

Preintubation Sequential Organ Failure Assessment Score for Predicting COVID-19 Mortality: External Validation Using Electronic Health Record From 86 U.S. Healthcare Systems to Appraise Current Ventilator Triage Algorithms*

OBJECTIVES: Prior research has hypothesized the Sequential Organ Failure Assessment (SOFA) score to be a poor predictor of mortality in mechanically ventilated patients with COVID-19. Yet, several U.S. states have proposed SOFA-based algorithms for ventilator triage during crisis standards of care. Using a large cohort of mechanically ventilated patients with COVID-19, we externally validated the predictive capacity of the preintubation SOFA score for mortality prediction with and without other commonly used algorithm elements.

DESIGN: Multicenter, retrospective cohort study using electronic health record data.

SETTING: Eighty-six U.S. health systems.

PATIENTS: Patients with COVID-19 hospitalized between January 1, 2020, and February 14, 2021, and subsequently initiated on mechanical ventilation.

INTERVENTIONS: None.

MEASUREMENTS AND MAIN RESULTS: Among 15,122 mechanically ventilated patients with COVID-19, SOFA score alone demonstrated poor discriminant accuracy for inhospital mortality in mechanically ventilated patients using the validation cohort (area under the receiver operating characteristic curve [AUC], 0.66; 95% CI, 0.65–0.67). Discriminant accuracy was even poorer using SOFA score categories (AUC, 0.54; 95% CI, 0.54–0.55). Age alone demonstrated greater discriminant accuracy for inhospital mortality than SOFA score (AUC, 0.71; 95% CI, 0.69–0.72). Discriminant accuracy for mortality improved upon addition of age to the continuous SOFA score (AUC, 0.74; 95% CI, 0.73–0.76) and categorized SOFA score (AUC, 0.72; 95% CI, 0.71–0.73) models, respectively. The addition of comorbidities did not substantially increase model discrimination. Of 36 U.S. states with crisis standards of care guidelines containing ventilator triage algorithms, 31 (86%) feature the SOFA score. Of these, 25 (81%) rely heavily on the SOFA score (12 exclusively propose SOFA; 13 place highest weight on SOFA or propose SOFA with one other variable).

CONCLUSIONS: In a U.S. cohort of over 15,000 ventilated patients with COVID-19, the SOFA score displayed poor predictive accuracy for short-term mortality. Our findings warrant reappraisal of the SOFA score's implementation and weightage in existing ventilator triage pathways in current U.S. crisis standards of care guidelines.

KEY WORDS: COVID-19; intubation; mechanical ventilation; mortality; predictive models; sequential organ failure assessment

Michael B. Keller, MD^{1,2}

Jing Wang, MS³

Martha Nason, PhD⁴

Sarah Warner, PhD¹

Dean Follmann, PhD⁴

Sameer S. Kadri, MD, MS¹

The COVID-19 pandemic has caused surges in hospital caseloads worldwide, placing strain on affected healthcare systems (1). Patient caseloads have often exceeded a hospital's capacity to provide standard-of-care,

*See also p. 1150.

Copyright © 2022 by the Society of Critical Care Medicine and Wolters Kluwer Health, Inc. All Rights Reserved.

DOI: 10.1097/CCM.0000000000005534

necessitating contingency standards, and in extreme situations, crisis standards of care (CSC). The latter may result in scenarios, whereby parsimonious allocation of life-saving resources becomes pivotal (2). Methods to adequately predict and maximize survival are paramount to inform CSC triage guidelines. Several guidelines have been developed to guide resource allocation under such circumstances (3–8). In addition to elements intended to predict survival, these guidelines include components intended to predict survival and also identify those at risk for high resource consumption such as from poor functional outcomes or prolonged mechanical ventilation. However, there is considerable variation in elements included in ventilator triage algorithms across state CSC guidelines as well as the quality of evidence underpinning their inclusion, raising ethical concerns around the adequacy of ventilator allocation offered by current algorithms (9).

Many mechanical ventilator triage protocols in the United States include the Sequential Organ Failure Assessment (SOFA) score to predict short-term survival (9–11), including two U.S. states that recently declared CSC due to COVID-19 surges (12, 13). However, the degree to which ventilator triage decisions would hinge on the score has received less attention. Two prior studies in cohorts of ICU patients with sepsis have reported an area under the receiver operating characteristic curve (AUC) of 0.74 and 0.75 of the SOFA score for predicting survival (14, 15). The SOFA score assigns equal weightage to its six organ system components; however, respiratory failure tends to be the predominant organ failure among acutely ill patients with COVID-19, and these patients display less variability in SOFA score than those with conditions such as bacterial sepsis (16). Hence, despite its inclusion in several triage protocols nationwide, it is unclear whether the SOFA score adequately predicts mortality in mechanically ventilated patients with COVID-19.

A recent hypothesis-generating study suggests that the discriminant accuracy of the SOFA score for predicting inhospital mortality in mechanically ventilated COVID-19 patients is poor (16). However, the study was relatively small (675 ventilated patients), was regional, and did not assess for model calibration or the predictive capacity of combining SOFA with other relevant predictors featured in existing triage protocols. As suggested by a recent expert consensus panel, there

is need for additional, larger studies to validate the predictive accuracy of existing algorithms and formulate better prediction tools (17). Hence, in this study, we: 1) examine implementation and weightage of the SOFA score in State CSC ventilator triage algorithms nationally and 2) leverage a large electronic health record (EHR) database of U.S. hospitals to externally validate the hypothesis that preintubation SOFA score is a poor predictor for inhospital mortality in COVID-19 patients requiring mechanical ventilation.

MATERIALS AND METHODS

Study Design and Data Source

We performed a multicenter, retrospective cohort study using the Cerner COVID-19 Deidentified Data cohort. This repository contains EHR data from 86 U.S. healthcare systems that share data with Cerner (Kansas City, MO) and includes billing records, medication orders, laboratory results, vitals, and other physiologic variables (18). Data were accessed and analyzed on Cerner HealthIntent (Cerner), a cloud-based management platform following a data use agreement (no. 1-70WNSGX) between the National Institutes of Health (NIH) and Cerner. Data refreshes were provided quarterly, enabling incorporation of new cases. Downstream curation of study-specific variables and algorithms was performed by NIH-contracted informaticists under the guidance of study investigators (M.K., S.S.K.) and study design feedback offered by all investigators. Given the deidentified nature of the data, the study was deemed exempt from ethics board review based on the policy of the NIH Office of Human Subjects Research Protections.

