Journal of the American Heart Association

ORIGINAL RESEARCH

Optimized Risk Score to Predict Mortality in Patients With Cardiogenic Shock in the Cardiac Intensive Care Unit

Eric Yamga, MD; Sreekar Mantena, AB; Darin Rosen, MD; Emily M. Bucholz, MD, PhD, MPH; Robert W. Yeh, MD, MSc; Leo A. Celi, MD; Berk Ustun, PhD; Neel M. Butala, MD, MBA

BACKGROUND: Mortality prediction in critically ill patients with cardiogenic shock can guide triage and selection of potentially high-risk treatment options.

METHODS AND RESULTS: We developed and externally validated a checklist risk score to predict in-hospital mortality among adults admitted to the cardiac intensive care unit with Society for Cardiovascular Angiography & Interventions Shock Stage C or greater cardiogenic shock using 2 real-world data sets and Risk-Calibrated Super-sparse Linear Integer Modeling (RiskSLIM). We compared this model to those developed using conventional penalized logistic regression and published cardiogenic shock and intensive care unit mortality prediction models. There were 8815 patients in our training cohort (in-hospital mortality 13.4%) and 2237 patients in our validation cohort (in-hospital mortality 22.8%), and there were 39 candidate predictor variables. The final risk score (termed BOS,MA_2) included maximum blood urea nitrogen ≥25 mg/dL, minimum oxygen saturation <88%, minimum systolic blood pressure <80 mm Hg, use of mechanical ventilation, age ≥60 years, and maximum anion gap ≥14 mmol/L, based on values recorded during the first 24 hours of intensive care unit stay. Predicted in-hospital mortality ranged from 0.5% for a score of 0 to 70.2% for a score of 6. The area under the receiver operating curve was 0.83 (0.82–0.84) in training and 0.76 (0.73–0.78) in validation, and the expected calibration error was 0.9% in training and 2.6% in validation.

CONCLUSIONS: Developed using a novel machine learning method and the largest cardiogenic shock cohorts among published models, BOS,MA_2 is a simple, clinically interpretable risk score that has improved performance compared with existing cardiogenic-shock risk scores and better calibration than general intensive care unit risk scores.

Key Words: cardiogenic shock ■ CICU ■ machine learning ■ mortality ■ risk score ■ SCAI shock

Mortality for cardiogenic shock remains high, with fewer than 70% of patients surviving to hospital discharge.[1,2] Randomized trials have largely been unsuccessful in identifying strategies to improve mortality for patients with cardiogenic shock,[3–5] aside from culprit-vessel revascularization for myocardial infarction (MI).[6] Nevertheless, a spectrum of treatment options exists, ranging from emergent mechanical circulatory support to palliation.[7] The variation in outcomes for patients with cardiogenic shock may partly stem from differences in illness severity.[1] In this setting, tools to stratify patients with cardiogenic shock can provide important prognostic information and guide the appropriate triage and selection of therapies.

The Society for Cardiovascular Angiography & Interventions (SCAI) Shock Stage classification system was developed to indicate cardiogenic shock severity[8] but represents only 1 component of cardiogenic shock mortality risk prediction and must be used in concert with tools to account for risk modifiers in cardiogenic

Correspondence to: Neel M. Butala, MD, MBA, Rocky Mountain Regional VA Medical Center, 1700 N Wheeling St, Aurora, CO 80045, Cardiology 111B, 720-723-6197, USA. Email: neel.butala@cuanschutz.edu

This article was sent to Hani Jneid, MD, Associate Editor, for review by expert referees, editorial decision, and final disposition.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.122.029232

For Sources of Funding and Disclosures, see page 11.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- Developed using 2 large real-world data sets (n=8815 for training, n=2237 for validation) and a machine learning method, the 6-component BOS,MA$_2$ (blood urea nitrogen ≥25 mg/dL, minimum oxygen saturation <88 %, minimum systolic blood pressure <80 mm Hg, use of mechanical ventilation, age ≥60 years, and maximum anion gap) checklist risk score has improved performance compared with existing models for predicting mortality in patients with cardiogenic shock in the cardiac intensive care unit.

What Are the Clinical Implications?

- The BOS,MA$_2$ risk score is a simple, clinically interpretable risk score that can guide clinical decision-making at the bedside for patients with cardiogenic shock in the cardiac intensive care unit.
- This score can be used to assess the impact of treatment strategies on expected mortality, can enable the design of future clinical trials with more homogenous populations, and can serve as a model for developing future risk scores in cardiology.

Nonstandard Abbreviations and Acronyms

CICU	cardiac intensive care unit
ECE	expected calibration error
PLR	penalized logistic regression
RiskSLIM	Risk-calibrated Super-sparse Linear Integer Model

shock evaluation and prognostication.[9] Generic models for predicting mortality in the intensive care unit (ICU) are ill suited to predicting mortality in the cardiogenic shock population,[10] as they were developed using large heterogeneous ICU patient populations.[1] Conversely, models to predict mortality specifically for patients with cardiogenic shock are derived from small observational or clinical trial data sets with strict exclusion criteria, which limit their accuracy and generalizability.[11–17]

To address this knowledge gap, we developed and validated a specialized risk score to predict mortality among patients with cardiogenic shock in the cardiac intensive care unit (CICU). We leveraged 2 large real-world databases and a machine learning method developed to fit simple additive risk scores.[18] Such a risk score can quickly be calculated at the bedside and inform shared decision-making regarding treatment options for critically ill patients.

METHODS

Data Sources

We developed and validated our model using independent data sets from 2 publicly available clinical data repositories. Our training data set was derived from the Philips electronic ICU (eICU) database (eICU-CRD v2.0), which is composed of 200 859 patient encounters for 139 367 unique patients admitted between 2014 and 2015 in 1 of 335 units in 208 hospitals located throughout the United States.[19] Our validation data set was derived from Medical Information Mart for Intensive Care III (MIMIC-III), which is composed of 61 532 adult hospital admissions for 53 423 distinct patients admitted to critical care units between 2001 and 2012 at Beth Israel Deaconess Medical Center.[20] Because of the sensitive nature of the data collected for this study, access to these data sets by qualified researchers trained in human subject confidentiality protocols may be obtained by following policies listed on PhysioNet (https://physionet.org/about/database/). Methods to replicate the findings of this study are available from the corresponding author upon reasonable request.

The use of the eICU data set for this study with waiver of informed consent was exempt from institutional review board approval, and the use of the MIMIC-III with waiver of informed consent was approved by the institutional review board of the Massachusetts Institute of Technology and Beth Israel Deaconess Medical Center.

All data extraction and analyses were conducted using the cloud platform Google BigQuery, Python version 3.7, and R version 4.0.3.

Inclusion Criteria

We included all adult patients (age ≥18 years) admitted to a CICU who met the criteria for SCAI Shock Stage C or greater upon admission.[8] The SCAI Shock staging system was designed partly to enable the inclusion of a more homogeneous set of patients for enrollment in clinical trials and retrospective studies.[21–24] We included patients with SCAI Shock C or higher as those are categories representing patients with overt manifestations signs of shock.[8] Specifically, to be classified as SCAI Shock C, patients were required to have evidence of hypoperfusion, defined as a doubling of creatinine within 24 hours, blood lactate level >2.0 mmol/L, or the use of vasoactive medications. Our approach was similar to the operationalization of the SCAI Shock classification scheme in other real-world data sets that

included all patients admitted to CICUs regardless of admission diagnosis.[21,24]

Variables

The outcome of interest was in-hospital mortality. We considered 39 distinct variables present in both data sets for potential inclusion in our risk model, including demographics (age, sex), an array of comorbidities, vital signs within the first 24 hours (heart rate, systolic blood pressure, respiratory rate, oxygen saturation), laboratory values within the first 24 hours of ICU admission (anion gap, bicarbonate, chloride, glucose, hematocrit/hemoglobin, platelets, potassium, international normalized ratio, sodium, creatinine, blood urea nitrogen, white blood cell count, and red cell distribution width), and use of critical care therapies (renal replacement therapy, mechanical ventilation, intra-aortic balloon pump, and vasopressors). All variables were directly extracted from the databases except for comorbidities, for which *International Classification of Diseases, Ninth Revision* (*ICD-9*) codes were mapped to comorbidities based on the Clinical Classifications Software categories.

As laboratory results and vital signs were recorded multiple times in the first 24 hours, we considered their minimum and maximum values as potential features in our model. We also evaluated various cutoffs for the dichotomization of quantitative variables as distinct features for our model.

All candidate variables had <25% missing values, and simple imputation using predictive mean matching was applied independently to the training and validation data sets. A complete list of all variables available in both data sets and their corresponding missingness is available (Table S1).

We compared the baseline characteristics of our patient population according to the outcome of interest using standardized differences, with a threshold of at least 10% used to define a meaningful difference.[25]

Statistical Analysis and Model Development

We built our risk score using Risk-Calibrated Supersparse Linear Integer Model (RiskSLIM),[18] a machine learning method designed to fit simple customized risk scores optimized to yield calibrated risk estimates with few terms and small integer coefficients that have been used in clinical applications.[26,27]

Risk-Calibrated Super-sparse Linear Integer Modeling

RiskSLIM uses modern optimization techniques to fit the best logistic regression model with small integer weights and a limited number of risk factors.[28,29] Compared with models developed heuristically, this technique can fit risk scores with better risk calibration and area under the receiver operating curve (AUC) by combining logistic regression with feature selection and continuous variable dichotomization techniques in a single step. In this application, the model was constrained to use unit weights to allow for quick computation at the bedside as a checklist.

Comparator Methods

We compared the results of the RiskSLIM models to those developed using conventional penalized logistic regression (PLR) to evaluate the potential loss in accuracy due to feature selection and the use of integer coefficients.

Training Procedure

We trained all models using the eICU data set because it includes patients from a more heterogenous population (ie, across 208 hospitals)[30] and used the MIMIC data set for validation. We evaluated the performance of each model internally using 5-fold cross-validation. Platt scaling was employed in the final model to improve the reliability of estimates.

Performance Evaluation

We evaluated all models by rank accuracy and risk calibration.[31] We assessed rank accuracy via AUC. We assessed risk calibration by constructing a reliability diagram plotting the observed mortality compared with the predicted mortality and by reporting the expected calibration error (ECE), which reflects how close the predicted mortality risk is to the actual mortality risk.

Comparison to Other Risk Scores

We compared the performance of our RiskSLIM model to that of other published risk scores in our cohorts by comparing the AUC and ECE. We calculated 2 generic ICU scores, the Sequential Organ Failure Assessment score and Oxford Acute Severity of Illness Score, and the cardiogenic shock-specific CardShock score among all patients with available data. All were computed using data from the first 24 hours of admission.

