

Empirical Assessment of U.S. Coronavirus Disease 2019 Crisis Standards of Care Guidelines

OBJECTIVES: To establish the feasibility of empirically testing crisis standards of care guidelines.

DESIGN: Retrospective single-center study.

SETTING: ICUs at a large academic medical center in the United States.

SUBJECTS: Adult, critically ill patients admitted to ICU, with 27 patients admitted for acute respiratory failure due to coronavirus disease 2019 and 37 patients admitted for diagnoses other than coronavirus disease 2019.

INTERVENTIONS: Review of electronic health record.

MEASUREMENTS AND MAIN RESULTS: Many U.S. states released crisis standards of care guidelines with algorithms to allocate scarce healthcare resources during the coronavirus disease 2019 pandemic. We compared state guidelines that represent different approaches to incorporating disease severity and comorbidities: New York, Maryland, Pennsylvania, and Colorado. Following each algorithm, we calculated priority scores at the time of ICU admission for a cohort of patients with primary diagnoses of coronavirus disease 2019 and diseases other than coronavirus disease 2019 ($n = 64$). We assessed discrimination of 28-day mortality by area under the receiver operating characteristic curve. We simulated real-time decision-making by applying the triage algorithms to groups of two, five, or 10 patients. For prediction of 28-day mortality by priority scores, area under the receiver operating characteristic curve was 0.56, 0.49, 0.53, 0.66, and 0.69 for New York, Maryland, Pennsylvania, Colorado, and raw Sequential Organ Failure Assessment score algorithms, respectively. For groups of five patients, the percentage of decisions made without deferring to a lottery were 1%, 57%, 80%, 88%, and 95% for New York, Maryland, Pennsylvania, Colorado, and raw Sequential Organ Failure Assessment score algorithms, respectively. The percentage of decisions made without lottery was higher in the subcohort without coronavirus disease 2019, compared with the subcohort with coronavirus disease 2019.

CONCLUSIONS: Inclusion of comorbidities does not consistently improve an algorithm's performance in predicting 28-day mortality. Crisis standards of care algorithms result in a substantial percentage of tied priority scores. Crisis standards of care algorithms operate differently in cohorts with and without coronavirus disease 2019. This proof-of-principle study demonstrates the feasibility and importance of empirical testing of crisis standards of care guidelines to understand whether they meet their goals.

KEY WORDS: crisis triage; ethical triage; intensive care; intensive care unit; medical ethics

The coronavirus disease 2019 (COVID-19) pandemic has renewed debate over crisis standards of care (CSC) (1, 2). Several states in the United States have released guidelines for allocating healthcare resources, like ventilators and ICU beds, if the number needed exceeds supply (3). Although U.S. hospitals largely avoided implementation of these guidelines, careful scrutiny is needed given

Maheetha Bharadwaj, MS, MPhil¹

Julia L. Jezmir, MD, MBA^{1,2}

Sandeep P. Kishore, MD, PhD¹⁻³

Marisa Winkler, MD, PhD^{1,4}

Bradford Diephus, MD, MBA^{1,2}

Hibah Haider, MD⁵

Conor P. Crowley, AG-ACNP⁵

Mayra Pinilla-Vera, MD⁵

Jack Varon, MD^{1,5}

Rebecca M. Baron, MD^{1,5}

William B. Feldman, MD, DPhil^{1,5,6}

Edy Y. Kim, MD, PhD^{1,5}

Copyright © 2021 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of the Society of Critical Care Medicine. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/CCE.000000000000496

the ongoing global COVID-19 pandemic and possible future pandemics. CSC guidelines aim to maximize population benefits rather than allocate critical care resources on an “all-come, all-served basis” (1, 2). Most guidelines rely upon Sequential Organ Failure Assessment (SOFA) scores as a measure of disease severity to identify those likely to survive hospitalization and maximize the number of lives saved (3). However, states have differing approaches on whether to account for patient comorbidities. Each state guideline specifies an algorithm that assigns a priority score, and the patient with the lower priority score receives the scarce resources (**Table 1**) (**Supplemental Table 1**, <http://links.lww.com/CCX/A724>). For example, in the Pennsylvania system (as initially devised), if two patients with the same SOFA score require a ventilator, but one patient has fewer comorbidities, then the patient with fewer comorbidities would receive the ventilator. Each state guideline also outlines tiebreaker procedures that are decided by age and/or lottery, and some take into account healthcare worker status and pregnancy (4–7).