Study Population

Patients greater than or equal to 18 years old with COVID-19 admitted as inpatients between January 1, 2020, and February 14, 2021, who underwent mechanical ventilation were included. For each patient, one admission was randomly selected for inclusion in the analysis. Patients admitted with COVID-19 were identified by an *International Classification of Diseases*, 10th Edition (ICD-10) diagnosis code for COVID-19 (U07.1), a positive polymerase chain reaction (PCR) test for severe acute respiratory syndrome coronavirus 2 (SARS-CoV2), or positive serology for COVID-19

antibodies. The ICD-10 diagnosis code for COVID-19 captures patients positive for SARS-CoV2 on PCR with a sensitivity of 98%, specificity of 99%, and a positive predictive value of 92% (19). Encounters prior to March 2020 were identified using a legacy coding strategy that leverages coding for generic coronaviruses (B97.29) (19). Patients who received invasive mechanical ventilation were identified by ICD-10 invasive mechanical ventilation procedure codes and Logical Observation Identifiers Names and Codes. Patients on mechanical ventilation within 24 hours of admission and those with a designation of do-not-resuscitate (DNR) status present at admission, respectively, were excluded.

Study Variables

The primary outcome was inhospital mortality, defined as death during hospitalization or discharge to hospice.

The highest SOFA score was calculated within 24 hours prior to initiation of mechanical ventilation, signifying a time point at which ventilator triage is likely to occur based on current CSC protocols (3, 5, 7). Cerner HealthIntent contains all components necessary to compute the SOFA score except urine output and vasopressor dose. Therefore, as previously described, we used creatinine levels to assign points for renal dysfunction and the number of vasopressors to assign points for cardiovascular dysfunction (20). If values for P_{O_2} in arterial blood to fractional concentration of oxygen ratio ($P_{aO_2}:F_{iO_2}$) were missing, we used the saturation of blood oxygen to F_{iO_2} ratio ($S_{aO_2}:F_{iO_2}$) to assign points for respiratory dysfunction (21). Daily SOFA score was computed using the worst scoring criteria for each component on each day. If no values were available on a day, we used the closest value within 5 days looking backward (20). If there was no value within the prior 5 days, we assigned the SOFA score component as 0 (missing-as-normal), as previously described (15, 22, 23). For Glasgow Coma Scale (GCS) score, the lowest value was taken for a given day. We then carried that value forward until a new value was present on another day (20). If the first GCS score occurred several days into hospitalization, the GCS score was assigned as 0 each day leading up to that day. Further details regarding calculation of daily SOFA scores are in **eMethods (Online Supplement, <http://links.lww.com/CCM/H88>)**. We evaluated SOFA score as both a continuous (count) variable ranging from 0

to 24 and a categorized variable, based on strata (<6, 6–8, 9–11, >11) commonly implemented in existing CSC guidelines (3, 5, 7, 12, 13).

Select patient-level covariates were identified, including age, sex, obesity, diabetes, and hypertension, based on prior data linking these covariates with poor outcomes (24–29). Patient-level comorbidities were identified using respective ICD-10 codes. Aggregate comorbidity burden was assessed using the Elixhauser comorbidity index (30, 31).

Evaluation of U.S. State-Adopted CSC Guidelines

We next performed a cross-sectional analysis of state-adopted CSC guidelines to examine the prevalence of SOFA score utilization and degree of representation in current CSC models. One study investigator (M.K.) performed a search on three separate dates between October first, 2021, and October 14, 2021, for state-adopted CDC guidelines, providing guidance on triage of mechanical ventilation or scarce resources as previously described (32). State-adopted CSC guidelines were identified as those written by or in coordination with the state's department of public health. CSC guidelines that were revoked or not written in coordination with the state's department of health (33, 34) were excluded. Guidelines that directly mention COVID-19 or were written after March 1, 2020, were deemed "COVID-19 specific" (further details of the search methods are outlined in the eMethods, Online Supplement, <http://links.lww.com/CCM/H88>). We categorized each CSC guideline's level of reliance on the SOFA score as follows:

- 1) No reliance
- 2) Low reliance—SOFA score is mentioned but not directly involved in triage of mechanical ventilation
- 3) Heavy reliance—SOFA score used alone or as a major component (SOFA indicated as holding the greatest weight or used with one other variable) in assigning patients to priority tiers for receipt of mechanical ventilation

We calculated the prevalence of each category of SOFA score reliance for CSC state-adopted protocols in the United States.

Statistical Analysis

We calculated descriptive statistics of patient characteristics overall or by patient groups. All characteristics were reported at admission to hospital except for

preintubation SOFA score. To assess the difference between patient groups, Mann-Whitney nonparametric tests were used for continuous variables, and chi-square or Fisher exact tests were used for categorical variables.

To investigate preintubation SOFA score and/or age as predictors, inhospital mortality, with or without adjusting for other covariates, logistic regression, and conditional classification trees with Bonferroni adjustments (35) were fit using the derivation set, which is two thirds of the entire dataset. The other third of the data were saved to validate selected models. Derivation/validation cohort splitting was stratified by hospital.

To evaluate discriminant accuracy, we computed the AUC with 95% CIs and performed Delong test to compare the AUCs. We considered an AUC below 0.7 to be poor accuracy, AUC 0.7–0.8 moderate, 0.8–0.9 good, and greater than 0.9 excellent (36). We also generated calibration belts, followed by conducting Hosmer-Lemeshow test to assess calibration (37–39).

We conducted sensitivity analyses: 1) excluding patients with chronic kidney disease (CKD) and end-stage renal disease (ESRD) (as the renal score component of SOFA used creatinine rather than urine output, thus potentially effecting model performance), 2) excluding patients with missing SOFA score values after our substitution method (imputed as normal=0), 3) excluding patients who had an ICD-10 Major Operating Room procedure code on the same day as intubation (to account for potential preoperative rather than critical illness-related intubation), and 4) using tree-based instead of logistic regression models.

Among survivors, logistic regression was performed to investigate preintubation SOFA score as a predictor of discharge to long-term acute care (LTAC) (secondary outcome) facilities, as these patients may represent a population at risk for high resource utilization.

Analyses were conducted using R 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria). A *p* value of less than 0.05 was considered statistically significant. A link to statistical code is included in the **Online Supplement** (page 8, <http://links.lww.com/CCM/H88>).

RESULTS

Between January 1, 2020, and February 14, 2021, 101,985 patients (109,285 inpatient encounters) with COVID-19 were admitted to 86 U.S. healthcare

systems. Of those encounters without an ICD diagnosis code or positive PCR, only 930 (0.85%) were selected based on positive serology. Of the 101,985 patients, 24,908 patients (24%) were mechanically ventilated, but 9,043 patients were mechanically ventilated within 24 hours of admission and 743 patients were DNR at admission and excluded, leaving 15,122 patients in the final analysis divided into a derivation ($n = 10,085$) and a validation ($n = 5,037$) cohort (**Fig. 1**). Of 15,122 ventilated patients, 7,568 (50.0%) died or were discharged to hospice; among 7,554 survivors, 501 (6.6%) were discharged to LTAC. The mean preintubation SOFA score was 2.77 (SD, 1.91). Respiratory SOFA subscore had a mean of 1.39 compared with 0.53 for hepatic, 0.29 for cardiovascular, 0.24 for neurologic, 0.23 for coagulation, and 0.09 for renal. A density plot illustrating SOFA scores for re-admissions is presented in **Supplementary Figure 1** (<http://links.lww.com/CCM/H88>). A total of 7,568 patients (50%) died or were discharged to hospice. Patients who died or were discharged to hospice tended to be older and male, and display higher SOFA and Elixhauser scores, respectively (**Table 1**).