Supplemental Analysis

To assess the effect of the potential inclusion of a heterogeneous patient population in our cohort, we evaluated the final trained model on various subgroups of the validation set. The 4 subgroups were as follows: patients with a cardiovascular primary ICU admission diagnosis (cardiogenic shock, heart failure, angina,

and MI), patients without a cardiovascular primary ICU admission diagnosis, the presence of MI necessitating acute coronary revascularization (percutaneous coronary intervention or coronary artery bypass grafting during hospitalization), and the absence of MI necessitating acute coronary revascularization. Admission diagnoses in the MIMIC-III were assigned by clinicians and stored as free text.

Primary admission diagnosis was not used to define our study cohort, given that this was not included in prior operationalizations of the SCAI Shock classification scheme in real-world data sets. In addition, admission diagnoses in eICU and MIMIC are recorded inconsistently. Using them would limit our model's applicability due to the inconsistent capture and accuracy of admission diagnoses in those 2 databases.

Key reporting elements for machine learning analyses in clinical research are summarized in a standardized format in Table S2.[32]

RESULTS

Patient Characteristics

eICU Cohort

Of 29 626 patients admitted to a CICU in the eICU database, 8815 (29.8%) met our inclusion criteria for our final training cohort (Figure 1). The overall in-hospital mortality rate in the eICU cohort was 13.5% (Table 1).

Patients with in-hospital mortality were more likely to be older (68 versus 64 years) and to have atrial fibrillation (19% versus 14%), solid neoplasm (16% versus 12%), congestive heart failure (27% versus 22%), chronic kidney disease (24% versus 17%), chronic obstructive pulmonary disease (22% versus 16%), or metastatic cancer (3% versus 1%). Additionally, patients with in-hospital mortality were more likely to be mechanically ventilated (57% versus 28%).

On average, patients with in-hospital mortality had a higher maximum heart rate (130 bpm versus 113 bpm), lower minimum systolic blood pressure (71 mm Hg versus 88 mm Hg), higher maximum respiratory rate (36/min versus 32/min), and lower minimum oxygen saturation (69% versus 86%) despite the use of all available therapies over the first 24 hours of CICU stay. On average, patients with in-hospital mortality were also more likely to have hyperglycemia (average minimum glucose 129 versus 118), higher maximum anion gap (18 versus 13), lower minimum bicarbonate (19 versus 23), higher maximum potassium (5.0 versus 4.6), higher maximum international normalized ratio (2.1 versus 1.5), higher maximum blood urea nitrogen (44 versus 30), and higher maximum white blood cell count (19 000 versus 14 000), and were more likely to have a 50% increase in creatinine (34% versus 19%) in the first 24 hours.

MIMIC Cohort

Of the 6802 patients admitted to a CICU in the MIMIC database, 2237 (32.9%) met the inclusion criteria and were included in the final validation cohort (Figure 1). The overall in-hospital mortality rate in the MIMIC cohort was 22.8% (Table S3).

Cardiogenic Shock Risk Score

After fitting several models using RiskSLIM for model size constraints between 1 and 10 and comparing performance on AUC and calibration (Figure S1), the RiskSLIM model with 6 variables was chosen based on both statistical performance and clinical interpretability. For similar accuracy, the 6-variable model had better calibration compared with the 4-, 5-, and 7-variable models.

In our final risk score, termed BOS,MA$_2$, a patient would receive a point for each of the following: maximum blood urea nitrogen ≥25 mg/dL, minimum oxygen saturation <88%, minimum systolic blood pressure <80 mm Hg, any use of mechanical ventilation, age ≥60 years, and maximum anion gap ≥14 mmol/L, based on values recorded during the first 24 hours of ICU stay (Figure 2). Predicted in-hospital mortality was 0.5% for a score of 0, 1.4% for a score of 1, 3.9% for a score of 2, 10.0% for a score of 3, 23.5% for a score of 4, 46% for a score of 5, and 70.2% for a score of 6. Patients with all risk score values were well represented in both the training and validation cohorts (Figure S2).

Model Validation

The AUC for the BOS,MA$_2$ risk score model was 0.83 (95% CI, 0.82–0.84) in the training data set and 0.76 (95% CI, 0.73–0.78) in the validation data set (Table 2; Figure 3A). This was only slightly lower than the comparator full PLR model with 51 variables (0.80 [95% CI, 0.78–0.82]), though similar to a parsimonious PLR model including only the same 6 variables in the risk score (0.76 [95% CI, 0.73–0.78]; Table S4). Subsequent inclusion of the BOS,MA$_2$ risk score itself into comparator full or parsimonious PLR models did not enhance prediction over baseline clinical characteristics alone (Table S4).

The calibration error for the BOS,MA$_2$ risk score model was 0.9% in the training data set and 2.6% in the validation data set (Table 2; Figure 3B). The calibration error for the BOS,MA$_2$ risk score model on the validation data set was lower than both the full PLR model and the parsimonious PLR model with the same features (Table S4).

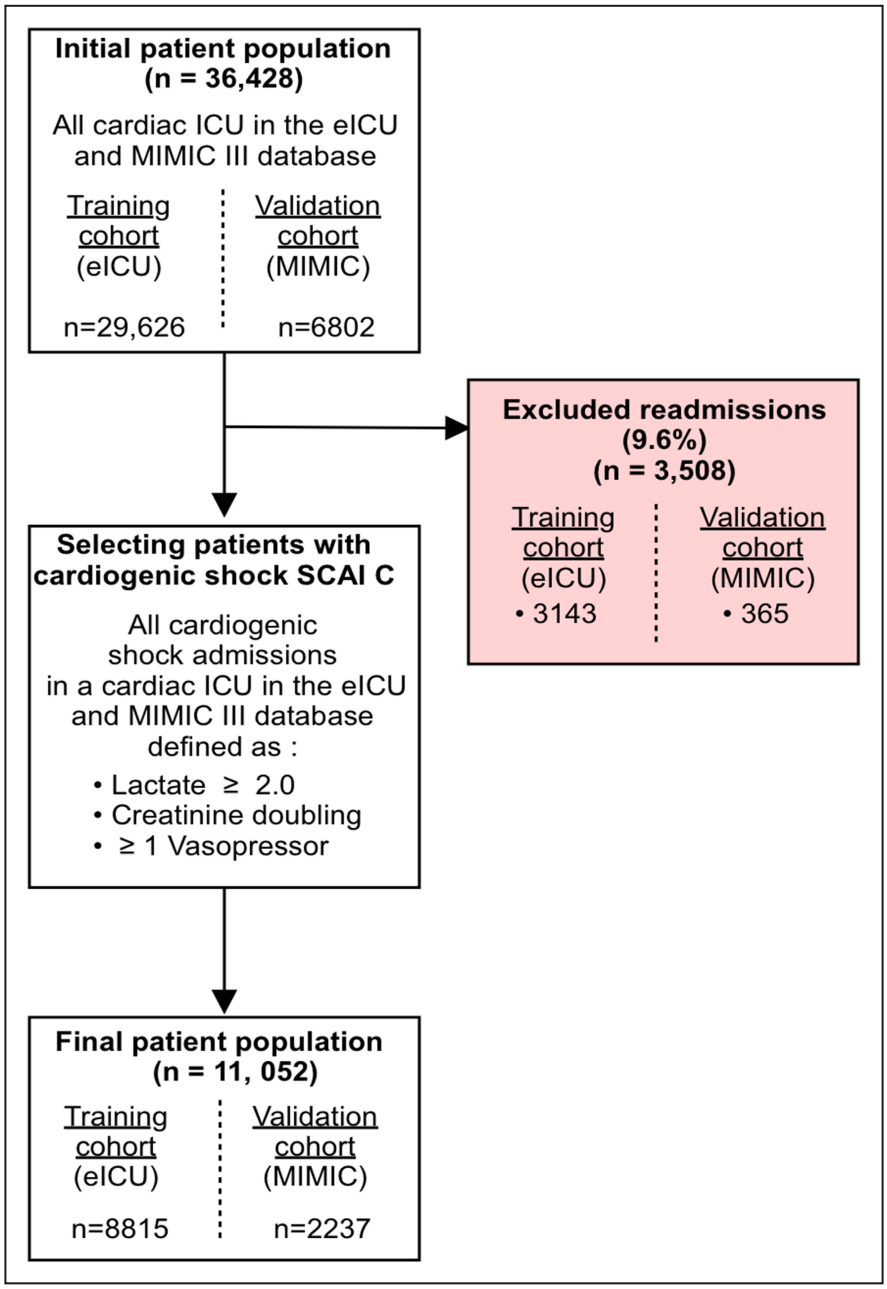

Figure 1. Study patient flow diagram.
eICU, indicates electronic intensive care unit; ICU, intensive care unit; MIMIC, Medical Information Mart for Intensive Care; and SCAI C, Society for Cardiovascular Angiography & Interventions Shock Stage C.

Comparison to Other Risk Scores

In our validation cohort, the AUC for the BOS,MA$_2$ score (0.76 [95% CI, 0.73–0.78]) was greater than that for the cardiogenic shock-specific CardShock score (0.66 [95% CI, 0.63–0.69]) and similar to general ICU risk scores (Sequential Organ Failure Assessment 0.75 [95% CI, 0.73–0.78]; Oxford Acute Severity of Illness Score 0.77 [95% CI, 0.74–0.79]; Table 2). However, the validation cohort ECE for the BOS,MA$_2$ score (2.6%) was lower than all other scores examined (CardShock 11.4%, Sequential Organ Failure Assessment 4.7%, Oxford Acute Severity of Illness Score 4.1%).

