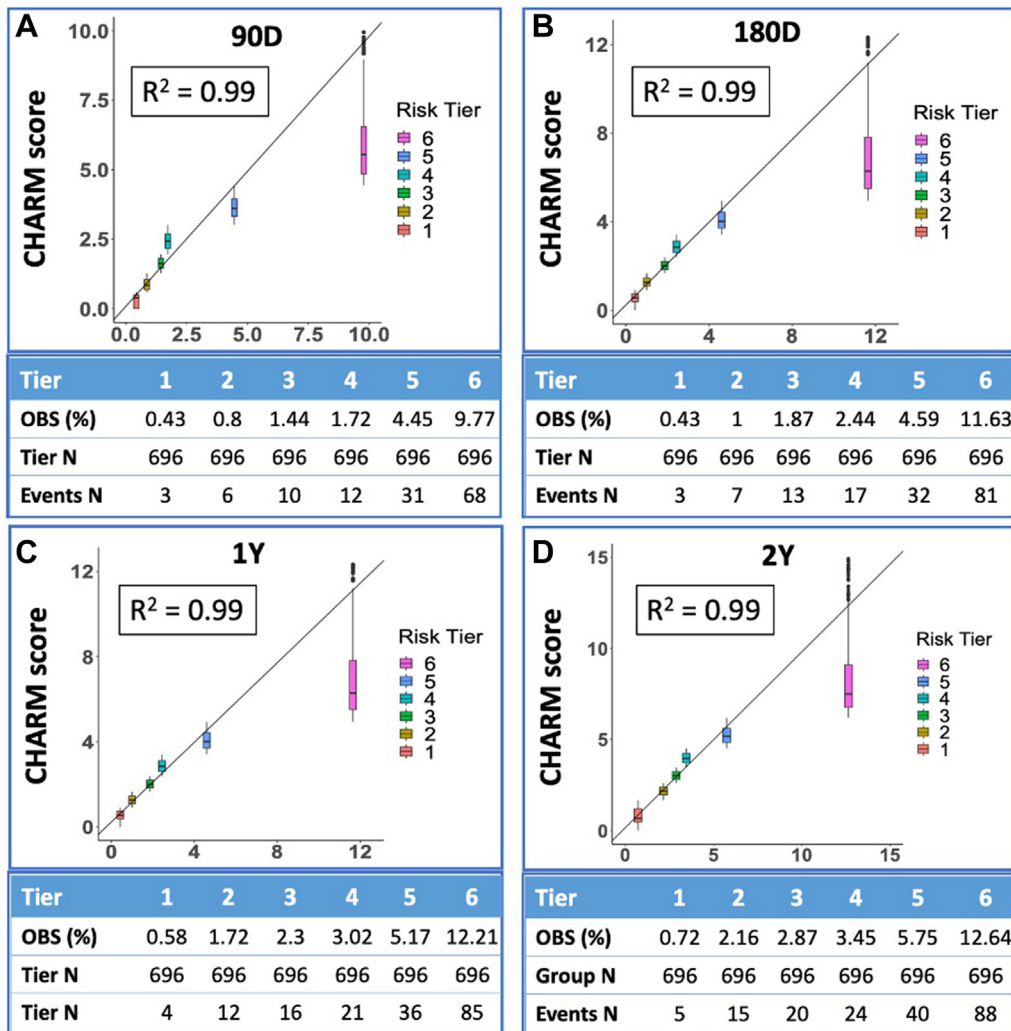
Patient mortality probabilities were calculated for all patients in the cohort (N = 4,176) and are provided as a function of observed patient mortality rate per tier for the (A) 90-day, (B) 180-day, (C) 1-year, and (D) 2-year models. The goodness-of-fit was calculated with 6 tiers using the Hosmer-Lemeshow methodology and was greater than 0.99 for all 4 models. The observed morality rate (OBS), number of patients per group (Tier N), and the number of waitlist patient deaths (Events N) are reported for each tier and model. CHARM = Colorado Heart failure Acuity Risk Model.