Discrimination of Mortality Risk by SOFA Score

Using logistic regression models, the SOFA score demonstrated poor discriminant accuracy for mortality in mechanically ventilated patients (AUC, 0.66; 95% CI, 0.64–0.67). Discriminant accuracy was even poorer using categorized SOFA scores (AUC, 0.54; 95% CI, 0.54–0.55) (**Table 2**) and SOFA as a dichotomous variable with score greater than 11 (AUC, 0.50; 95% CI, 0.50–0.50). Discrimination of respiratory SOFA subscore alone for mortality was also poor (AUC, 0.67; 95% CI, 0.65–0.68) as were each of the other SOFA subscores (**Supplementary Index Table 1a**, <http://links.lww.com/CCM/H88>). In addition, the discriminant accuracy for SOFA score at admission (AUC, 0.61; 95% CI, 0.60–0.62) was poor as was change in SOFA score from admission to intubation (AUC, 0.54; 95% CI, 0.52–0.55; **Supplementary Index Table 1b**, <http://links.lww.com/CCM/H88>). Results were similar across the derivation and validation cohorts.

Discrimination by Age and Comorbidities, Alone and in Combination With SOFA Score

Age alone demonstrated better discriminant accuracy for mortality than SOFA score (AUC, 0.71; 95%

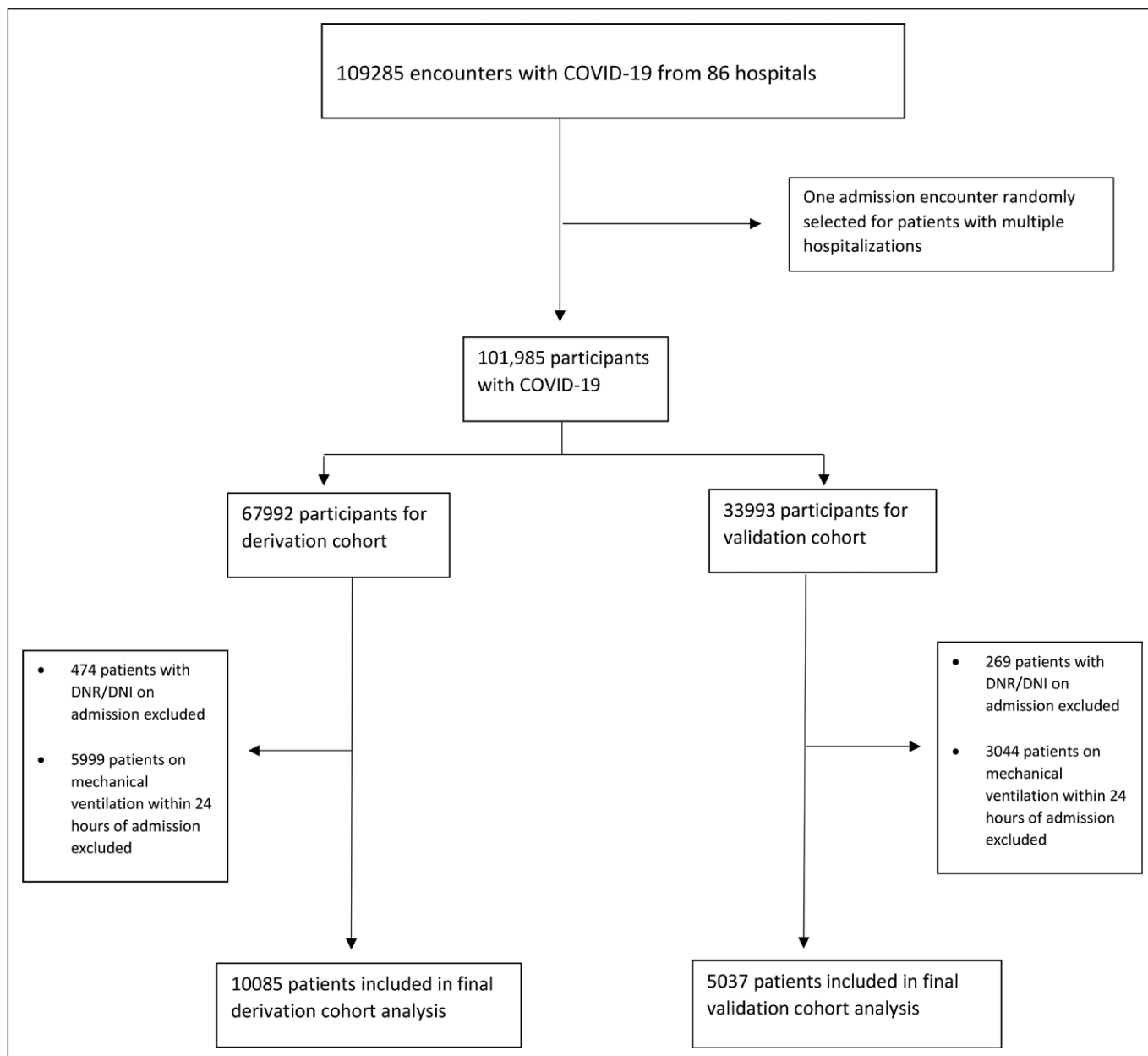


Figure 1. Study flowchart depicting the exclusion of patients with do-not-resuscitate (DNR)/do-not-intubate (DNI) status and mechanical ventilation on within 24 hr of admission; 15,122 patients were included in the final analysis.

CI, 0.69–0.72). Discriminant accuracy for mortality improved upon addition of age to the continuous SOFA score (AUC, 0.74; 95% CI, 0.73–0.76) and categorized SOFA score (AUC, 0.72; 95% CI, 0.71–0.73) models, respectively. The addition of other covariates (gender, obesity, diabetes, hypertension, or Elixhauser score) did not meaningfully improve discrimination beyond that offered by SOFA + age. Models without age, that is, utilizing SOFA and comorbidities (SOFA + Elixhauser score), had poor discriminant accuracy for both continuous SOFA (AUC, 0.66; 95% CI, 0.65–0.68) and categorized SOFA scores (AUC, 0.56; 95% CI, 0.55–0.58) (Table 2).