Although state CSC guidelines have been developed with considerable debate, their validity has not, to our knowledge, been empirically examined in mixed COVID-19 and non-COVID-19 cohorts, although limited testing has been performed with COVID-19 patients alone and with non-COVID-19 patients alone (8–10). Yet, the ethical justification for these guidelines depends, in part, on the assumption that priority scores will correlate with clinical outcomes and efficiently allocate scarce resources. In this study, we propose a proof-of-concept model to empirically test CSC guidelines. We compared four representative U.S. state guidelines that all used SOFA scores but had varying approaches to patient comorbidities: New York, Maryland, Pennsylvania, and Colorado (Table 1) (Supplemental Table 1, <http://links.lww.com/CCX/A724>) (3, 11). We also assessed a hypothetical triage algorithm that uses raw (i.e., ungrouped) SOFA scores and age as a tiebreaker. We first assessed whether the priority scores produced by state guidelines correlated with 28-day outcomes. Then, we assessed how often the algorithms led to tied priority scores requiring a lottery.

MATERIALS AND METHODS

Study Cohort

Our retrospective cohort study included 64 critically ill patients, of whom 27 had a primary diagnosis of

COVID-19 pneumonia and 37 had a non-COVID-19 diagnosis. The patients with COVID-19 were consecutively admitted to the ICU at Brigham and Women's Hospital between March 12, 2020, and April 3, 2020, with COVID-19 infection under the Mass General Brigham (MGB) Institutional Review Board (IRB)-approved protocol number 2020P001139. COVID-19 infection was defined by severe acute respiratory syndrome coronavirus 2 real-time polymerase chain reaction testing of nasopharyngeal swabs. The 37 non-COVID-19 patients were drawn from the Brigham and Women's Hospital Registry of Critical Illness, a well-established convenience cohort of patients enrolled within 48 hours of admission to the medical ICU under MGB IRB-approved protocol number 2008P000495 and described in references (12–17). These patients were enrolled prior to the COVID-19 pandemic and had a variety of admission diagnoses, including sepsis, cardiogenic shock, acute hypoxemic respiratory failure, chronic obstructive pulmonary disease exacerbation, or upper airway obstruction. We included consecutive patients enrolled from November 3, 2018, to May 7, 2019, a time range selected to overlap with the calendar months of the COVID-19 cohort while accruing patients in a 3:2 ratio with the COVID-19 cohort. This ratio simulates one realistic situation in which there is a large number of COVID-19 patients, but the majority of patients have non-COVID-19 diagnoses. Patients were independently reviewed for inclusion by two board-certified pulmonary and critical care attending physicians.

We applied each of the four state's CSC guidelines to the cohort to calculate priority scores at the time of patients' admission to the ICU, a key triage decision point if resources are limited. Outcomes were either dead (0) or alive (1) at 28 days post ICU admission (18). Two independent reviewers manually abstracted data from electronic health records, and conflicts were adjudicated by a third reviewer, simulating a typical workflow by which triage teams assess priority scores.