Supplemental Analysis

There was no reduction in the model's performance among the validation cohort subgroups examined (Table S5). Notably, among 305 patients with a cardiovascular primary admission diagnosis in the MIMIC

Table 1. Baseline Characteristics of eICU Cohort Stratified by Outcome

	Survived to discharge (n=7627)	In-hospital mortality (n=1188)	Standardized difference
Demographics			
Age, y, mean (SD)	63.69 (14.64)	67.80 (13.55)	0.291
Male sex (%)	4543 (59.6)	691 (58.2)	0.028
Comorbidities			
Cerebral vascular disease (%)	830 (10.9)	152 (12.8)	0.059
Anemia (%)	40 (0.5)	11 (0.9)	0.047
Atrial fibrillation (%)	1065 (14.0)	224 (18.9)	0.132
Blood malignancy (%)	104 (1.4)	32 (2.7)	0.094
Solid neoplasm (%)	876 (11.5)	192 (16.2)	0.136
Congestive heart failure (%)	1703 (22.3)	323 (27.2)	0.113
Chronic kidney disease (%)	1302 (17.1)	282 (23.7)	0.166
Chronic obstructive pulmonary disease (%)	1192 (15.6)	266 (22.4)	0.173
Coronary artery disease (%)	2099 (27.5)	287 (24.2)	0.077
Diabetes (%)	2575 (33.8)	409 (34.4)	0.014
Valvulopathy (%)	824 (10.8)	98 (8.2)	0.087
Hypertension (%)	4778 (62.6)	718 (60.4)	0.045
Metastatic cancer (%)	92 (1.2)	33 (2.8)	0.113
Prior myocardial infarction (%)	1191 (15.6)	156 (13.1)	0.071
Cardiac ICU therapies			
Renal replacement therapy(%)	375 (4.9)	84 (7.1)	0.091
Mechanical ventilation (%)	2099 (27.5)	672 (56.6)	0.616
Intra-aortic balloon pump (%)	363 (4.8)	80 (6.7)	0.085
≥1 vasopressor (%)	6917 (90.7)	1040 (87.5)	0.101
≥1 inotrope (%)	1156 (15.2)	201 (16.9)	0.048
Vital signs			
Heart rate min, mean (SD)	63.19 (13.78)	59.85 (20.53)	0.191
Heart rate max, mean (SD)	113.12 (25.30)	129.80 (27.79)	0.628
Systolic BP min, mean (SD)	88.12 (19.18)	70.67 (20.45)	0.880
Systolic BP max, mean (SD)	158.73 (27.69)	156.33 (33.45)	0.078
Respiratory rate min, mean (SD)	10.44 (4.93)	8.22 (7.22)	0.359
Respiratory rate max, mean (SD)	32.26 (9.37)	36.29 (9.69)	0.423
Oxygen saturation min, mean (SD)	86.31 (13.97)	69.23 (24.93)	0.845
Laboratory results			
Glucose min, mean (SD)	117.98 (45.25)	128.73 (68.16)	0.186
Anion gap max, mean (SD)	12.89 (5.85)	17.87 (7.60)	0.734
Bicarbonate min, mean (SD)	22.58 (4.99)	18.97 (6.18)	0.642
Chloride max, mean (SD)	105.98 (6.39)	105.99 (7.92)	0.002
Hematocrit max, mean (SD)	37.15 (6.50)	36.87 (7.41)	0.041
Hemoglobin min, mean (SD)	10.58 (2.37)	10.11 (2.57)	0.188
Platelet min, mean (SD)	181.49 (89.18)	173.15 (98.23)	0.089
Potassium max, mean (SD)	4.61 (0.81)	4.99 (1.00)	0.405
International normalized ratio max, mean (SD)	1.51 (0.88)	2.11 (1.71)	0.439
Sodium min (mean (SD))	135.61 (5.17)	135.33 (6.09)	0.049
Blood urea nitrogen max, mean (SD)	29.70 (22.38)	44.15 (28.00)	0.570
White blood cell max, mean (SD)	14.02 (9.10)	18.61 (16.95)	0.338
Red cell distribution width max, mean (SD)	15.32 (2.34)	16.60 (3.03)	0.472
Creatinine >1.5× baseline (%)	1439 (18.9)	409 (34.4)	0.358

(*Continued*)

Table 1. Continued

	Survived to discharge (n=7627)	In-hospital mortality (n=1188)	Standardized difference
Outcomes			
ICU length of stay, d, mean (SD)	1.48 (2.09)	1.77 (2.25)	0.133
Hospital length of stay, d, mean (SD)	9.17 (9.68)	7.35 (8.85)	0.197
Time to death, d, mean (SD)	—	5.39 (6.98)	—

BP indicates blood pressure; and ICU, intensive care unit.

data set, the BOS,MA$_2$ risk score model had similar discrimination (AUC 0.78 [95% CI, 0.73–0.75]) and calibration (ECE 1.05%) compared with when applied to the entire validation cohort. Similarly, among 208 patients with cardiogenic shock associated with MI necessitating acute coronary revascularization and 478 patients with cardiogenic shock without MI necessitating acute coronary revascularization, the model's performance was comparable with an AUC of 0.83 (95% CI, 0.77–0.88) and 0.79 (95% CI, 0.75–0.83), respectively. Notably, the ECE of the BOS,MA$_2$ risk score was significantly lower than that of the general CICU risk scores in looking at these cardiovascular-specific subgroups (Table S6).

DISCUSSION

We developed and externally validated a checklist-based risk score to predict mortality among patients with cardiogenic shock admitted to the CICU using a machine learning algorithm and distinct large real-world training and validation data sets. The BOS,MA$_2$ risk score is methodologically robust, generalizable, and readily applicable at the bedside. In addition, this clinical tool has direct practical implications for managing patients with cardiogenic shock.

The BOS,MA$_2$ risk score model outperforms existing cardiogenic shock and general ICU mortality risk scores in predicting mortality among patients with cardiogenic shock in the CICU. The BOS,MA$_2$ risk score achieves higher discrimination in external validation than the published CardShock, and IABP-SHOCK II (Intraaortic Balloon Pump in Cardiogenic Shock II) risk score models (Table 3),[11,1733,34] and performs better than the CardShock score in our own external validation cohort (AUC 0.76 versus 0.66; ECE 2.6% versus 11.4%). Although general ICU risk scores can have high AUCs in external validation in all-comer ICU cohorts, they have similar discrimination as the BOS,MA$_2$ risk score in our cardiogenic shock cohort, indicating

Variable	Points	
maximum **B**UN ≥25 mg/dL	1	+ ___
minimum **O**xygen saturation < 88%	1	+ ___
minimum **S**ystolic blood pressure <80 mm Hg	1	+ ___
Mechanical ventilation	1	+ ___
Age≥ 60 years	1	+ ___
maximum **A**nion gap ≥14 mmol/L	1	+ ___
	Score	= ___

SCORE	0	1	2	3	4	5	6
RISK	0.5%	1.4%	3.9%	10.0%	23.5%	46.0%	70.2%

Figure 2. BOS,MA$_2$ risk score for cardiogenic shock mortality in the cardiac intensive care unit.
A patient receives 1 point for meeting each of the criteria as specified in the risk score. The BOS,MA$_2$ risk score is calculated by summing the number of points. The mortality risk corresponds to the number of points in the reference table. BUN indicates blood urea nitrogen. BOS,MA$_2$ risk score defined as maximum BUN ≥25 mg/dL, minimum oxygen saturation <88%, minimum systolic blood pressure <80 mm Hg, any use of mechanical ventilation, age ≥60 years, and maximum anion gap ≥14 mmol/L.

Table 2. BOS,MA$_2$ Risk Score Model Performance Compared With Other Validated Risk Scores

Cohort	eICU (training)		MIMIC-III (validation)	
Risk score	AUC (95% CI)*	ECE	AUC (95% CI)*	ECE
BOS,MA$_2$	0.83 (0.82–0.84)	0.9%	0.76 (0.73–0.78)	2.6%
CardShock score†			0.66 (0.63–0.69)	11.4%
Sequential Organ Failure Assessmentscore, day 1	0.76 (0.74–0.78)	6%	0.75 (0.73–0.78)	4.7%
OASIS score, day 1			0.77 (0.74–0.79)	4.1%

AUC indicates area under the curve; BOS,MA$_2$, blood urea nitrogen ≥25 mg/dL, minimum oxygen saturation <88%, minimum systolic blood pressure <80 mm Hg, any use of mechanical ventilation, age ≥60 years, and maximum anion gap ≥14 mmol/L; ECE, expected calibration error; eICU, electronic intensive care unit; MIMIC, Medical Information Mart for Intensive Care; and OASIS, Oxford Acute Severity of Illness Score.

Variables to calculate CardShock and OASIS scores not available in eICU (training) data set.

*95% CI computed using the DeLong statistic.

†Calculated based on 1269 patients with complete information available for all variables.

the challenges of predicting mortality in this complex cardiogenic shock population. Notably, the BOS,MA$_2$ risk score model has better calibration than Sequential Organ Failure Assessment score and Oxford Acute Severity of Illness Score in our validation cohort, particularly in looking at cardiovascular-specific subgroups, such as those with MI necessitating acute coronary revascularization. Calibration error in clinical populations in which a score will be used is an important, yet often overlooked, feature of risk score performance in cardiology.[35] Better calibration justifies using the BOS,MA$_2$ score over general ICU mortality risk scores for patients in cardiogenic shock in the CICU.

The BOS,MA$_2$ risk score contains certain elements present in other cardiogenic shock risk prediction models but also includes unique features (Table 3). Age has been consistently associated with an increased risk of death in cardiogenic shock, though cutoffs vary across studies.[11,13–17,36–39] The RiskSLIM method used in our study enabled us to consider a range of cutoffs simultaneously, and age over 60 was deemed the most discriminant. Kidney dysfunction, as measured by blood urea nitrogen in our study, is also predictive of cardiogenic shock mortality in other studies in the form of creatinine level[15–17] or glomerular filtration rate.[11] Many existing models have also included systolic blood pressure[15,37] or other proxies for hypoperfusion, such as lactate,[11,16,17,36] altered mental status,[11,15] and anion gap.[40] This is consistent with our model, which also includes both systolic blood pressure and an elevated anion gap.

The BOS,MA$_2$ risk score includes 2 distinct markers of respiratory failure: mechanical ventilation and hypoxemia (SpO_2 < 88%). Only 1 other model predicting outcomes among cardiogenic shock patients included respiratory failure. However, this model was built on a more heterogenous population of patients admitted to the CICU, and only 14% of patients in the development cohort had shock.[40] Most published cardiogenic shock mortality risk scores did not consider mechanical ventilation or oxygen saturation as candidate predictors during development, suggesting that respiratory function may frequently be overlooked despite the pathophysiologic mechanisms linking respiratory and cardiac dysfunction in cardiogenic shock. The use of large real-world data sets with a wide variety of candidate input parameters enabled the detection of respiratory function inputs as critical predictors of cardiogenic shock mortality using the RiskSLIM method.

Our model has implications for the clinical care patients with cardiogenic shock and for designing future clinically relevant risk scores. The BOS,MA$_2$ risk score can quickly be calculated at the bedside as a checklist of 6 objective variables that are often readily available, making it easy to remember and implement in clinical decision-making situations, such as while rounding on patients in the CICU. Additionally, given that it was derived from raw electronic health record data, this risk score can easily be integrated in the electronic health record to facilitate its adoption into clinical practice, which has been an issue with adoption of other clinical risk prediction tools.[41] The prediction tool can be used in conjunction with the SCAI Shock classification system to help triage patients efficiently, obtain reliable prognostic information, and guide clinical decision-making for a challenging patient population for which a myriad of potentially high-risk therapeutic options are available.[8,9] Additionally, this risk score can be used to assess the impact of treatment strategies on expected mortality and can enable the design of future clinical trials with more homogenous populations. Our model also has broader implications for the development of risk scores in clinical cardiology. The machine learning algorithm used in this article can be used in other clinical contexts. Our article thus serves as a model for developing interpretable risk scores in cardiology using the RiskSLIM methodology.