Sensitivity Analysis

Sensitivity analysis separately excluding patients with CKD and ESRD, patients with missing SOFA score values after our substitution method (imputed as normal—0), and those who had procedure codes for major operating room procedure and intubation on the same day all yielded results consistent with our primary analysis. Using tree-based models as a sensitivity analysis generated similar results, with slightly poorer discriminant accuracy in comparison with logistic regression models (**Supplementary Index Tables 2–5**, <http://links.lww.com/CCM/H88>).

TABLE 1.
Patient Characteristics

Characteristics	Overall (n = 15,122)	Alive (n = 7,554)	Dead (n = 7,568)	P
Age, mean (sd)	64.47 (15.28)	58.89 (15.74)	70.05 (12.54)	< 0.001
< 65 (%)	6,914 (46)	4,599 (61)	2,315 (31)	< 0.001
≥ 65 (%)	8,208 (54)	2,955 (39)	5,253 (69)	
Diabetes mellitus (%)	3,292 (22)	1,564 (21)	1,728 (23)	0.002
Hypertension (%)	3,813 (25)	1,854 (25)	1,959 (26)	0.06
Obesity (%)	1,455 (9.6)	908 (12)	547 (7.2)	< 0.001
Sex, female %	10,010 (40)	5,129 (41)	4,881 (39)	< 0.001
Preintubation SOFA score, mean (sd)	2.77 (1.91)	2.26 (1.63)	3.29 (2.03)	< 0.001
SOFA subscores, mean (sd)				
Respiratory	1.39 (0.82)	1.27 (0.86)	1.51 (0.76)	< 0.001
Coagulation	0.23 (0.56)	0.17 (0.47)	0.29 (0.64)	< 0.001
Hepatic	0.53 (1.01)	0.36 (0.86)	0.70 (1.11)	< 0.001
Cardiovascular	0.29 (0.57)	0.26 (0.55)	0.32 (0.60)	< 0.001
Neurologic	0.24 (0.61)	0.14 (0.46)	0.34 (0.72)	< 0.001
Renal	0.09 (0.39)	0.06 (0.30)	0.12 (0.46)	< 0.001
Elixhauser score mean (sd)	1.39 (1.65)	1.30 (1.60)	1.48 (1.70)	0.001

SOFA = Sequential Organ Failure Assessment.

Secondary Outcome

Using logistic regression in the validation cohort, pre-intubation SOFA displayed an AUC of 0.53 (0.49–0.57; odds ratio, 0.99 [0.93–1.04]) for predicting disposition to LTAC.

Model Calibration

The calibration belt for continuous SOFA score (Fig. 2A) demonstrated significant miscalibration, overestimating mortality risk for patients with observed mortality of 81–95%. The SOFA + age model was well calibrated (Fig. 2B). All other models also showed good calibration (Supplementary Fig. 2, <http://links.lww.com/CCM/H88>).

Nationwide Distribution of SOFA Score Implementation and Weightage in State CSC Ventilation Triage Algorithms

Our search revealed 36 states with state-adopted CSC guidelines in place (Fig. 3). Twenty-six of these guidelines are COVID-19-specific and 10 are adopted from prior influenza pandemics. The SOFA score features in 31/36 (86%) of these CSC guideline's triage protocols.

Of these 31 guidelines that feature SOFA score, 25/31 (81%) are heavily reliant on SOFA, with all utilizing the categorized SOFA score (12 of the protocols propose categorized SOFA as the only variable for ventilator triage). Six CSC guidelines place low reliance on SOFA score, excluding it as a main focus of their triage algorithm in place of greater emphasis on clinical judgment (Supplementary Index Table 6, <http://links.lww.com/CCM/H88>).

DISCUSSION

External validation of prediction models is a vital step that ensures models are reproducible, generalizable, and reliable for application in real-world decision-making. Using a cohort of over 15,000 mechanically ventilated COVID-19 patients at 86 U.S. healthcare systems, our study externally validates findings from a prior hypothesis-generating study of 675 ventilated COVID-19 patients and offers further confirmation that preintubation SOFA score is a poor predictor of inhospital mortality in COVID-19 patients. Despite a preponderance of respiratory failure in COVID-19, our study also found that the discriminant accuracy for inhospital mortality remains poor

TABLE 2.**Area Under the Receiver Operating Characteristic Curve for Prediction Models on Both Derivation and Validation Cohorts**

Model	Variable	AUC (95% CI), Derivation Cohort (n = 10,085)	AUC (95% CI), Validation Cohort (n = 5,037)	Logistic Regression, OR (95% CI) ^a
SOFA ^b	SOFA	0.66 (0.65–0.67)	0.66 (0.64–0.67)	1.39 (1.35–1.42)
SOFA categories ^c	≥ 6 and < 9	0.55 (0.54–0.55)	0.54 (0.54–0.55)	3.42 (2.89–4.06)
	≥ 9 and < 12			4.73 (3–7.82)
	≥ 12			5.89 (1.96–25.32)
Age	Age	0.71 (0.7–0.72)	0.71 (0.69–0.72)	1.06 (1.06–1.06)
Age + SOFA categories	Age	0.73 (0.72–0.74)	0.72 (0.71–0.73)	1.06 (1.05–1.06)
	≥ 6 and < 9			3.18 (2.67–3.82)
	≥ 9 and < 12			4.92 (3.05–8.3)
	≥ 12			6.66 (2.13–29.31)
Age + SOFA	Age	0.75 (0.74–0.76)	0.74 (0.73–0.76)	1.06 (1.05–1.06)
	SOFA			1.33 (1.3–1.36)
SOFA + age + covariates ^d	SOFA	0.75 (0.74–0.76)	0.74 (0.73–0.76)	1.33 (1.29–1.36)
	Age			1.06 (1.05–1.06)
	Gender (male vs female)			1.15 (1.05–1.25)
	Obesity			0.92 (0.79–1.08)
	Diabetes			1.18 (1.05–1.32)
	Hypertension			0.84 (0.75–0.93)
SOFA + age + Elixhauser score	Age	0.75 (0.74–0.76)	0.74 (0.73–0.76)	1.06 (1.05–1.06)
	Elixhauser score			1 (0.97–1.03)
	SOFA			1.33 (1.3–1.36)
SOFA + Elixhauser score	SOFA	0.66 (0.65–0.67)	0.66 (0.65–0.68)	1.38 (1.35–1.42)
	Elixhauser score			1.04 (1.01–1.07)
Categories SOFA + Elixhauser score	≥ 6 and < 9	0.57 (0.56–0.58)	0.56 (0.55–0.58)	3.33 (2.82–3.96)
	≥ 9 and < 12			4.59 (2.91–7.59)
	≥ 12			5.75 (1.91–24.76)
	Elixhauser score			1.06 (1.03–1.09)

AUC = area under the receiver operating curve, OR = odds ratio, SOFA = Sequential Organ Failure Assessment.