(180-day), 0.779 (0.730-0.829) (1-year), and 0.766 (0.718-0.814) (2-year), respectively (Figure 3). All 30 random validation sets based on selection splits fell within the 95% CIs for all 4 models. In summary, these models and the tiered-risk system provide a reliable and highly accurate methodology for ranking the short-term survival of waitlisted heart transplant patients.

TIME-DEPENDENT VALIDATION OF THE CHARM AND 6-TIERED RISK INDEX SYSTEM. We applied the Cox proportional hazard model as a multivariable and univariable validation test. In the multivariable model, all 13 original independent variables were used. In the univariable model, the independent variable was the risk index, or tier. We found a significant difference in survival times by risk tier. For the multivariable model, the survival concordance values were 78.8% (90D), 77.2% (180D), 75.3% (1Y), and 73.7% (2Y). For the univariable model, the survival concordance values were 72.6% (90D), 76.1% (180D), 73.7% (1Y), and 72.1% (2Y). The standard error was below 0.021 for all 4 concordance values.

FIGURE 3 Area Under the Receiver Operating Characteristic Curves Using Sample Hold-Out Validation

Logistic regression was performed on the training set (N = 2,088) to calculate each model's coefficients. Predictions were made using the test set (N = 2,088) for the (A) 90-day, (B) 180-day, (C) 1-year, and (D) 2-year models. Upper and lower 95% CIs were provided for each model in brackets. AUC = area under the curve.
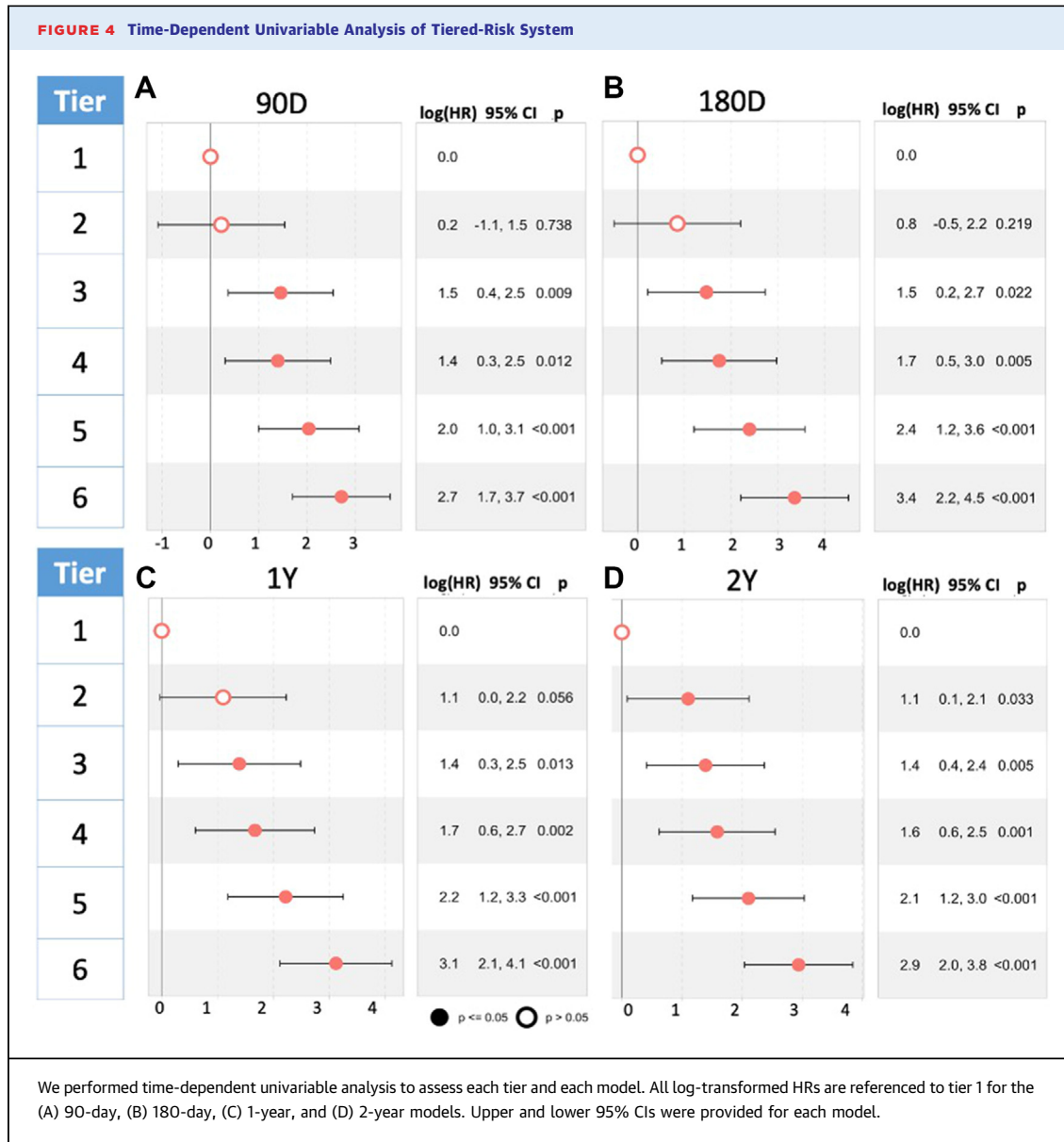
Figure 4 provides the log (HR) values for each tier and model. The HR was transformed to the logarithmic scale so that the risk tiers could be compared in linear space. For example, in the 90D model, Tier 5 was twice as likely, on average, to die on the waitlist as Tier 1. Accordingly, Tier 5 had a 50% increase in the likelihood of death as compared to Tier 3.