Our study has several key strengths. First, we used 2 large real-world cohorts (n=8815 for training, n=2237 for external validation), which are the largest cohorts of patients used to develop a cardiogenic shock mortality prediction model to date. Other scores have mainly

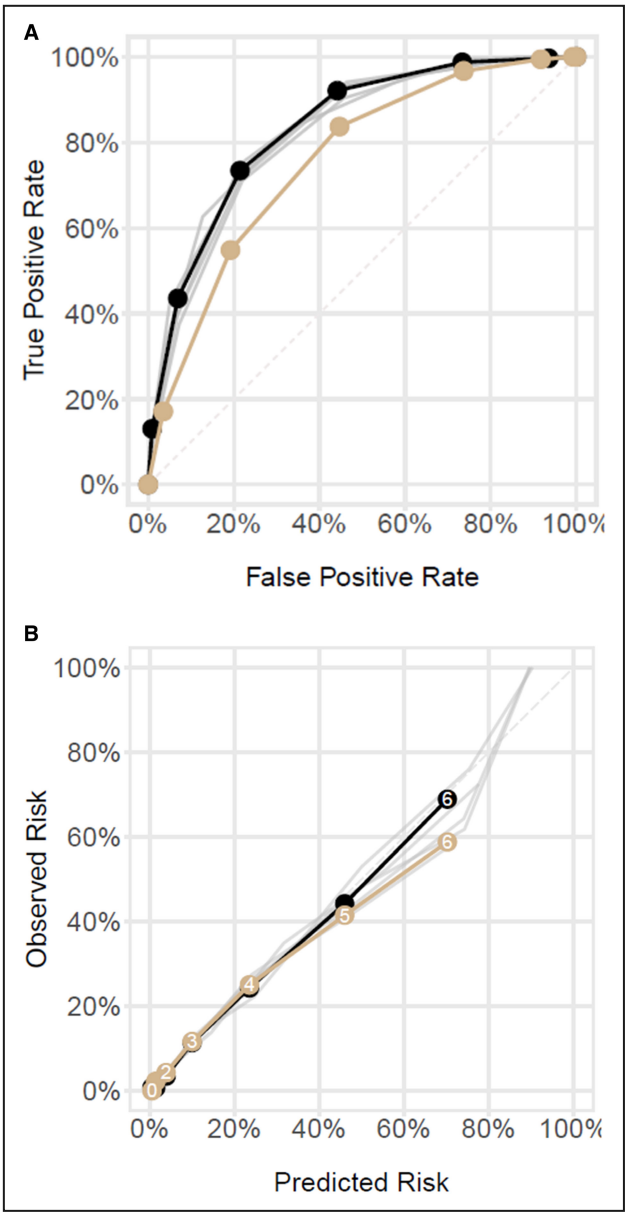

Figure 3. BOS,MA$_2$ risk score performance.
A, Receiver operating curve of BOS,MA$_2$ risk score model. **B,** Calibration plot of BOS,MA$_2$ risk score model. BOS,MA$_2$ risk score defined as maximum blood urea nitrogen ≥25 mg/dL, minimum oxygen saturation <88%, minimum systolic blood pressure <80 mm Hg, any use of mechanical ventilation, age ≥60 years, and maximum anion gap ≥14 mmol/L. Black=training cohort, gray=5-fold cross validation, brown=validation cohort.

used much smaller cohorts derived from clinical trials or with narrower inclusion criteria focusing on a single cause of cardiogenic shock or specific mechanical circulatory support,[14,17,39] which may have limited generalizability when applied to real-world settings. Second, we employed a machine learning algorithm specifically designed to fit simple customized clinically interpretable risk scores optimized to yield calibrated risk estimates. This is in stark contrast to other machine learning methods, which often offer only minor improvements in calibration at the expense of clinical interpretability when applied to generating clinical risk scores, such as in the case of acute MI mortality prediction.[42] In comparison to traditional logistic regression methods, we found similar or improved performance despite using only 6 terms as a checklist, which can aid in calculation at the bedside.

Our findings must be interpreted in the context of their limitations. First, the in-hospital mortality observed in our training cohort is lower than that observed in our validation cohort, likely reflecting the lower severity of patients in the CICUs represented in the eICU cohort, which are primarily community hospitals, relative to those in the singular large tertiary care center represented by the MIMIC database. However, the similar performance of our model across both cohorts reflects the generalizability of our risk score and substantiates its broad applicability across a range of clinical settings. Second, although we included only patients admitted to CICUs, it is possible that we have included patients with septic or mixed shock in our cohorts given that we did not consider admission diagnosis in defining our primary cohort. However, model performance was similar in subgroups of patients with a cardiovascular primary ICU admission diagnosis and with cardiogenic shock with MI necessitating acute revascularization. This suggests that inclusion of such patients was unlikely to affect our results and that our model is valid in these subpopulations. Third, we may not have captured all patients in our real-world data sets meeting the SCAI Shock C classification criteria due to missing data. However, the mortality rate in our validation cohort was similar to that in another real-world tertiary care center cohort that operationalized the SCAI shock classification criteria,[24] suggesting that the population of patients included in our study likely reflects the true population of patients with cardiogenic shock in CICUs. Fourth, several clinically relevant variables in other risk scores were unavailable in our data set due to inconsistent capture. Specifically, the presence of mechanical circulatory support beyond the intra-aortic balloon pump was not documented accurately in the eICU database and was likely rare, given that eICU hospitals were primarily community hospitals. However, we included a comprehensive range of other candidate variables, which captured similar clinical constructs (such as the use of anion gap in place of blood lactate), and our model performance exceeded that of other published models. Fifth, our model is developed based on data from the first 24 hours of CICU admission only. Although other variables may be more critical for prognosis at other time points in a patient's

Table 3. Comparison of Established Cardiogenic Shock Risk Scores

	BOS,MA$_2$ score	CardShock score[11]	IABP-SHOCK II[17]	SHOCK trial registry[15]
Training cohort sample size	8815	219	480	1217
Validation cohort sample size	2237	384	235	None
Validation	External	External	External	None
Area under the curve internal, (external)	0.83 (0.76)	0.85 (0.71)	0.79 (0.73)	0.76 (–)
Total number of predictors	6	7	6	8
Past medical history				
History of coronary artery bypass graft		X		X
History of stroke			X	
Admission diagnosis				
Acute coronary syndrome on admission		X		X
Shock on admission				X
Clinical variables				
Age	X	X	X	X
Renal function	X	X	X	X
Glucose >10.6 mmol/L		X	X	X
Anion gap >14	X			
End-organ hypoperfusion		X (lactate)	X (lactate)	X (clinical end point)
Neurologic dysfunction		X (confusion)		X (anoxic brain damage)
Mechanical ventilation	X			
Systolic blood pressure	X			X
Oxygen saturation <88	X			
Catheterization laboratory and hemodynamic measurements				
Thrombolysis in myocardial infarction flow			X	
Left ventricular ejection fraction		X		

BOS,MA$_2$ indicates blood urea nitrogen ≥25 mg/dL, minimum oxygen saturation <88%, minimum systolic blood pressure <80 mm Hg, any use of mechanical ventilation, age ≥60 years, and maximum anion gap ≥14 mmol/L; IABP-SHOCK II, Intraaortic Balloon Pump in Cardiogenic Shock II; and SHOCK trial, Should We Emergently Revascularize Occluded Coronaries in Cardiogenic Shock.

CICU course, we felt that the first 24 hours was the period during which many major triage and therapeutic decisions are made in the CICU. Understanding the relative importance of other variables at different time points in a patient's CICU course remains a rich area for future inquiry. Finally, as the population of patients in the CICU changes over time, it is very likely that this score will need recalibration in the future, which can likely be accomplished using newer versions of the validation data set in the future.[43]

CONCLUSIONS

In conclusion, we used 2 large real-world data sets and a machine learning method to develop an externally validated mortality prediction risk score for patients with cardiogenic shock in the cardiac intensive care unit. The BOS,MA$_2$ risk score is a simple, clinically-interpretable score that can guide clinical decision-making for patients with cardiogenic shock. This score can serve as a model for developing future risk scores in cardiology.

ARTICLE INFORMATION

Received December 22, 2022; accepted May 1, 2023.

Affiliations

Department of Medicine, Centre Hospitalier de l'Université de Montréal (CHUM), Montreal, QC, Canada (E.Y.); Harvard Medical School, Boston, MA (S.M., R.W.Y., L.A.C.); Johns Hopkins School of Medicine, Baltimore, MD (D.R.); University of Colorado School of Medicine, Aurora, CO (E.M.B., N.M.B.); Heart Institute, Children's Hospital of Colorado, Aurora, CO (E.M.B.); Richard A. and Susan F. Smith Center for Outcomes Research in Cardiology, Division of Cardiology, Beth Israel Deaconess Medical Center, Boston, MA (R.W.Y.); Laboratory for Computational Physiology, MIT Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA (L.A.C.); Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA (L.A.C.); Halıcıoğlu Data Science Institute, University of California, San Diego, CA (B.U.); and Rocky Mountain Regional VA Medical Center, Aurora, CO (N.M.B.).

Acknowledgments

The computer code used to generate the analyses is available on GitHub (https://github.com/ustunb/cshock), including preprocessing and modeling steps.

Sources of Funding

Dr. Butala is supported by the Boettcher Foundation Webb-Waring Biomedical Research Grant. Dr. Ustun is supported by the National Science Foundation IIS 2040880.

Disclosures

Dr Butala reports consulting fees and ownership interest in Hilabs, Inc. Dr Yeh reports grant/contracts support from Abiomed, AstraZeneca, BD Bard, Boston Scientific, Cook Medical, Medtronic, and Philips, and consulting fees from Abbott, Boston Scientific, Edwards Lifesciences, Medtronic, and Shockwave Medical outside the submitted work. The other authors report no conflicts.

Supplemental Material

Tables S1–S6.
Figures S1–S2.