^aOutcome variable has two levels: deceased and discharged, where discharged is served as the reference level.

^bSOFA and all components are the scores recorded within the 24 hr prior to the start of ventilation.

^cVariable SOFA category is created based on SOFA variable with the cutoffs of: < 6, ≥ 6 and < 9, ≥ 9 and < 12, and ≥ 12. In the logistic regression model, SOFA < 6 is served as the reference group.

^dCovariates include age, gender, obesity, hypertension, and diabetes.

when considering respiratory subscore alone. The combination of SOFA score and age provided moderate mortality prediction; however, the addition to select common comorbidities or aggregate comorbidity burden did not meaningfully improve the predictive accuracy. Poor prediction of mortality was consistently observed across various parameterizations of the SOFA score (including as a continuous variable,

categorical variable, component score, score change, and across different time periods) and in multiple sensitivity analysis. Even among ventilated patients with COVID-19 who survived hospitalization, SOFA score was a poor predictor of requiring LTAC. Furthermore, we demonstrated that 81% of current state CSC triage algorithms using SOFA score propose heavily reliance on the categorized SOFA score to assign patients into

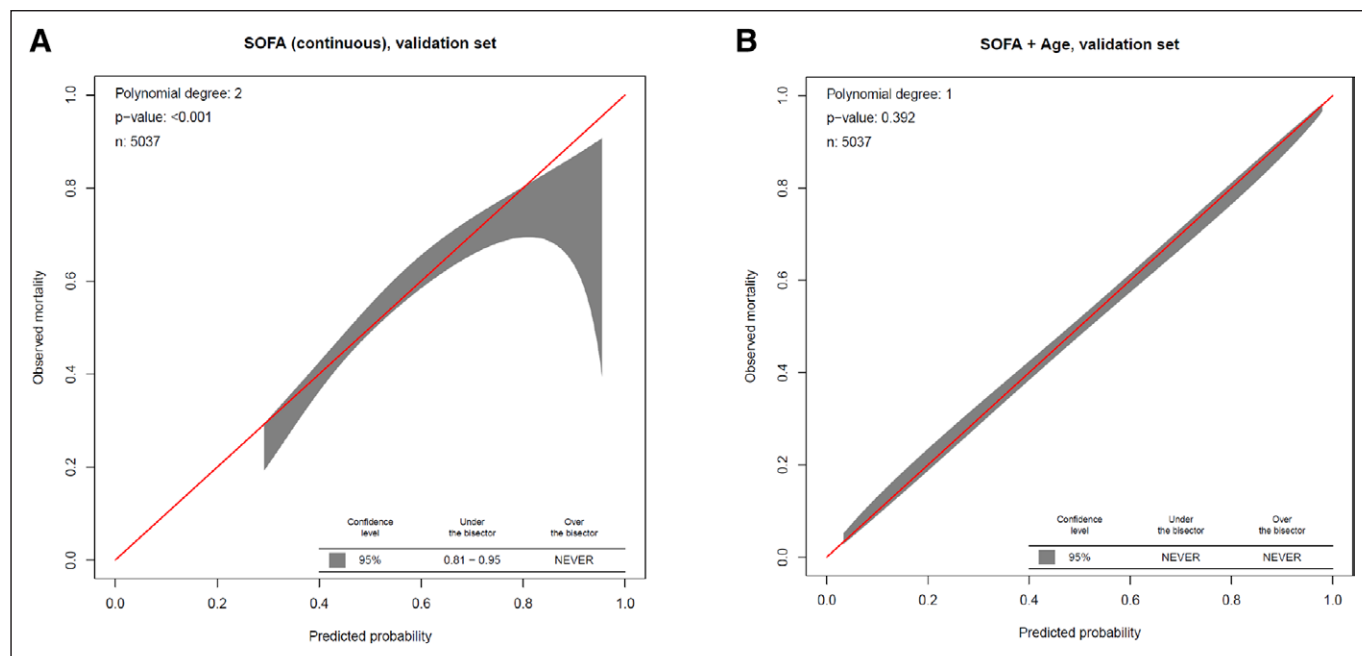


Figure 2. Calibration belts for mortality prediction scores. **A**, Continuous Sequential Organ Failure Assessment (SOFA) score. **B**, SOFA score + age. The range of values for which the predicted mortality overestimates mortality (the observed mortality values are significantly under the bisector) or underestimates mortality (observed mortality lies above the bisector) based on the shaded 95% CI is reported at the *bottom* of each graph.

priority tiers for receipt of mechanical ventilation despite its poor predictive accuracy (AUC, 0.54). Our findings build upon observations of previous studies by utilizing expansive, well harmonized EHR data from a large cohort of U.S. hospitals and suggest that significant reliance on the SOFA score in ventilator triage protocols warrants reappraisal (16).

The SOFA score, as acknowledged by its developers, was not intended to predict outcomes but to provide a quantitative description of the degree of organ dysfunction in critically ill patients (40, 41). As the degree of organ dysfunction is associated with mortality, several studies have validated its predictive accuracy for mortality in critically ill patients (14, 15, 42, 43). However, these populations may not be generalizable to patients with COVID-19, which primarily presents with respiratory failure and less variation in SOFA scores (16). In addition, many of these studies used distinct SOFA criteria often not represented in triage protocols, such as a change in SOFA score, that were assessed at snapshots in time (not necessarily at the point of triage). A recent study evaluating the predictive accuracy for in-hospital mortality of the SOFA score in patients with sepsis and acute respiratory failure prior to the COVID-19 pandemic also revealed poor discriminant accuracy (44). Concerningly, this study also found that using

the SOFA score for prognostic evaluation may lead to racial disparities in resource allocation by overestimating mortality for Black patients, potentially diverting scarce resources from this population without warrant.