Calculating SOFA Scores

We tested the first versions of the CSC guidelines released during the COVID-19 pandemic. All state guidelines had prioritization algorithms that assign priority points based on groupings of raw SOFA scores. New York's algorithm included SOFA score groupings alone,

TABLE 1.
Crisis Standard of Care State Guidelines

Scoring Criteria	Guidelines by State			
	Maryland	Pennsylvania	Colorado	New York ^f
SOFA prioritization ^a	≤ 8: 1 point 9–11: 2 points 12–14: 3 points > 14: 4 points	< 6: 1 point 6–8: 2 points 9–11: 3 points ≥ 12: 4 points	< 6: 1 point 6–9: 2 points 10–12: 3 points > 12: 4 points	< 7: 1 point 8–11: 2 points > 11: 3 points
Comorbidities scoring ^b	Severe comorbidities: 3 points	Major comorbidities: 2 points Severe comorbidities: 4 points	Modified Charlson Comorbidity Index ^c	None
Special considerations	Pregnancy: ^d –1 point	Pregnancy: –2 points Healthcare worker: –1 point	None	None
Priority score calculation	SOFA prioritization + comorbidities score + special considerations	SOFA prioritization + comorbidities score + special considerations	SOFA prioritization + Charlson Comorbidity Index Score	SOFA prioritization
Priority grouping based on priority score	None	High priority: 1–3 Intermediate priority: 4–5 Low priority: ≥ 6	None	High priority: 1 Intermediate priority: 2 Low priority: 3
Tiebreakers ^e	First tiebreaker: Life cycle Second tiebreaker: Lottery	First tiebreaker: Life cycle Second tiebreaker: SOFA prioritization Third tiebreaker: Lottery	First tiebreaker: Children, healthcare workers, and/or first responders Second tiebreaker: Life cycle, pregnancy, and/or sole caretakers for elderly Third tiebreaker: Lottery	First tiebreaker: Children Second tiebreaker: Lottery

SOFA = Sequential Organ Failure Assessment.

^aPatients assigned lower priority scores are more likely to receive the scarce resources.

^bPlease refer to text and Table 3 for lists of major and severe comorbidities for each algorithm.

^cThe modified Charlson Comorbidity Index is provided in Table 3.

^dFor Maryland, only pregnancy with a “viable fetus” is considered for a point reduction.

^eLife cycle groupings were different for each algorithm. Maryland: 0–49 = 1 (highest), 50–69 = 2, 70–84 = 3, 85+ = 4; Pennsylvania: 0–40 = 1 (highest), 41–60 = 2, 61–75 = 3, 76+ = 4; Colorado: 0–49 = 1 (highest), 50–59 = 2, 60–69 = 3, 70–79 = 4, 80+ = 5.

^fThe New York Algorithm exclusion criteria include the following: 1) unwitnessed cardiac arrest, recurrent arrest without hemodynamic stability, arrest unresponsive to standard interventions and measures, trauma-related arrest; 2) irreversible age-specific hypotension unresponsive to fluid resuscitation and vasopressor therapy; 3) traumatic brain injury with no motor response to painful stimulus (i.e., best motor response = 1); 4) severe burns where predicted survival ≤ 10% even with unlimited aggressive therapy; and 5) any other conditions resulting in immediate or near-immediate mortality even with aggressive therapy. None of the patients in this cohort fell into this exclusion criteria.

whereas the other three state algorithms (Maryland, Pennsylvania, and Colorado) included comorbidities when assigning priority points. The manner by which comorbidities were incorporated differed in Maryland, Pennsylvania, and Colorado. To calculate

the final priority score for these three states, the priority points from SOFA scores and the priority points from comorbidities were added. For Pennsylvania, priority scores are also converted into three priority categories (Table 1), as specified in their guidelines.

Each algorithm outlines procedures for tiebreakers when two or more patients have the same priority score. Maryland and Colorado use their own unique age categories (Table 1) as tiebreakers and then a lottery if ties remain. Colorado also specifies that one's status as a child, healthcare worker, first responder, pregnant woman, or caretaker for the elderly may enter into tiebreaker considerations. Pennsylvania uses its own age categories (Table 1) as the first tiebreaker, followed by raw priority scores as the second tiebreaker, followed by lottery for any remaining ties. New York does not use age and goes directly to a lottery when priority scores are tied. In addition to these four algorithms, we also tested a hypothetical algorithm that used raw SOFA scores with age as a tiebreaker.