REFERENCES

1. van Diepen S, Katz JN, Albert NM, Henry TD, Jacobs AK, Kapur NK, Kilic A, Menon V, Ohman EM, Sweitzer NK, et al. Contemporary management of cardiogenic shock: a scientific statement from the American Heart Association. *Circulation*. 2017;136:e232–e268. doi: 10.1161/CIR.0000000000000525
2. Berg DD, Bohula EA, van Diepen S, Katz JN, Alviar CL, Baird-Zars VM, Barnett CF, Barsness GW, Burke JA, Cremer PC, et al. Epidemiology of shock in contemporary cardiac intensive care units. *Circ Cardiovasc Qual Outcomes*. 2019;12:e005618. doi: 10.1161/CIRCOUTCOMES.119.005618
3. Thiele H, Zeymer U, Neumann F-J, Ferenc M, Olbrich H-G, Hausleiter J, Richardt G, Hennersdorf M, Empen K, Fuernau G, et al. Intraaortic balloon support for myocardial infarction with cardiogenic shock. *N Engl J Med*. 2012;367:1287–1296. doi: 10.1056/NEJMoa1208410
4. TRIUMPH Investigators; Alexander JH, Reynolds HR, Stebbins AL, Dzavik V, Harrington RA, Van de Werf F, Hochman JS. Effect of tilarginine acetate in patients with acute myocardial infarction and cardiogenic shock: the TRIUMPH randomized controlled trial. *JAMA*. 2007;297:1657–1666. doi: 10.1001/jama.297.15.joc70035
5. Thiele H, Jobs A, Ouweneel DM, Henriques JPS, Seyfarth M, Desch S, Eitel I, Pöss J, Fuernau G, de Waha S. Percutaneous short-term active mechanical support devices in cardiogenic shock: a systematic review and collaborative meta-analysis of randomized trials. *Eur Heart J*. 2017;38:3523–3531. doi: 10.1093/eurheartj/ehx363
6. Hochman JS, Sleeper LA, Webb JG. Early revascularization in acute myocardial infarction complicated by cardiogenic SHOCK. SHOCK investigators. Should We Emergently Revascularize Occluded Coronaries for Cardiogenic Shock. *N Engl J Med*. 1999;341:625–634. doi: 10.1056/NEJM199908263410901
7. Peura JL, Colvin-Adams M, Francis GS, Grady KL, Hoffman TM, Jessup M, John R, Kiernan MS, Mitchell JE, O'Connell JB, et al. Recommendations for the use of mechanical circulatory support: device strategies and patient selection: a scientific statement from the American Heart Association. *Circulation*. 2012;126:2648–2667. doi: 10.1161/CIR.0b013e3182769a54
8. Baran DA, Grines CL, Bailey S, Burkhoff D, Hall SA, Henry TD, Hollenberg SM, Kapur NK, O'Neill W, Ornato JP, et al. SCAI clinical expert consensus statement on the classification of cardiogenic shock: this document was endorsed by the American College of Cardiology (ACC), the American Heart Association (AHA), the Society of Critical Care Medicine (SCCM), and the Society of Thoracic Surgeons (STS) in April 2019. *Catheter Cardiovasc Interv*. 2019;94:29–37. doi: 10.1002/ccd.28329
9. Naidu SS, Baran DA, Jentzer JC, Hollenberg SM, van Diepen S, Basir MB, Grines CL, Diercks DB, Hall S, Kapur NK, et al. SCAI SHOCK stage classification expert consensus update: a review and incorporation of validation studies. *J Am Coll Card*. 2022;79:933–946. doi:10.1016/j.jacc.2022.01.018
10. Kellner P, Prondzinsky R, Pallmann L, Siegmann S, Unverzagt S, Lemm H, Dietz S, Soukup J, Werdan K, Buerke M. Predictive value of outcome scores in patients suffering from cardiogenic shock complicating AMI: APACHE II, APACHE III, Elebute-Stoner, SOFA, and SAPS II. *Med Klin Intensivmed Notfmed*. 2013;108:666–674. doi: 10.1007/s00063–013–0234–2
11. Harjola V-P, Lassus J, Sionis A, Køber L, Tarvasmäki T, Spinar J, Parissis J, Banaszewski M, Silva-Cardoso J, Carubelli V, et al. Clinical picture and risk prediction of short-term mortality in cardiogenic shock. *Eur J Heart Fail*. 2015;17:501–509. doi: 10.1002/ejhf.260
12. Granger CB, Goldberg RJ, Dabbous O, Pieper KS, Eagle KA, Cannon CP, Van De Werf F, Avezum A, Goodman SG, Flather MD, et al. Predictors of hospital mortality in the global registry of acute coronary events. *Arch Intern Med*. 2003;163:2345–2353. doi: 10.1001/archinte.163.19.2345
13. Hasdai D, Holmes DR, Califf RM, Thompson TD, Hochman JS, Pfisterer M, Topol EJ. Cardiogenic shock complicating acute myocardial infarction: predictors of death. GUSTO investigators. Global Utilization of Streptokinase and Tissue-Plasminogen Activator for Occluded Coronary Arteries. *Am Heart J*. 1999;138:21–31. doi: 10.1016/S0002–8703(99)70241–3
14. Garcia-Alvarez A, Arzamendi D, Loma-Osorio P, Kiamco R, Masotti M, Sionis A, Betriu A, Brugada J, Bosch X. Early risk stratification of patients with cardiogenic shock complicating acute myocardial infarction who undergo percutaneous coronary intervention. *Am J Cardiol*. 2009;103:1073–1077. doi: 10.1016/j.amjcard.2008.12.033
15. Sleeper LA, Reynolds HR, White HD, Webb JG, Dzavík V, Hochman JS. A severity scoring system for risk assessment of patients with cardiogenic shock: a report from the SHOCK trial and registry. *Am Heart J*. 2010;160:443–450. doi: 10.1016/j.ahj.2010.06.024
16. Cheng JM, Helming AM, van Vark LC, Kardys I, Den Uil CA, Jewbali LSD, van Geuns R-J, Zijlstra F, van Domburg RT, Boersma E, et al. A simple risk chart for initial risk assessment of 30-day mortality in patients with cardiogenic shock from ST-elevation myocardial infarction. *Eur Heart J Acute Cardiovasc Care*. 2016;5:101–107. doi: 10.1177/2048872615568966
17. Pöss J, Köster J, Fuernau G, Eitel I, de Waha S, Ouarrak T, Lassus J, Harjola V-P, Zeymer U, Thiele H, et al. Risk stratification for patients in cardiogenic shock after acute myocardial infarction. *J Am Coll Cardiol*. 2017;69:1913–1920. doi: 10.1016/j.jacc.2017.02.027
18. Ustun B, Rudin C. Learning optimized risk scores. *Journal of Machine Learning Research*. 2019;20:1–75.
19. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU collaborative research database, a freely available multi-center database for critical care research. *Sci Data*. 2018;5:180178. doi: 10.1038/sdata.2018.178
20. Johnson AEW, Pollard TJ, Shen L, Lehman L-WH, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3:160035. doi: 10.1038/sdata.2016.35
21. Jentzer JC. Understanding cardiogenic shock severity and mortality risk assessment. *Circ Heart Fail*. 2020;13:e007568. doi: 10.1161/CIRCHEARTFAILURE.120.007568
22. Schrage B, Dabboura S, Yan I, Hilal R, Neumann JT, Sörensen NA, Goßling A, Becher PM, Grahn H, Wagner T, et al. Application of the SCAI classification in a cohort of patients with cardiogenic shock. *Catheter Cardiovasc Interv*. 2020;96:E213–E219. doi: 10.1002/ccd.28707
23. Thayer KL, Zweck E, Ayouty M, Garan AR, Hernandez-Montfort J, Mahr C, Morine KJ, Newman S, Jorde L, Haywood JL, et al. Invasive hemodynamic assessment and classification of In-hospital mortality risk among patients with cardiogenic shock. *Circ Heart Fail*. 2020;13:e007099. doi: 10.1161/CIRCHEARTFAILURE.120.007099
24. Jentzer JC, van Diepen S, Barsness GW, Henry TD, Menon V, Rihal CS, Naidu SS, Baran DA. Cardiogenic shock classification to predict mortality in the cardiac intensive care unit. *J Am Coll Cardiol*. 2019;74:2117–2128. doi: 10.1016/j.jacc.2019.07.077
25. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Commun Stat Simul Comput*. 2009;38:1228–1234. doi: 10.1080/03610910902859574
26. Ustun B, Adler LA, Rudin C, Faraone SV, Spencer TJ, Berglund P, Gruber MJ, Kessler RC. The World Health Organization adult attention-deficit/hyperactivity disorder self-report screening scale for DSM-5. *JAMA Psychiatry*. 2017;74:520–527. doi: 10.1001/jamapsychiatry.2017.0298
27. Struck AF, Ustun B, Ruiz AR, Lee JW, LaRoche SM, Hirsch LJ, Gilmore EJ, Vlachy J, Haider HA, Rudin C, et al. Association of an electroencephalography-based risk score with seizure probability in

hospitalized patients. *JAMA Neurol*. 2017;74:1419–1424. doi: 10.1001/jamaneurol.2017.2459
28. Gage BF, van Walraven C, Pearce L, Hart RG, Koudstaal PJ, Boode BSP, Petersen P. Selecting patients with atrial fibrillation for anticoagulation: stroke risk stratification in patients taking aspirin. *Circulation*. 2004;110:2287–2292. doi: 10.1161/01.CIR.0000145172.55640.93
29. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA*. 1993;270:2957–2963. doi: 10.1001/jama.1993.03510240069035
30. Alpaydin E. *Introduction to Machine Learning*. MIT Press; 2020. doi: 10.7551/mitpress/13811.001.0001
31. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17:230. doi: 10.1186/s12916–019–1466–7
32. Stevens LM, Mortazavi BJ, Deo RC, Curtis L, Kao DP. Recommendations for reporting machine learning analyses in clinical research. *Circ Cardiovasc Qual Outcomes*. 2020;13:e006556. doi: 10.1161/CIRCOUTCOMES.120.006556
33. Rivas-Lasarte M, Sans-Roselló J, Collado-Lledó E, González-Fernández V, Noriega FJ, Hernández-Pérez FJ, Fernández-Martínez J, Ariza A, Lidón R-M, Viana-Tejedor A, et al. External validation and comparison of the CardShock and IABP-SHOCK II risk scores in real-world cardiogenic shock patients. *Eur Heart J Acute Cardiovasc Care*. 2020;10:16–24. doi: 10.1177/2048872619895230
34. Miller RJH, Southern D, Wilton SB, James MT, Har B, Schnell G, van Diepen S, Grant ADM. Comparative prognostic accuracy of risk prediction models for cardiogenic shock. *J Intensive Care Med*. 2020;35:1513–1519. doi: 10.1177/0885066619878125
35. Blaha MJ. The critical importance of risk score calibration: time for transformative approach to risk score validation? *J Am Coll Cardiol*. 2016;67:2131–2134. doi: 10.1016/j.jacc.2016.03.479
36. Valente S, Lazzeri C, Vecchio S, Giglioli C, Margheri M, Bernardo P, Comeglio M, Chiocchini S, Gensini GF. Predictors of in-hospital mortality after percutaneous coronary intervention for cardiogenic shock. *Int J Cardiol*. 2007;114:176–182. doi: 10.1016/j.ijcard.2006.01.024
37. Katz JN, Stebbins AL, Alexander JH, Reynolds HR, Pieper KS, Ruzyllo W, Werdan K, Geppert A, Dzavik V, Van de Werf F, et al. Predictors of 30-day mortality in patients with refractory cardiogenic shock following acute myocardial infarction despite a patent infarct artery. *Am Heart J*. 2009;158:680–687. doi: 10.1016/j.ahj.2009.08.005
38. Zeymer U, Vogt A, Zahn R, Weber MA, Tebbe U, Gottwik M, Bonzel T, Senges J, Neuhaus K-L; Arbeitsgemeinschaft Leitende Kardiologische Krankenhausärzte (ALKK). Predictors of in-hospital mortality in 1333 patients with acute myocardial infarction complicated by cardiogenic shock treated with primary percutaneous coronary intervention (PCI); results of the primary PCI registry of the Arbeitsgemeinschaft Leitende Kardiologische Krankenhausärzte (ALKK). *Eur Heart J*. 2004;25:322–328. doi: 10.1016/j.ehj.2003.12.008
39. Klein LW, Shaw RE, Krone RJ, Brindis RG, Anderson HV, Block PC, McKay CR, Hewitt K, Weintraub WS;. American College of Cardiology National Cardiovascular Data Registry. Mortality after emergent percutaneous coronary intervention in cardiogenic shock secondary to acute myocardial infarction and usefulness of a mortality prediction model. *Am J Cardiol*. 2005;96:35–41. doi: 10.1016/j.amjcard.2005.02.040
40. Jentzer JC, Anavekar NS, Bennett C, Murphree DH, Keegan MT, Wiley B, Morrow DA, Murphy JG, Bell MR, Barsness GW. Derivation and validation of a novel cardiac intensive care unit admission risk score for mortality. *J Am Heart Assoc*. 2019;8:e013675. doi: 10.1161/JAHA.119.013675
41. Sharma V, Ali I, van der Veer S, Martin G, Ainsworth J, Augustine T. Adoption of clinical risk prediction tools is limited by a lack of integration with electronic health records. *BMJ Health Care Inform*. 2021;28:28. doi: 10.1136/bmjhci–2020–100253
42. Khera R, Haimovich J, Hurley NC, McNamara R, Spertus JA, Desai N, Rumsfeld JS, Masoudi FA, Huang C, Normand S-L, et al. Use of machine learning models to predict death after acute myocardial infarction. *JAMA Cardiol*. 2021;6:633–641. doi: 10.1001/jamacardio.2021.0122
43. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, Pollard TJ, Moody B, Gow B, Lehman L-WH, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023;10:1. doi: 10.1038/s41597–022–01899–x