DISCUSSION

In 2023, it was demonstrated that the OPTN's 6-status allocation system had only a moderate ability

(AUC = 0.67) to identify the short-term survival likelihood of waitlisted heart transplant candidates.[6] While the current system accounts for many patient factors, an objective model that can accurately stratify patients by medical urgency is needed to inform the new continuous distribution system. Formal pretransplant patient mortality models have been successfully developed and utilized in liver and kidney transplantation. Evans et al[30] demonstrated an overall 1-year survival rate increase of 18% in high-acuity patients in the 15 years following the national implementation of the MELD-Na score. A similar

FIGURE 4 Time-Dependent Univariable Analysis of Tiered-Risk System

We performed time-dependent univariable analysis to assess each tier and each model. All log-transformed HRs are referenced to tier 1 for the (A) 90-day, (B) 180-day, (C) 1-year, and (D) 2-year models. Upper and lower 95% CIs were provided for each model.

metric is needed for cardiac transplantation to identify and prioritize the most critically ill waitlisted patients.

To this end, we created the CHARM score. This score used a 6-tiered categorical risk model that can be designated to patients on the waiting list and can prioritize patients for waitlist selection. Our models precisely rank patient subgroups based on waitlist mortality, with holdout-validation yielding average AUCs of 0.825 (90-day), 0.805 (180-day), 0.779 (1-year), and 0.766 (2-year).

The time-dependent concordance values were 78.8% (90D), 77.2% (180D), 75.3% (1Y), and 73.7% (2Y). This prediction accuracy is the highest of any existing HF survival or pretransplant mortality score, including the U.S.-CRS (mean AUC 0.825 vs 0.78 at 90 days). Risk indices for all models were 99% correlated with the observed mortality rates.