We recognize the complexity that underlies decisions around continued use of current CSC models. We also recognize that these decisions are often not solely predicated on mortality but take into consideration other outcomes including poor long-term functional status and risk of high resource utilization. Nonetheless, given the limitations and uncertainty surrounding the adequacy of the SOFA score for ventilator triage decisions, more robust and reliable strategies for prognostication are needed. Although there has yet to be significant ventilator triage performed in United States at this point in time, even in overwhelmed hospital systems, with the recent surge in the Omicron Variant, hospitalizations are currently the highest they have been during the pandemic, and ICUs are nearing capacity in several U.S. states (45). As such, we must be adequately prepared for the possibility of needing ventilator triage in the future, especially if we are met with a variant with high transmissibility and a propensity for both evading immunity and causing severe disease. Twelve U.S. States rely exclusively on categorical SOFA score for ventilator triage. Our findings

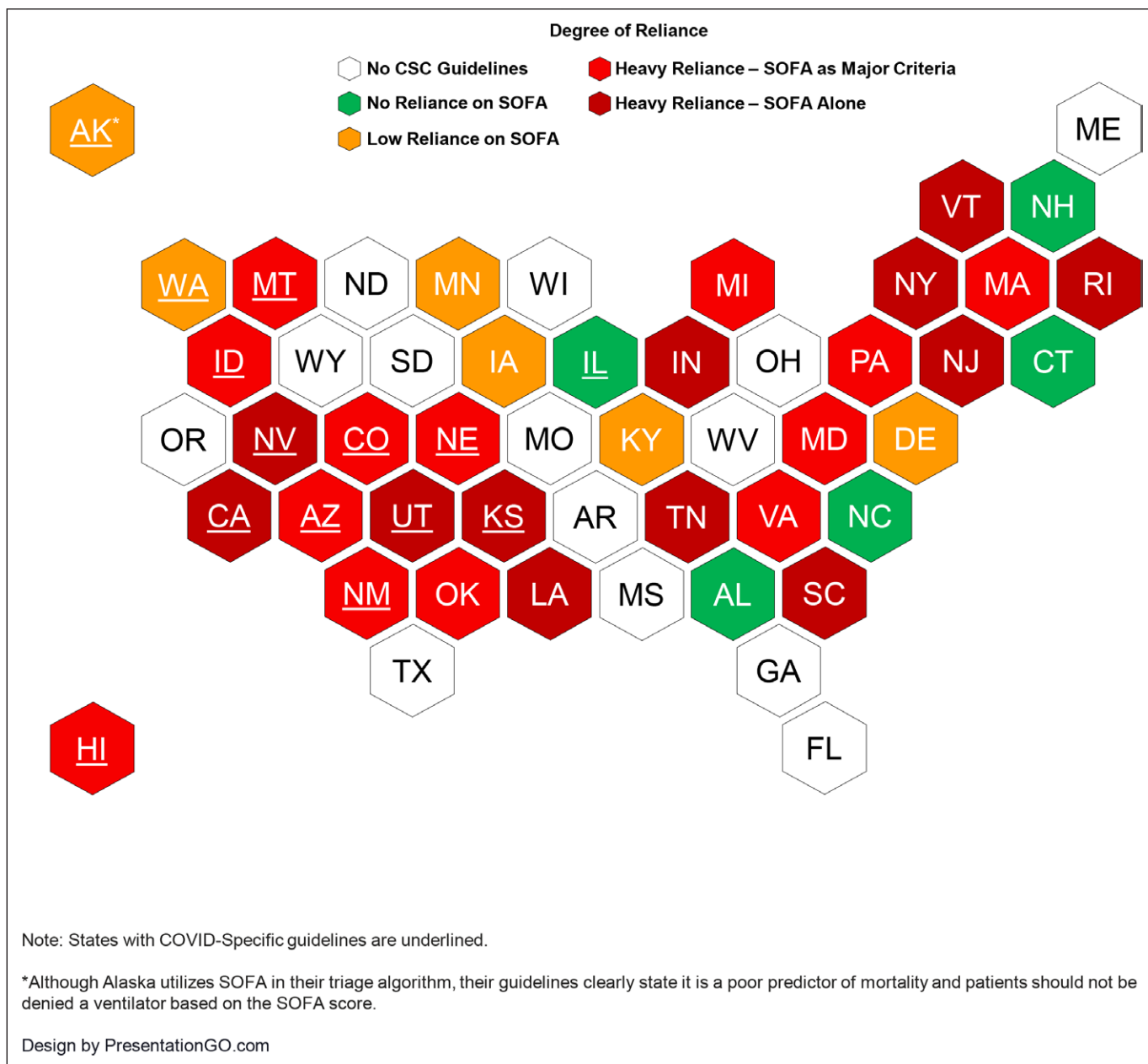


Figure 3. Heat map illustrating the availability of crisis standards of care (CSC) protocols in the United States by state and degree of reliance on Sequential Organ Failure Assessment (SOFA) score to guide scarce resource allocation. States with COVID-specific guidelines are underlined.

raise questions on the appropriateness of ventilator triage decisions that rely solely on the SOFA score to gauge short-term survivability. Our results indicate that the combination of SOFA and age provide better prediction for mortality than SOFA alone. However, the use of age in CSC triage algorithms as a primary or even as a tie-breaker element has been controversial; on the one hand, its inclusion may bias against the elderly (46), and on the other hand, may offer everyone an equal opportunity to achieve a normal lifespan (47). Comprehensive evaluation of the magnitude of

implicit triage decisions prevailing during a surge and their impact on the observed outcome in the elderly will further inform this decision. Some CSC guidelines do acknowledge the limitations of the SOFA score and place less emphasis on its role in the allocation of scarce resources (4, 48). Future studies should aim to develop more accurate and pragmatic triage protocols that incorporate novel predictors of mortality in patients with COVID-19. These protocols should also aim to balance maximizing survival with the equitable distribution of resources across society to mitigate

disparities in health outcomes (49, 50). In addition, it will be important to continue to elicit informed public opinion and engage in surveys and focus groups to ensure that triage decisions remain patient-centered.

This study has several strengths. We have externally validated the model used in a smaller, retrospective cohort study. The use of derivation and validation cohorts allowed for proper internal validation of our models. Model performance was further evaluated by assessing model calibration. The results of this study were robust to sensitivity analyses utilizing tree-based models, different parameterizations of SOFA score, and excluding patients in whom chronically elevated creatinine might bias model performance. These facets are congruent with recent best practice statements concerning the development and reporting of predictive models (51). Although prior studies have demonstrated the poor predictive accuracy of SOFA for mortality, our study is directly applicable to a time period and population more relevant to triage scenarios. Because ventilator triage was not sizably conducted in the United States during the time of our study, our findings are unlikely to be influenced by actual triage-based allocation of mechanical ventilation.

There are limitations to this study. Our cohort might not be nationally representative, limiting the generalizability of our findings. As this is a retrospective study evaluating EHR data, the exact time when mechanical ventilation was initiated on a given day cannot be certain. The accuracy and intensity of preexisting comorbid conditions may have been limited using diagnosis codes to identify them. Patient outcomes beyond hospital discharge were not assessed. About 40% of patients had a component of the SOFA score imputed as missing-as-normal. However, a sensitivity analysis excluding these patients remained consistent with our primary analysis, and a recent study demonstrating this technique provides similar results to other imputation techniques (52). More complex models with additional variables (including inflammatory biomarkers) may have provided better predictive capacity; however, our objective was to examine the adequacy of select elements commonly found in existing CSC triage algorithms.