SUPPLEMENTAL MATERIAL

Table S1. Variables in eICU and MIMIC cohorts with percent missing.

	eICU - Training Cohort (n=8815)	Missing (%)	**MIMIC III - Validation Cohort (n=2237)**	Missing (%)
Demographics				
Age (mean (SD))	64.07 (14.60)	2.7	69.79 (13.78)	0
BMI (mean (SD))	29.69 (9.70)	5	28.39 (6.69)	32.3
Sex = M (%)	5234 (59.4)	0	1314 (58.7)	0
Ethnicity (%)		0		0
Asian	123 (1.4)		43 (1.9)	
Black	1356 (15.4)		143 (6.4)	
Hispanic	357 (4.0)		37 (1.7)	
Other	486 (5.5)		433 (19.4)	
White	6493 (73.7)		1581 (70.7)	
Comorbidities				
AIDS/HIV (%)	0 (0)	0	16 (0.7)	0
Acute cerebral vascular disease (%)	82 (11.1)	0	87 (3.9)	0
Anemia (%)	51 (0.6)	0	623 (27.8)	0
Atrial Fibrillation (%)	1289 (14.6)	0	911 (40.7)	0
Blood Malignancy (%)	136 (1.5)	0	63 (2.8)	0
CAD (%)	2386 (27.1)	0	1303 (58.2)	0
COPD (%)	1458 (16.5)	0	345 (15.4)	0
Chronic Kidney Disease (%)	1584 (18.0)	0	373 (16.7)	0
Congestive Heart Failure (%)	2026 (23.0)	0	1290 (57.7)	0
Diabetes Mellitus II (%)	2984 (33.9)	0	798 (35.7)	0
Valvulopathy (%)	922 (10.5)	0	548 (24.5)	0
Hypertension (%)	5496 (62.3)	0	1319 (59.0)	0
Prior MI (%)	1347 (15.3)	0	1054 (47.1)	0
Peripheral Vascular Disease (%)	0 (0)	0	285 (12.7)	0
Dementia (%)	0 (0)	0	25 (1.1)	0
Rheumatoid Arthritis/Collagen Vascular Disease (%)	0 (0)	0	66 (3.0)	0
Peptic Ulcer Disease (%)	0 (0)	0	30 (1.3)	0
Metastatic Cancer (%)	125 (1.4)	0	42 (1.9)	0
Mild Liver Disease (%)	0 (0)	0	203 (9.1)	0
Sever Liver Disease (%)	0 (0)	0	39 (1.7)	0
Solid Neoplasm (%)	1068 (12.1)	0	116 (5.2)	0
Charlson Score (%)		0		0
0	2747 (31.2)		146 (6.5)	
1	2423 (27.5)		469 (21.0)	
2	1661 (18.8)		690 (30.8)	
3	1010 (11.5)		523 (23.4)	
4	501 (5.7)		274 (12.2)	
5	230 (2.6)		97 (4.3)	
6	80 (0.9)		31 (1.4)	
7	32 (0.4)		7 (0.3)	
8	64 (0.7)		-	
9	33 (0.4)		-	
10	18 (0.2)		-	
11	15 (0.2)		-	
12	1 (0.0)		-	
CICU therapies				
Renal Replacement Therapy (%)	459 (5.2)	0	92 (4.1)	0
Mechanical Ventilation (%)	2771 (31.4)	0	1289 (57.6)	0
IABP (%)	430 (5.8)	15.7	607 (27.1)	0
≥ 1 Vasopressor (%)	7957 (90.3)	0	1725 (77.1)	0
≥ 1 Inotrope (%)	1357 (15.4)	0	1119 (50.0)	0
Total pressors within day 1 (%)		0		0
0	858 (9.7)		512 (22.9)	
1	5632 (63.9)		1165 (52.1)	
2	1650 (18.7)		357 (16.0)	
3	473 (5.4)		127 (5.7)	
4	140 (1.6)		60 (2.7)	

5	58 (0.7)		12 (0.5)	
6	3 (0.0)		3 (0.1)	
7	1 (0.0)		1 (0.0)	
Dobutamine (%)	557 (6.3)	0	179 (8.0)	0
Dopamine (%)	631 (7.2)	0	912 (40.8)	0
Epinephrine (%)	816 (9.3)	0	93 (4.2)	0
Milrinone (%)	300 (3.4)	0	151 (6.8)	0
Norepinephrine (%)	2298 (26.1)	0	576 (25.7)	0
Phenylephrine (%)	733 (8.3)	0	547 (24.5)	0
Vasopressin (%)	708 (8.0)	0	127 (5.7)	0
Total Pressors within first hour(%)		31.1		0
0	3452 (56.9)		1539 (68.8)	
1	2024 (33.3)		587 (26.2)	
2	475 (7.8)		95 (4.2)	
3	87 (1.4)		12 (0.5)	
4	24 (0.4)		4 (0.2)	
5	9 (0.1)		-	
Vital Signs				
Heart Rate (min) (mean (SD))	62.70 (14.83)	5.7	66.17 (16.71)	0.6
Heart Rate (max) (mean (SD))	115.46 (26.38)	5.7	108.13 (24.04)	0.6
Heart Rate (mean) (mean (SD))	84.59 (14.20)	5.7	84.20 (16.33)	0.6
Systolic BP (min) (mean (SD))	85.78 (20.36)	8.4	78.06 (16.89)	0.6
Systolic BP (max) (mean (SD))	158.38 (28.58)	8.4	145.31 (25.39)	0.6
Systolic BP (mean) (mean (SD))	119.44 (17.04)	8.4	109.24 (14.90)	0.6
Diastolic BP (min) (mean (SD))	43.95 (13.29)	8.5	37.60 (11.32)	0.7
Diastolic BP (max) (mean (SD))	95.72 (22.13)	8.5	83.31 (16.89)	0.7
Diastolic BP (mean) (mean (SD))	65.15 (10.13)	8.5	57.41 (10.06)	0.7
MAP (min) (mean (SD))	58.01 (16.07)	15	51.04 (13.99)	0.6
MAP (max) (mean (SD))	111.62 (23.47)	15	106.67 (29.22)	0.6
MAP (mean) (mean (SD))	81.19 (12.43)	15	74.90 (10.66)	0.6
Respiratory Rate (min) (mean (SD))	10.19 (5.35)	10.1	12.02 (3.69)	0.6
Respiratory Rate (max) (mean (SD))	32.78 (9.57)	10.1	28.67 (7.14)	0.6
Respiratory Rate (mean) (mean (SD))	19.67 (3.58)	10.1	19.21 (3.92)	0.6
Temperature (min) (mean (SD))	35.79 (1.14)	1.7	35.84 (0.98)	2.6
Temperature (max) (mean (SD))	37.68 (0.85)	1.7	37.56 (1.01)	2.6
Temperature (mean) (mean (SD))	36.76 (0.57)	1.7	36.75 (0.82)	2.6
SpO2 (min) (mean (SD))	84.15 (16.69)	9.3	88.72 (12.01)	1
SpO2 (max) (mean (SD))	99.64 (1.15)	9.3	99.63 (1.50)	1
SpO2 (mean) (mean (SD))	96.51 (2.57)	9.3	96.83 (3.39)	1
Urine Output within 24 hours (mean (SD))	689.86 (828.15)	29.5	2011.47 (1428.94)	5.1
First GCS score (mean (SD))	12.40 (4.03)	24.8	14.61 (1.59)	0.9
Laboratory Results				
Glucose (min) (mean (SD))	119.38 (49.08)	3.2	110.28 (42.07)	1.9
Glucose (max) (mean (SD))	193.48 (122.88)	3.2	215.82 (108.43)	1.9
Anion Gap (min) (mean (SD))	9.79 (4.60)	17.1	13.09 (3.44)	0.9
Anion Gap (max) (mean (SD))	13.54 (6.30)	17.1	18.44 (5.34)	0.9
Albumin (min) (mean (SD))	3.06 (0.73)	30.5	3.20 (0.65)	40.4
Albumin (max) (mean (SD))	3.30 (0.68)	30.5	3.36 (0.59)	40.4
Bands (min) (mean (SD))	10.65 (11.48)	91	6.96 (7.77)	81.9
Bands (max)(mean (SD))	13.96 (13.80)	91	9.65 (10.13)	81.9
Bicarbonate (min) (mean (SD))	22.00 (5.32)	5.7	20.55 (5.18)	0.8
Bicarbonate (max) (mean (SD))	25.58 (4.71)	5.7	25.77 (4.94)	0.8
Bilirubin (min) (mean (SD))	0.90 (1.44)	32.3	0.95 (2.37)	30.3
Bilirubin (max) (mean (SD))	1.08 (1.81)	32.3	1.32 (3.43)	30.3
Creatinine (min) (mean (SD))	1.56 (1.56)	3.2	1.36 (1.12)	0.3
Creatinine (max) (mean (SD))	2.02 (2.02)	3.2	1.97 (1.70)	0.3
Creatinine >1.5x baseline (%)	1820 (21.3)	3.2	762 (34.2)	0.3
Chloride (min) (mean (SD))	101.21 (6.72)	3.2	100.31 (6.31)	0.6
Chloride (max) (mean (SD))	105.96 (6.60)	3.2	106.95 (6.20)	0.6
Hematocrit (min) (mean (SD))	31.51 (7.29)	1.8	29.95 (6.21)	0.1
Hematocrit (max) (mean (SD))	37.11 (6.63)	1.8	38.13 (5.44)	0.1
Hemoglobin (min) (mean (SD))	10.50 (2.40)	1.9	10.16 (2.16)	0.4
Hemoglobin (max) (mean (SD))	12.23 (2.31)	1.9	12.67 (1.91)	0.4
Lactate (min) (mean (SD))	2.25 (2.52)	55.9	2.17 (2.05)	34.6