The CHARM score is intended to be used to prioritize cardiac transplant candidates at the time of listing, with the recommendation to renew data points and recalculate every 90 days. The 1Y and 2Y

models are intended to better inform the prioritization and timing of the most critically ill patients to be included on the waitlist. Other models, such as the U.S.-CRS, utilized shorter censoring periods in their predictions (6 weeks), which did not include the majority of patient risk and could lead to a misrepresentation of the cummulative mortality rate, as there were relatively fewer patient deaths observed within the 6-week time period. The 90D time frame will allow for frequent reevaluation while maintaining a high degree of prediction accuracy. In addition, this timeframe better falls within the median waitlist time of 3 to 6 months reported by SRTR in 2019. We also included additional time lengths of 180D, 1Y, and 2Y models to demonstrate the cumulative patient risk of death on the waitlist, which has not been included in any other risk calculator.

We anticipate the CHARM score could be used to inform the heart continuous distribution system, as it utilizes a framework that is point-based rather than status-based, in which candidates are prioritized for transplant through the designation of a composite score from a variety of attributes. Staged implementation of the continuous distribution system is currently in use for other organs, with anticipated completion of the heart allocation system within the next few years. The CHARM score provides a simple, accurate measure of pretransplant mortality that can easily be incorporated into a heart transplant composite score.

The 13 independent patient variables incorporated into the CHARM score were chosen for their objectivity, clinical availability, and relevance to cardiac illness. Each variable was also determined to significantly contribute to the predictive value of waitlist mortality. Laboratory values were selected for their evaluation of crucial organ function in the setting of severe HF. A large meta-analysis of 64 models that predicted death or hospitalization from HF determined renal function is one of the most significant factors in these outcomes.[31] Renal function was included through serum sodium and creatinine. Multiple studies have demonstrated worsened short-term mortality for HF patients related to low serum sodium.[32,33] While estimated glomerular filtration rate, was considered a measure of renal function, it is not a value directly recorded in the SRTR database and can be calculated differently by institutions. BNP serves as an objective marker of cardiac stretch since it is influenced by the level of end-diastolic volume, and it can also be an indicator of responsiveness to

diuretic management.[34] In addition to physiologic data, indicator variables such as previous cardiac surgery, previous cardiac transplant, ventilator support, and life support proved to be large contributors to pretransplant mortality. These interventions often serve as a "bridge" to transplant, reserved for the most critically ill patients. Previous cardiac surgeries or transplants are also indicative of a more extensive history of cardiac illness.

One unique patient variable added to the CHARM score is that of accepting a hepatitis C virus (HCV) positive donor heart. This element was added for multiple reasons. Patients who elect to receive a HCV+ heart spend less time on the waitlist than those who reject an HCV+ donor. Thus, HCV acceptance acts as a surrogate variable for waitlist time. We specifically chose not to include variables that depend upon a physician's practice or measurement, such as right heart catheterization data or the level of inotropic support. These variables were excluded with the intent of reducing bias based on treatment variation. We also intentionally excluded age and time on the waitlist to reduce bias.

The next steps for further utilization of the CHARM score will include simulation modeling and a prospective, multicenter validation study in which these additional variables are collected and analyzed. We will also examine time-dependent and ensemble classifiers. To move the CHARM model into the national spotlight, we will continue to prospectively study waitlist mortality and work with all stakeholders to develop models using complete national patient databases. We strongly advocate for increasing data reporting requirements among waitlist candidates at the national level. The frequency with which data points are updated, along with the lack of granularity within reported patient variables, has posed a challenge for the creation of CHARM and other similar risk models that utilized the SRTR database. Nonetheless, this work proves that we can still provide accurate and reliable medical acuity metrics.

STUDY LIMITATIONS. Though we present accurate predictive models, they must be viewed within the confines of the study limitations, including available data within the national database. The variables chosen for the model were limited by the type of data recorded within SRTR. As such, certain serum markers were not readily available within the database and therefore could not be accurately used. Because the coefficients of the serology predictors are