Primary Analysis

We calculated an area under the receiver operating characteristic (AUROC) curve to assess the accuracy of each CSC algorithm in discriminating 28-day mortality, with admission to the ICU serving as day 1. For this analysis, we used priority scores of algorithms "without" adding tiebreakers. Patients who were discharged alive from the hospital prior to 28 days were considered to be alive at 28 days (the validity of this assumption was verified in a subset of patients, described elsewhere) (18).

Groups of Two, Five, and 10 Analysis

To assess how the prioritization algorithms may function clinically, we compared the priority scores for groups of two, five, or 10 patients using a bootstrap method. Each "group" represented a situation in which a triage algorithm would have to choose one individual of the group to receive scarce resources based on the priority score assigned by the algorithm. For each group of patients, a "winner" with the "best" priority score (i.e., lowest priority point total) was selected, and the "winner's" 28-day outcome (survivor or deceased) was noted. The group was considered "tied" if two or more patients tied for the "best" (lowest) priority point total.

We performed 100 iterations of a computational simulation in which we randomly selected 1,000 groups of two, five, or 10 patients (**Supplemental Fig. 1**, <http://links.lww.com/CCX/A726>). We excluded patient groups in which all the patients had the same outcome (i.e., all survivors or all deceased), since we

cannot assess if the algorithm correctly selects a patient with a better outcome if all the patients in that group of two to 10 patients shared the same outcome. For each simulation of 1,000 patient groups, we calculated the percentage of groups for which the algorithm chose a patient who survived, and we computed the percentage of groups in which the algorithm required a tiebreaking lottery. These simulations yielded a distribution of the "percentage of algorithm decisions selecting a survivor" and "percentage of decisions requiring a lottery." An unpaired *t* test was used to calculate significant differences between the distributions.

A hypothetical example of this method is illustrated in Supplemental Figure 1 (<http://links.lww.com/CCX/A726>). In step 1, a simulation of 1,000 groups of five patients resulted in 952 groups in which at least one patient in the group of five patients had a different outcome than the other patients. The 48 groups in which all patients survived or all died were excluded from further analysis. The state CSC algorithms were applied to the 952 groups. For one group of five patients shown as an example in step 2, CSC algorithms resulted in one of three decisions: 1) a tie (e.g., New York, Maryland); 2) selection of a patient "winner" who was a survivor (e.g., Pennsylvania); or 3) selection of a patient "winner" who died (e.g., Colorado, raw SOFA + age). By applying the CSC algorithms to the other 951 groups of five patients, we determined the percentage of patient groups with ties and the percentage of patient groups for which the algorithm correctly selected a survivor at 28 days (step 3). We then repeated this simulation for a total of 100 iterations (of 1,000 patient groups in each iteration) to determine a distribution of the percentage of decisions resulting in ties and the percentage of decisions selecting a surviving patient (step 4). We also conducted these analyses for groups of two or 10 patients. We ran these analyses for the entire cohort of mixed COVID-19 and non-COVID-19 patients, and we ran these analyses for the COVID-19 subcohort and non-COVID-19 subcohort separately.

Statistical Analysis

The primary outcome was mortality at 28 days post hospital admission. For patients who were discharged alive from the hospital before 28 days, their 28-day outcome was imputed as surviving. Continuous variables were assessed for normality by the Shapiro-Wilk

test, and unpaired *t* test applied for normal variables and Mann-Whitney *U* tests for nonparametric variables, both two-tailed. Categorical variables were compared by Fisher exact test (two-sided). AUROC curve was used to assess each triage algorithm's discrimination of 28-day mortality (18). The simulations of patient groups of two, five, or 10 were performed as described above. SE was calculated using the DeLong et al method, and CIs by exact binomial test. Statistical tests were performed in IBM SPSS Statistics Version 25.0 (IBM, Armonk, NY) and R Version 3.6 (The R Project for Statistical Computing, Vienna, Austria).