Lactate (max) (mean (SD))	4.04 (3.91)	55.9	4.37 (3.73)	34.6
Platelet (min) (mean (SD))	180.05 (90.54)	3.5	189.74 (85.52)	0.3
Platelet (max) (mean (SD))	226.56 (103.54)	3.5	268.25 (111.74)	0.3
Potassium (min) (mean (SD))	3.81 (0.60)	1.1	3.64 (0.56)	0.1
Potassium (max) (mean (SD))	4.67 (0.85)	1.1	5.03 (0.98)	0.1
PTT (min) (mean (SD))	32.85 (12.21)	35.8	34.04 (15.87)	4.2
PTT(max) (mean (SD))	44.00 (24.28)	35.8	73.81 (45.14)	4.2
INT (min) (mean (SD))	1.34 (0.62)	23.7	1.38 (0.57)	3.9
INR (max) (mean (SD))	1.60 (1.08)	23.7	2.03 (1.95)	3.9
PT (min) (mean (SD))	14.95 (6.34)	23.9	14.99 (4.64)	4.1
PT (max) (mean (SD))	17.50 (10.58)	23.9	18.98 (11.67)	4.1
Sodium (min) (mean (SD))	135.54 (5.30)	3.2	134.96 (5.08)	0.4
Sodium (max) (mean (SD))	139.66 (5.16)	3.2	140.58 (4.59)	0.4
BUN (min) (mean (SD))	24.89 (19.09)	3.2	26.98 (19.47)	0.4
BUN (max) (mean (SD))	31.79 (23.68)	3.2	38.47 (25.92)	0.4
WBC (min) (mean (SD))	10.40 (7.44)	2.2	9.95 (5.51)	0.5
WBC (max) (mean (SD))	14.66 (10.66)	2.2	15.68 (9.12)	0.5
Troponin I (min) (mean (SD))	4.18 (18.97)	57.1	7.87 (11.50)	92
Troponin I (max) (mean (SD))	11.36 (44.62)	57.1	13.26 (15.43)	92
RDW (min) (mean (SD))	15.06 (2.29)	6.1	14.61 (1.90)	0.7
RDW (max) (mean (SD))	15.49 (2.48)	6.1	15.23 (2.21)	0.7
Shock Index (mean (SD))	1.06 (0.39)	8.5	0.79 (0.21)	0.7
Outcomes				
In-hospital Mortality (%)	1188 (13.5)	0	509 (22.8)	0
ICU length of stay (days) (mean (SD))	1.52 (2.11)	0	5.55 (6.91)	0
Hospital length of stay (days) (mean (SD))	8.92 (9.59)	0	10.85 (11.75)	0
Time to death (days) (mean (SD))	5.39 (6.98)	0	8.90 (11.55)	0

SD=standard deviation, BMI=body mass index, AIDS/HIV=Acute Immunodeficiency Syndrome/Human Immunodeficiency Virus, COPD=chronic obstructive pulmonary disease, CAD=coronary artery disease, MI=myocardial infarction, IABP=intra-aortic balloon pump, BP=blood pressure, MAP=mean arterial pressure, SpO2=oxygen saturation, GCS=Glasgow Coma Scale, PTT=partial thromboplastin time, INR=international normalized ratio, PT=prothrombin time, BUN=blood urea nitrogen, WBC=white blood cell, RDW=red cell distribution width, ICU=intensive care unit

Table S2. Key reporting elements for machine learning analyses.

Study Design

1. **Clinical question:** Can a simple, clinically interpretable risk-score improve mortality prediction among patients with cardiogenic shock in the cardiac ICU?
2. **Intended use of results:** Build a simple, clinically interpretable risk-score that can be used to risk stratify patients with cardiogenic shock can provide important prognostic information and guide the appropriate triage and selection of therapies.
3. **Problem type:** Predictive classification model based on an unknown number of features present in two large real-world electronic medical record datasets
4. **Available data:**
    a. Philips eICU database (eICU-CRD v2.0): 200,859 patient encounters for 139,367 unique patients admitted between 2014 and 2015 in one of 335 units in 208 hospitals located throughout the US
    b. MIMIC-III: 61,532 adult hospital admissions for 53,423 distinct patients admitted to critical care units between 2001 and 2012 at the Beth Israel Deaconess Medical Center
5. **ML method and rationale:** RiskSLIM uses modern optimization techniques to fit the best logistic regression model with small integer weights and a limited number of risk factors. This technique can fit risk scores that have better risk-calibration and area under the curve compared to models developed heuristically (e.g., by combining logistic regression with techniques for feature selection and continuous variable dichotomization). The gain in performance stems from how RiskSLIM fits models in a single step without relying on approximations or heuristics. In this application, the model was constrained to use unit weights to allow for quick computation at the bedside as a checklist.
6. **Evaluation measures, training protocols, and validation**
    a. We evaluated all models by rank accuracy and risk calibration.
        i. We assessed rank accuracy via the area under the receiver operating characteristic curve (AUC).
        ii. We assessed risk calibration by constructing a reliability diagram plotting the observed mortality compared with the predicted mortality and by reporting the expected calibration error (ECE).
    b. The Philips eICU database was used for training given patients from a more heterogenous population.
    c. We evaluated the performance of each model internally, validating the performance of all models using 5-fold cross validation.
    d. The MIMIC-III database was used for external validation.
    e. Platt scaling was employed on the final models to improve reliability of estimates.

Data sources and preprocessing

1. **Population:**
    a. Philips eICU database (eICU-CRD v2.0): 200,859 patient encounters for 139,367 unique patients admitted between 2014 and 2015 in one of 335 units in 208 hospitals located throughout the US
    b. MIMIC-III: 61,532 adult hospital admissions for 53,423 distinct patients admitted to critical care units between 2001 and 2012 at the Beth Israel Deaconess Medical Center
2. **Sample record and measurement characteristics:** All structured data from ICU electronic medical records
    a. Contains demographics recorded from administrative data
    b. Comorbidities and CICU therapies recorded by care providers
    c. Vital signs recorded by nurses
    d. Core clinical laboratory panels
3. **Data collection and quality:**
    a. Data was collected as part of routine clinical care in the ICU
    b. Very low percentage missing across most variables (Table S1)
4. **Data structure and types**
    a. Categorical: demographics, comorbidities, CICU therapies
    b. Quantitative: vital signs, laboratory values -> summarized into categorical variables using cutoffs based on clinical judgment with multiple possible cutoffs of variables included as distinct candidate features

5. **Differences between evaluation and validation sets:** Please see Table 1 and Table S1
6. **Data preprocessing**
    a. Data subsets or aggregation: We restricted the observations to the values available within the first 24 hours of ICU admissions. For variables recorded multiple times over that timespan in the database, we kept its minimal and maximal value, and its average.
    b. Missing data: All candidate variables for which > 25% values were missing were excluded. For variables with < 25% missing values, we used simple imputation using predictive mean matching applied independently on the training and the validation datasets.
    c. Data transformation: We did not conduct any data transformation.
    d. Data label source: The primary outcome of our study, in-hospital mortality was directly recorded in the database and did not require any data processing or human input.
7. **Link to data or data request mechanism**
    a. Philips eICU database can be requested at: https://eicu-crd.mit.edu/
    b. MIMIC-III database can be requested at: https://mimic.mit.edu/

Model development and validation

1. **Hardware, software, and packages used**
    a. We fit RiskSLIM models using the RiskSLIM Python package, which is freely available at https://github.com/ustunb/risk-slim. This package uses the CPLEX 12.10 MIP solver, which is freely available to academic use through the IBM Academic Initiative. We fit each model for at most 20 minutes on a 3.33GHZ single-core CPU with 16 GB of RAM.
    b. We fit penalized logistic regression models using the glmnet package in R as a comparator.
2. **Model training and evaluation**
    a. Training completed using eICU database
    b. We created RiskSLIM models with successively increasing parameters (from 1 to 10) and examined performance via AUC and calibration error.
    c. Each RiskSLIM model was fit to optimize the logistic loss over a family of risk scores with model size and coefficient constraint. This model corresponds to risk score that attains the maximum possible calibration within the family of models – under the parametric assumption, true risk can be modeled using a logistic link function.
3. **Model parameters/hyperparameters:**
    a. For RiskSLIM models, we used a "checklist" style model where coefficients were restricted to 1 (count condition); 0 (do not consider condition); and -1 (count absence of condition)
    b. We fit penalized logistic regression models using the glmnet package in R. The free parameters for this model include: $\alpha \in [0, 1]$ (the elastic-net mixing parameter) and $\gamma \geq 0$ is a regularization penalty. We trained 1,100 PLR models by choosing 1,100 combinations of (α,γ): 11 values of alpha in 0.1,0.2,...1.0} x 100 values of γ (chosen automatically by glmnet for each α). This free parameter grid produces 1,100 PLR models that include models obtained by: (i) Lasso (l1-penalty), which corresponds to PLR when $\alpha = 1.0$; (ii) Ridge (l2-penalty), which corresponds to PLR when $\alpha = 0.0$; (iii) standard logistic regression, which corresponds to PLR when $\alpha = 0.0$ and γ is small.
4. **Features selected and input into the model (all binary 0/1)**
    a. maximum BUN ≥25
    b. minimum Oxygen saturation < 88
    c. minimum Systolic blood pressure <80
    d. Mechanical ventilation
    e. Age≥60
    f. maximum Anion gap ≥14
5. **Validation method and performance metrics**
    a. MIMIC-III database was used for external validation
    b. Training set (eICU): AUC 0.83 (0.82-0.84), ECE 0.9%
    c. Validation set (MIMIC III): AUC 0.76 (0.73-0.78), ECE 2.6%
6. **Reproducibility and code reuse**
    a. The computer code used to generate the analyses is available on GitHub (https://github.com/ustunb/cshock) including preprocessing and modeling steps.