relatively low, normally distributed, and not inflated, we expect minimal bias. Further studies will be needed to aggregate sufficient serology values. These efforts will greatly improve future models. Similarly to the US-CRS, we relied on imputation for serological predictors. Because the imputed laboratory variables are distributed normally and their coefficient values are very low, they do not inflate or bias the CHARM score. Furthermore, the specificity of some variables was limited. For example, the term "life support" is broad and includes anyone on inotropic support alone or varying degrees of temporary or durable mechanical support. Therefore, "life support" currently serves as a binary variable within the model, understanding that different modalities of mechanical support may contribute differently to patient risk. For the 1-year and 2-year models, we acknowledge that a subset of patients may not have accurate follow-up. Thus, we have validated the CHARM model using the time-dependent Cox proportional hazard survival regression analysis to derive a C-statistic that is more robust in the event of informative censoring.[35-37] In terms of statistical discrimination, logistic regression is often inferior to the random forest method, gradient boosting, and other ensemble methodologies.

ADDRESS FOR CORRESPONDENCE: Dr John Stephen Malamon, University of Colorado, Anschutz Medical Campus, 1635 Aurora Court, Aurora, Colorado 80045, USA. E-mail: john.malamon@cuanschutz.edu.

PERSPECTIVES

COMPETENCY IN SYSTEMS-BASED PRACTICE: The Colorado Heart failure Acuity Risk Model (CHARM) score provides a validated model with strong predictive ability for short-term mortality among patients waitlisted for cardiac transplantation. The CHARM score will be useful for prioritizing the selection of waitlisted patients.

TRANSLATIONAL OUTLOOK: We anticipate the CHARM score will be useful in the era of continuous distribution to standardize organ allocation by providing an objective and intuitive system for stratifying waitlisted heart failure patients based on medical urgency.

REFERENCES

1. Stevenson LW. Crisis awaiting heart transplantation: sinking the lifeboat. *JAMA Intern Med.* 2015;175(8):1406-1409.

2. Maitra NS, Dugger SJ, Balachandran IC, Civitello AB, Khazanie P, Rogers JG. Impact of the 2018 UNOS heart transplant policy changes on patient outcomes. *JACC Heart Fail.* 2023;11(5): 491-503.

3. Cogswell R, John R, Estep JD, et al. An early investigation of outcomes with the new 2018 donor heart allocation system in the United States. *J Heart Lung Transplant.* 2020;39(1):1-4.

4. Kilic A, Hickey G, Mathier MA, et al. Outcomes of the first 1300 adult heart transplants in the United States after the allocation policy change. *Circulation.* 2020;141(20):1662-1664.

5. Lazenby KA, Narang N, Pelzer KM, Ran G, Parker WF. An updated estimate of posttransplant survival after implementation of the new donor heart allocation policy. *Am J Transplant.* 2022;22(6):1683-1690.

6. Pelzer KM, Zhang KC, Lazenby KA, et al. The accuracy of initial U.S. Heart transplant candidate rankings. *JACC Heart Fail.* 2023;11(5):504-512.

7. Lin EY, Cohen HW, Bhatt AB, et al. Predicting outcomes using the heart failure survival score in adults with moderate or complex congenital heart disease. *Congenit Heart Dis.* 2015;10(5): 387-395.

8. Levy WC, Mozaffarian D, Linker DT, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation.* 2006;113(11):1424-1433.

9. Goda A, Williams P, Mancini D, Lund LH. Selecting patients for heart transplantation: comparison of the Heart Failure Survival Score (HFSS) and the Seattle heart failure model (SHFM). *J Heart Lung Transplant.* 2011;30(11): 1236-1243.

10. Lustig DB, Rodriguez R, Wells PS. Implementation and validation of a risk stratification method at The Ottawa Hospital to guide thromboprophylaxis in ambulatory cancer patients at intermediate-high risk for venous thrombosis. *Thromb Res.* 2015;136(6):1099-1102.

11. Pocock SJ, Ariti CA, McMurray JJ, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *Eur Heart J.* 2013;34(19):1404-1413.

12. Zhang KC, Narang N, Jasseron C, et al. Development and validation of a risk score predicting death without transplant in adult heart transplant candidates. *JAMA.* 2024;331(6):500-509.