CONCLUSIONS

In conclusion, among hospitalized patients with COVID-19, the SOFA score within 24 hours prior to

intubation shows inadequate discriminant accuracy for inhospital mortality. Caution should be taken in implementing the SOFA score in mechanical ventilator triage protocols for COVID-19, especially as the solitary or heavily weighted determinant seen in CSC guidelines currently endorsed by many U.S. states. More research is required to develop practical, accurate, and patient-centered scoring systems for inclusion in mechanical ventilator triage protocols for COVID-19 patients.

ACKNOWLEDGMENTS

We thank Mariam Noorulhuda, Christine Grady, and David Wendler of the National Institutes of Health (NIH) Bioethics Consultation Service for their thoughtful insight and guidance with the development of this article. This work used the computational resources of the NIH High Performance Computing Biowulf cluster (<http://hpc.nih.gov>).

- 1 *Critical Care Medicine Department, NIH Clinical Center, Bethesda, MD.*
- 2 *Department of Pulmonary and Critical Care Medicine, Johns Hopkins Hospital, Baltimore, MD.*
- 3 *Clinical Monitoring Research Program Directorate, Frederick National Laboratory for Cancer Research, Frederick, MD.*
- 4 *Biostatistics Research Branch, Division of Clinical Research, National Institute of Allergy and Infectious Diseases, Bethesda, MD.*

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (<http://journals.lww.com/ccmjjournal>).

Supported, in part, by the Intramural Research Program of the National Institutes of Health Clinical Center. Supported, in whole or in part, with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. 75N910D00024, Task Order No. 75N91019F00130.

Drs. Keller, Wang, Warner, Follmann, and Kadri received support for article research from the National Institutes of Health (NIH). Drs. Keller, Nason, Warner, Follmann, and Kadri disclosed government work. Dr. Wang disclosed work for hire. Dr. Kadri disclosed that he is an Associate Editor for Critical Care Medicine; he disclosed that this work was performed as part of official duties of the authors employed by the NIH.

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, or does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. The funders had no role in the design and conduct of the study or the collection, management, analysis, and interpretation of the data; preparation, review, or approval of the article; and decision to submit the article for publication. The

findings and conclusions in this study are those of the authors and do not necessarily represent the official position of the National Institutes of Health.

For information regarding this article, E-mail: sameer.kadri@nih.gov; michael.keller2@nih.gov

REFERENCES

- Kadri SS: association between caseload surge and COVID-19 survival in 558 U.S. hospitals, March to August 2020. *Ann Intern Med* 2021; 174:1240–1251
- Committee on Guidance for Establishing Crisis Standards of Care for Use in Disaster S, Institute of M: *In: Crisis Standards of Care: A Systems Framework for Catastrophic Disaster Response*. Washington, DC, National Academies Press, 2012
- New York State Task Force on Life and the Law New York State Department of Health: Ventilator allocation guidelines. 2015. Available at: https://www.health.ny.gov/regulations/task_force/reports_publications/docs/ventilator_guidelines.pdf. Accessed October 1, 2021
- Washington State Department of Health, Northwest Healthcare Response Network: Scarce resource management & crisis standards of care. 2020. Available at: https://nwhrn.org/wp-content/uploads/2020/03/Scarce_Resource_Management_and_Crisis_Standards_of_Care_Overview_and_Materials-2020-3-16.pdf. Accessed October 1, 2021
- University of Pittsburgh: Allocation of scarce critical care resources during a public health emergency. 2020. Available at: https://ccm.pitt.edu/sites/default/files/UnivPittsburgh_ModelHospitalResourcePolicy_2020_04_15.pdf. Accessed October 1, 2021
- New Mexico triage protocol for the allocation of scarce resources under COVID-19 crisis standards of care. Available at: https://cvmodeling.nmhealth.org/wp-content/uploads/sites/4/2020/05/NMMATAAssessmentTriageProtocol2020_05_22.pdf. Accessed October 1, 2021
- The Pandemic Influenza Ethics Initiative Work Group of the Veterans Health Administration's National Center For Ethics in Health Care: Meeting the Challenge Of Pandemic Influenza: Ethical Guidance For Leaders and Health Care Professionals in the Veterans Health Administration. 2010. Available at: https://www.ethics.va.gov/docs/policy/meeting_the_challenge_of_pan_flu_ethical_guidance_vha_2010_web_posting_2013.pdf. Accessed October 1, 2021
- National Academies of Sciences, Engineering, and Medicine 2020: Rapid Expert Consultation on Crisis Standards of Care for the COVID-19 Pandemic. Washington, DC, The National Academies Press, 2020
- Antommaria AHM, Gibb TS, McGuire AL, et al; Task Force of the Association of Bioethics Program Directors: Ventilator triage policies during the COVID-19 pandemic at U.S. hospitals associated with members of the Association of Bioethics Program Directors. *Ann Intern Med* 2020; 173:188–194
- Piscitello GM, Kapania EM, Miller WD, et al: Variation in ventilator allocation guidelines by US State during the coronavirus disease 2019 pandemic: A systematic review. *JAMA Netw Open* 2020; 3:e2012606
- Chelen JSC, White DB, Zaza S, et al: US ventilator allocation and patient triage policies in anticipation of the COVID-19 surge. *Health Secur* 2021; 19:459–467
- Alaska Department of Health and Social Services: Patient Care Strategies For Scarce Resource Situations: Alaska Department of Health and Social Services, Division of Public Health, Rural and Community Health Systems. 2021. Available at: http://dhss.alaska.gov/dph/Epi/id/SiteAssets/Pages/HumanCoV/SOA_DHSS_CrisisStandardsOfCare.pdf. Accessed October 1, 2021
- Idaho Department of Health & Welfare: Patient Care Strategies for Scarce Resource Situations: Idaho Department of Health and Welfare. 2020. Available at: https://coronavirus.idaho.gov/wp-content/uploads/2020/10/2020_ID_CSC_Strategies_v2_Final_Posted-1.pdf. Accessed October 1, 2021
- Seymour CW, Liu VX, Iwashyna TJ, et al: Assessment of clinical criteria for sepsis: For the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315:762–774
- Raith EP, Udy AA, Bailey M, et al; Australian and New Zealand Intensive Care Society (ANZICS) Centre for Outcomes and Resource Evaluation (CORE): Prognostic accuracy of the SOFA score, SIRS criteria, and qSOFA score for in-hospital mortality among adults with suspected infection admitted to the intensive care unit. *JAMA* 2017; 317:290–300
- Raschke RA, Agarwal S, Rangan P, et al: Discriminant accuracy of the SOFA score for determining the probable mortality of patients with COVID-19 pneumonia requiring mechanical ventilation. *JAMA* 2021; 325:1469–1470
- Maves RC, Downar J, Dichter JR, et al; ACCP Task Force for Mass Critical Care: Triage of scarce critical care resources in COVID-19 an implementation guide for regional allocation: An expert panel report of the task force for mass critical care and the American College of Chest Physicians. *Chest* 2020; 158:212–225
- Qeadan F, VanSant-Webb E, Tingey B, et al: Racial disparities in COVID-19 outcomes exist despite comparable Elixhauser comorbidity indices between Blacks, Hispanics, Native Americans, and Whites. *Sci Rep* 2021; 11:8738
- Kadri SS, Gundrum J, Warner S, et al: Uptake and accuracy of the diagnosis code for COVID-19 among US hospitalizations. *JAMA* 2020; 324:2553–2554
- Rhee C, Zhang Z, Kadri SS, et al; CDC Prevention Epicenters Program: Sepsis surveillance using adult sepsis events simplified eSOFA criteria versus sepsis-3 sequential organ failure assessment criteria. *Crit Care Med* 2019; 47:307–314
- Rice TW, Wheeler AP, Bernard GR, et al; National Institutes of Health, National Heart, Lung, and Blood Institute ARDS Network: Comparison of the SpO₂/FIO₂ ratio and the PaO₂/FIO₂ ratio in patients with acute lung injury or ARDS. *Chest* 2007; 132:410–417
- Soo A, Zuege DJ, Fick GH, et al: Describing organ dysfunction in the intensive care unit: A cohort study of 20,000 patients. *Crit Care* 2019; 23:186
- Freund Y, Lemachatti N, Krastinova E, et al; French Society of Emergency Medicine Collaborators Group: Prognostic accuracy of sepsis-3 criteria for in-hospital mortality among