RESULTS

Characteristics of Our Cohort

A total of 64 patients with COVID-19 ($n = 27$) and without COVID-19 ($n = 37$) met criteria for inclusion. At 28 days post ICU admission, 18 patients (28%) had died (Table 2). Half of the cohort was female. Mean age \pm SD was 62.3 ± 16.9 for the entire cohort, 68.1 ± 13.1 for the COVID-only cohort, and 58.2 ± 18.3 for the non-COVID-only cohort. Mean raw SOFA scores \pm SD were 4.6 ± 3.5 for the entire cohort, 3.2 ± 1.4 for COVID-only cohort, and 5.6 ± 4.2 for non-COVID-only cohort. Among our entire cohort, the prevalence of comorbidities was active malignancy (23%), chronic pulmonary disease (22%), chronic renal disease (19%), congestive heart failure (14%), and diabetes with complications (9%). Priority scores were calculated at ICU admission, a key decision point for many crisis scenarios.

Accuracy of CSC Algorithms in Discriminating 28-Day Mortality

For the entire cohort, the AUROC (95% CI) was 0.56 (0.39–0.72) for New York's algorithm, 0.49 (0.33–0.64) for Maryland's algorithm, 0.53 (0.37–0.70) for Pennsylvania's algorithm, 0.66 (0.52–0.80) for Colorado's algorithm, and 0.69 (0.56–0.82) for our hypothetical algorithm that employed raw SOFA scores (Fig. 1C).

For the COVID-19 subcohort, the AUROC (95% CI) was 0.50 (0.27–0.73) for New York's algorithm, 0.55 (0.32–0.78) for Maryland's algorithm, 0.52 (0.30–0.75) for Pennsylvania's algorithm, 0.72 (0.53–0.92) for Colorado's algorithm, and 0.80 (0.63–0.97) for our hypothetical algorithm that employed raw SOFA scores (Fig. 1A). For the non-COVID-19 subcohort,

the AUROC (95% CI) was 0.69 (0.47–0.90) for New York's algorithm, 0.49 (0.27–0.70) for Maryland's algorithm, 0.58 (0.35–0.82) for Pennsylvania's algorithm, 0.70 (0.52–0.89) for Colorado's algorithm, and 0.76 (0.57–0.95) for our hypothetical algorithm that employed raw SOFA scores (Fig. 1B).

The Performance of CSC Algorithms in Patient Groups of Two to 10

In simulations applying CSC algorithms to groups of two, five, or 10 patients, the size of the patient groups strongly affected the ability of CSC algorithms to select a patient "winner" without depending on a lottery as a tiebreaker. New York's algorithm selected a patient without a lottery in 38% (95% CI, 34–40%) of decisions in patient groups of two, 1% (0.1–1.1%) of decisions in patient groups of five, and 0% (0–0%) in patient groups of 10 (Table 3). For Maryland's algorithm, the percentage of decisions made without lottery (95% CI) were 84% (83–87%), 57% (54–60%), and 42% (40–45%) for patient groups of two, five, and 10, respectively. For Pennsylvania's algorithm, the percentage of decisions made without lottery (95% CI) were 95% (94–97%), 91% (89–93%), and 82% (80–84%) for patient groups of two, five, and 10, respectively. For Colorado's algorithm, the percentage of decisions made without lottery (95% CI) were 95% (94–97%), 88% (85–90%), and 79% (76–81%) for patient groups of two, five, and 10, respectively. For our hypothetical algorithm using raw SOFA scores with age as a tiebreaker, the percentage of decisions made without lottery (95% CI) were 99% (98–100%), 95% (93–96%), and 92% (91–94%) for patient groups of two, five, and 10, respectively.