13. Austin MT, Poulose BK, Ray WA, Arbogast PG, Feurer ID, Pinson CW. Model for end-stage liver disease: did the new liver allocation policy affect waiting list mortality? *Arch Surg.* 2007;142(11): 1079-1085.

14. OPTN. A guide to calculating and interpreting the estimated post-transplant survival (EPTS) score used in the kidney allocation system (KAS). 2020. Accessed August 5, 2022. https://optn.transplant.hrsa.gov/media/1511/guide_to_calculating_interpreting_epts.pdf

15. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Software.* 2011;45(3):1-67.

16. Johnston R, Jones K, Manley D. Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Qual Quant.* 2018;52(4):1957-1976.

17. Cowie MR, Mendez GF. BNP and congestive heart failure. *Prog Cardiovasc Dis.* 2002;44(4): 293-321.

18. Kuwahara K. The natriuretic peptide system in heart failure: diagnostic and therapeutic implications. *Pharmacol Ther.* 2021;227:107863.

19. Moe GW. B-type natriuretic peptide in heart failure. *Curr Opin Cardiol.* 2006;21(3):208-214.

20. Vuolteenaho O, Ala-Kopsala M, Ruskoaho H. BNP as a biomarker in heart disease. *Adv Clin Chem.* 2005;40:1-36.

21. Colvin M, Smith JM, Ahn Y, et al. OPTN/SRTR 2020 annual data report: heart. *Am J Transplant.* 2022;22(Suppl 2):350-437.

22. Lüdecke D. Performance: An R package for assessment, comparison and testing of statistical models. *J Open Source Softw*. 2021;6(3139).

23. Hosmer DW, Hosmer T, Lemeshow S. A goodness-of-fit tests for the multiple logistic regression model. *Commun Stat*. 1980;10:1043–1069.

24. Abidi MZ, Schold JD, Kaplan B, Weinberg A, Erlandson KM, Malamon JS. Patient years lost due to cytomegalovirus serostatus mismatching in the scientific registry of transplant recipients. *Front Immunol*. 2024;14:1292648.

25. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf*. 2011;12:77.

26. Cox DR. Regression models and life-tables. In: *Breakthroughs in statistics methodology and distribution*. New York: Springer; 1992:527–541.

27. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247(18):2543–2546.

28. *Team RC. R: A language and environment for statistical computing (Version 4.0. 2)*. R Foundation for Statistical Computing, Scientific Research Publishing: Wuhan, China; 2020.

29. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med*. 2015;13:1.

30. Evans MD, Diaz J, Adamusiak AM, et al. Predictors of survival after liver transplantation in patients with the highest acuity (MELD >/=40). *Ann Surg*. 2020;272(3):458–466.

31. Rahimi K, Bennett D, Conrad N, et al. Risk prediction in patients with heart failure: a systematic review and analysis. *JACC Heart Fail*. 2014;2(5):440–446.

32. Gheorghiade M, Abraham WT, Albert NM, et al. Relationship between admission serum sodium concentration and clinical outcomes in patients hospitalized for heart failure: an analysis from the OPTIMIZE-HF registry. *Eur Heart J*. 2007;28(8):980–988.

33. Klein L, O'Connor CM, Leimberger JD, et al. Lower serum sodium is associated with increased short-term mortality in hospitalized patients with worsening heart failure: results from the Outcomes of a Prospective Trial of Intravenous Milrinone for Exacerbations of Chronic Heart Failure (OPTIME-CHF) study. *Circulation*. 2005;111(19):2454–2460.

34. Mueller C, McDonald K, de Boer RA, et al. Heart Failure Association of the European Society of Cardiology practical guidance on the use of natriuretic peptide concentrations. *Eur J Heart Fail*. 2019;21(6):715–731.

35. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med*. 1999;18(17-18):2529–2545.

36. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000;56(2):337–344.

37. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med*. 2011;30(10):1105–1117.

KEY WORDS failure, heart, mortality, prediction, transplant, waitlist

---

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