- patients with suspected infection presenting to the emergency department. *JAMA* 2017; 317:301–308
24. Liang W, Liang H, Ou L, et al; China Medical Treatment Expert Group for COVID-19: Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern Med* 2020; 180:1081–1089
 25. Hendren NS, de Lemos JA, Ayers C, et al: Association of body mass index and age with morbidity and mortality in patients hospitalized with COVID-19: Results from the American Heart Association COVID-19 cardiovascular disease registry. *Circulation* 2021; 143:135–144
 26. Incerti D, Rizzo S, Li X, et al: Prognostic model to identify and quantify risk factors for mortality among hospitalised patients with COVID-19 in the USA. *BMJ Open* 2021; 11:e047121
 27. Estiri H, Strasser ZH, Klann JG, et al: Predicting COVID-19 mortality with electronic medical records. *NPJ Digit Med* 2021; 4:15
 28. Guan WJ, Liang WH, Zhao Y, et al: Comorbidity and its impact on 1590 patients with COVID-19 in China: A nationwide analysis. *Eur Respir J* 2020; 55:2000547
 29. Jordan RE, Adab P, Cheng KK: Covid-19: Risk factors for severe disease and death. *BMJ* 2020; 368:m1198
 30. Elixhauser A, Steiner C, Harris DR, et al: Comorbidity measures for use with administrative data. *Med Care* 1998; 36:8–27
 31. Li B, Evans D, Faris P, et al: Risk adjustment performance of Charlson and Elixhauser comorbidities in ICD-9 and ICD-10 administrative databases. *BMC Health Serv Res* 2008; 8:12
 32. Hantel A, Marron JM, Casey M, et al: US State Government crisis standards of care guidelines: Implications for patients with cancer. *JAMA Oncol* 2021; 7:199–205
 33. Ohio Hospital Association: Guidelines for allocation of scarce medical resources. 2020. Available at: <https://ohiohospitals.org/OHA/media/OHA-Media/Documents/Patient%20Safety%20and%20Quality/COVID19/Ohio-Guidelines-for-Allocation-of-Scarce-Medical-Resources-CLEAN-FINAL.pdf>. Accessed October 1, 2021
 34. Missouri Hospital Association: A framework for managing the 2020 COVID-19 pandemic response and implementing crisis standards of care. 2021. Available at: https://medicine.missouri.edu/sites/default/files/A_Framework_for_Managing_2020_COVID.pdf. Accessed October 1, 2021
 35. Hothorn T, Hornik K, Zeileis A: Unbiased recursive partitioning: A conditional inference framework. *J Comput Graph Stat* 2006; 15:651–674
 36. Youngstrom EA: A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *J Pediatr Psychol* 2014; 39:204–221
 37. Robin X, Turck N, Hainard A, et al: pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; 12:77
 38. Hosmer DW, Lemeshow S: Goodness of fit tests for the multiple logistic regression model. *Commun Stat Theory Methods* 1980; 9:1043–1069
 39. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 1988; 44:837–845
 40. Vincent JL, Moreno R, Takala J, et al: The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996; 22:707–710
 41. Lambden S, Laterre PF, Levy MM, et al: The SOFA score-development, utility and challenges of accurate assessment in clinical trials. *Crit Care* 2019; 23:374
 42. Ferreira FL, Bota DP, Bross A, et al: Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA* 2001; 286:1754–1758
 43. Pettilä V, Pettilä M, Sarna S, et al: Comparison of multiple organ dysfunction scores in the prediction of hospital mortality in the critically ill. *Crit Care Med* 2002; 30:1705–1711
 44. Ashana DC, Anesi GL, Liu VX, et al: Equitably allocating resources during crises: Racial differences in mortality prediction models. *Am J Respir Crit Care Med* 2021; 204:178–186
 45. Hick JL, Hanfling D, Wynia M, et al: Crisis standards of care and COVID-19: What did we learn? How do we ensure equity? What should we do? *NAM Perspect* 2021; 2021:10.31478/202108e
 46. Altman MC: A consequentialist argument for considering age in triage decisions during the coronavirus pandemic. *Bioethics* 2021; 35:356–365
 47. Hospital Utilization: HHS protect public data hub. 2015. Available at: <https://protect-public.hhs.gov/pages/hospital-utilization>. Accessed October 1, 2021
 48. Minnesota Department of Health: Patient Care Strategies For Scarce Resource Situations: Minnesota Health Care Preparedness Program. 2021. Available at: <https://www.health.state.mn.us/communities/ep/surge/crisis/standards.pdf>. Accessed October 1, 2021
 49. White DB, Lo B: Mitigating inequities and saving lives with ICU triage during the COVID-19 pandemic. *Am J Respir Crit Care Med* 2021; 203:287–295
 50. Kesler SM, Wu JT, Kalland KR, et al: Operationalizing ethical guidance for ventilator allocation in Minnesota: Saving the most lives or exacerbating health disparities? *Crit Care Explor* 2021; 3:e0455
 51. Leisman DE, Harhay MO, Lederer DJ, et al: Development and reporting of prediction models: Guidance for authors from editors of respiratory, sleep, and critical care journals. *Crit Care Med* 2020; 48:623–633
 52. Schenck EJ, Hoffman KL, Oromendia C, et al: A comparative analysis of the respiratory subscore of the Sequential Organ Failure Assessment scoring system. *Ann Am Thorac Soc* 2021; 18:1849–1860