Table S3. **Baseline characteristics of eICU and MIMIC cohorts.**

	eICU - Training Cohort (n=8815)	MIMIC III - Validation Cohort (n=2237)	Standardized Difference
Demographics			
Age (mean (SD))	64.07 (14.60)	69.79 (13.78)	0.391
Male sex (%)	5234 (59.4)	1314 (58.7)	0.013
Comorbidities			
Acute cerebral vascular disease (%)	982 (11.1)	87 (3.9)	0.278
Anemia (%)	51 (0.6)	623 (27.8)	0.848
Atrial Fibrillation (%)	1289 (14.6)	911 (40.7)	0.61
Blood Malignancy (%)	136 (1.5)	63 (2.8)	0.087
Solid Neoplasm (%)	1068 (12.1)	116 (5.2)	0.248
Congestive Heart Failure (%)	2026 (23.0)	1290 (57.7)	0.756
Chronic Kidney Disease (%)	1584 (18.0)	373 (16.7)	0.034
COPD (%)	1458 (16.5)	345 (15.4)	0.031
CAD (%)	2386 (27.1)	1303 (58.2)	0.664
Diabetes Mellitus (%)	2984 (33.9)	798 (35.7)	0.038
Valvulopathy (%)	922 (10.5)	548 (24.5)	0.376
Hypertension (%)	5496 (62.3)	1319 (59.0)	0.069
Metastatic Cancer (%)	125 (1.4)	42 (1.9)	0.036
Prior MI (%)	1347 (15.3)	1054 (47%)	0.732
CICU therapies			
Renal Replacement Therapy (%)	459 (5.2)	92 (4.1)	0.052
Mechanical Ventilation (%)	2771 (31.4)	1289 (57.6)	0.546
IABP (%)	430 (5.8)	607 (27.1)	0.631
≥ 1 Vasopressor (%)	7957 (90.3)	1725 (77.1)	0.362
≥ 1 Inotrope (%)	1357 (15.4)	1119 (50.0)	0.794
Vital signs			
Heart Rate Min (mean (SD))	62.70 (14.83)	66.17 (16.71)	0.216
Heart Rate Max (mean (SD))	115.46 (26.38)	108.13 (24.04)	0.288
Systolic BP Min (mean (SD))	85.78 (20.36)	78.06 (16.89)	0.024
Systolic BP Max (mean (SD))	158.38 (28.58)	145.31 (25.39)	0.487
Respiratory Rate Min (mean (SD))	10.19 (5.35)	12.02 (3.69)	0.409
Respiratory Rate Max (mean (SD))	32.78 (9.57)	28.67 (7.14)	0.492
SpO2 Min (mean (SD))	84.15 (16.69)	88.72 (12.01)	0.32
Laboratory results			
Glucose Min (mean (SD))	119.38 (49.08)	110.28 (42.07)	0.188
Anion Gap Max (mean (SD))	13.54 (6.30)	18.44 (5.34)	0.83
Bicarbonate Min (mean (SD))	22.00 (5.32)	20.55 (5.18)	0.292
Chloride Max (mean (SD))	105.96 (6.60)	106.95 (6.20)	0.15
Hematocrit Max (mean (SD))	37.11 (6.63)	38.13 (5.44)	0.168
Hemoglobin Min (mean (SD))	10.50 (2.40)	10.16 (2.16)	0.157
Platelet Min (mean (SD))	180.05 (90.54)	189.74 (85.52)	0.106
Potassium Max (mean (SD))	4.67 (0.85)	5.03 (0.98)	0.4
INR Max (mean (SD))	1.60 (1.08)	2.03 (1.95)	0.275
Sodium Min (mean (SD))	135.54 (5.30)	134.96 (5.08)	0.116
BUN Max (mean (SD))	31.79 (23.68)	38.47 (25.92)	0.274
WBC Max (mean (SD))	14.66 (10.66)	15.68 (9.12)	0.105
RDW Max (mean (SD))	15.49 (2.48)	15.23 (2.21)	0.113
Creatinine >1.5x baseline (%)	1820 (21.3)	762 (34.2)	0.298
Outcome			
In-hospital Mortality (%)	1188 (13.5)	509 (22.8)	0.243
ICU length of stay (days) (mean (SD))	1.52 (2.11)	5.55 (6.91)	0.789
Hospital length of stay (days) (mean (SD))	8.92 (9.59)	10.85 (11.75)	0.179
Time to death (days) (mean (SD))	5.39 (6.98)	8.90 (11.55)	0.368

SD=standard deviation, COPD=chronic obstructive pulmonary disease, CAD=coronary artery disease, MI=myocardial infarction, IABP=intra-aortic balloon pump, SpO2=oxygen saturation, INR=international normalized ratio, BUN=blood urea nitrogen, WBC=white blood cell, RDW=red cell distribution width

Table S4. Performance of BOS,MA$_2$ risk score model compared to penalized logistic regression models.

Model	eICU (Training)	MIMIC (Validation)
Penalized logistic regression model with all features		
AUC	0.87 (0.86, 0.88)	0.80 (0.78, 0.82)
Calibration error	0.50%	4.00%
Penalized logistic regression model with same features		
AUC	0.84 (0.83, 0.85)	0.76 (0.73, 0.78)
Calibration error	0.50%	3.40%
BOS,MA$_2$ model (RiskSLIM)		
AUC	0.83 (0.82, 0.84)	0.75 (0.73, 0.78)
Calibration error	0.90%	2.60%
Penalized logistic regression model with all features + BOS,MA$_2$		
AUC	0.86 (0.85, 0.87)	0.79 (0.77, 0.81)
Penalized logistic regression model with same features + BOS,MA$_2$		
AUC	0.84 (0.83, 0.85)	0.76 (0.74, 0.78)

Table S5. **BOS,MA$_2$ risk score subgroup analysis.**

Subgroup	Subgroup size	Positive cases	AUC*	ECE (%)
Presence of a primary cardiovascular admission diagnosis	305	69	0.78 (0.73-0.83)	1.05
Presence of a primary admission diagnosis other than CV	383	101	0.81 (0.77-0.86)	1.00
Presence of MI necessitating acute coronary revascularization	208	42	0.83 (0.77-0.88)	1.06
Absence of MI necessitating acute coronary revascularization	478	128	0.79 (0.75-0.83)	1.00

*95 CI calculated using DeLong statistics

Table S6. BOS,MA_2 risk score model performance compared to other validated risk scores in cardiovascular-specific subgroups of the validation cohort.

Cohort	Presence of a primary cardiovascular admission diagnosis (n=305)		Presence of myocardial infarction necessitating acute coronary revascularization (n=208)	
Risk score	AUC (95% CI) [a]	ECE	AUC (95% CI) [a]	ECE
BOS,MA_2	0.78 (0.73-0.83)	1.05	0.83 (0.77-0.88)	1.06
SOFA score (day 1)	0.76 (0.70-0.82)	4%	0.83 (0.77-0.89)	7.7%
OASIS score (day 1)	0.79 (0.74-0.85)	8.1%	0.82 (0.76-0.89)	5.7%
AUC=area under the curve; CI=confidence interval; ECE=expected calibration error [a] 95% CI computed using the DeLong statistic				

Figure S1. Candidate model performance based on model size.

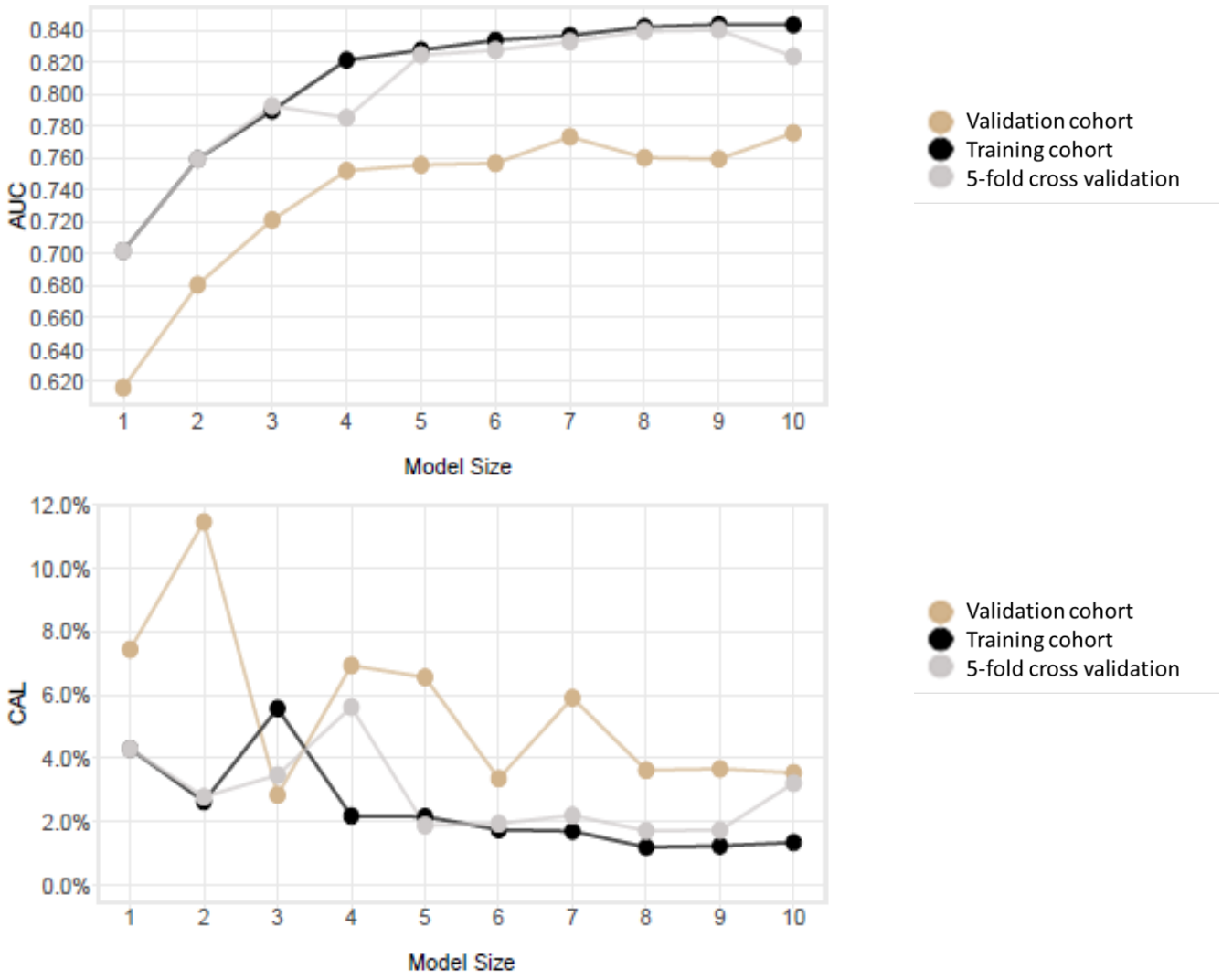

Figure S2. Observed BOS,MA2 risk score values across datasets.