Next in this simulation analysis, we assessed the percent of nonlottery decisions in which the algorithm made a "correct" choice and chose a patient with the better outcome (i.e., survival). The size of the patient groups facing triage decisions strongly affected algorithm performance. New York's algorithm chose a patient who survived in 64% (95% CI, 59–69%) of patient groups of two and 72% (24–100%) of groups of five. All decisions went to lottery in patients groups of 10. For Maryland's algorithm, the percentage (95% CI) of decisions where a surviving patient was selected was 58% (54–61%), 87% (84–91%), and 99% (98–100%) for patient groups of two, five, and 10, respectively. For Pennsylvania's algorithm, the percentage (95% CI) of decisions where a

TABLE 2.
Demographics, Clinical Characteristics, and Outcomes^a

Patient Characteristics	All Patients (N = 64)	COVID-19 (N = 27)	Non-COVID-19 (N = 37)	p ^a
Age, mean ± SD, yr ^b	62.3 ± 16.9	68.1 ± 13.1	58.2 ± 18.3	0.019
Gender, n (%)				
Male	31 (48)	15 (55)	16 (43)	0.45
Female	32 (50)	12 (45)	20 (54)	
Other	1 (2)	0 (0)	1 (3)	
Raw Sequential Organ Failure Assessment Scores, mean ± SD	4.6 ± 3.5	3.2 ± 1.4	5.6 ± 4.2	0.06
Comorbidities, n (%) ^c				
Congestive heart failure	9 (14)	2 (7)	7 (19)	0.28
Chronic pulmonary disease	14 (22)	4 (15)	10 (27)	0.36
Chronic renal disease	12 (19)	7 (26)	5 (14)	0.33
Dementia	2 (3)	2 (7)	0 (0)	0.17
Active malignancy	15 (23)	4 (15)	11 (30)	0.24
Diabetes with complications	6 (9)	3 (11)	3 (8)	0.69
Chronic liver disease	2 (3)	0 (0)	2 (5)	0.51
28 d outcome, n (%)				
Death	18 (28)	10 (37)	8 (22)	0.26
Alive	46 (72)	17 (63)	29 (78)	

COVID-19 = coronavirus disease 2019.

^aIn statistical comparison of COVID-19 and non-COVID-19 subcohorts, for age and Sequential Organ Failure Assessment (SOFA) score, normality was assessed by Shapiro-Wilk test, unpaired *t* test was used for age and Mann-Whitney *U* test used for SOFA score (both two-tailed). Fisher exact test (two-sided) was used for the other variables.

^bChildren, pregnant women, essential workers, and healthcare workers were not part of the cohort.

^cPlease refer to Table 3 to see each algorithm's definition of major and severe comorbidities.

surviving patient was selected was 67% (66–72%), 83% (80–86%), and 92% (91–95%) for patient groups of two, five, and 10, respectively. For Colorado's algorithm, the percentage (95% CI) of decisions where a surviving patient was selected was 67% (63–69%), 89% (87–91%), and 98% (97–99%) for patient groups of two, five, and 10, respectively. For the hypothetical algorithm using raw SOFA with age, the percentage (95% CI) of decisions where a surviving patient was selected was 71% (67–74%), 92% (90–93%), and 98% (98–99%) for patient groups of two, five, and 10, respectively.

The Effect of Heterogeneity in Algorithm Performance

We next evaluated the effect of cohort homogeneity on CSC algorithm performance. We divided our

cohort into a more homogenous subcohort of only patients with COVID-19 pneumonia (*n* = 27) and a more heterogenous subcohort of patients with a mixture of non-COVID-19 diagnoses (*n* = 37). We performed the simulation of groups of two, five, and 10 patients on these subcohorts (**Supplemental Table 2**, <http://links.lww.com/CCX/A725>). For groups of two or five patients, more decisions resulted in ties requiring lottery in the COVID-19 subcohort compared with the non-COVID-19 subcohort. For groups of five patients, the percent of decisions made without lottery for New York's algorithm were 0% (95% CI, 0–0%) in the COVID-19 subcohort and 5% (4–6%) in the non-COVID-19 subcohort. For Maryland's, Pennsylvania's, and Colorado's algorithms, the percent of decisions made without lottery were 47% (42–51%),

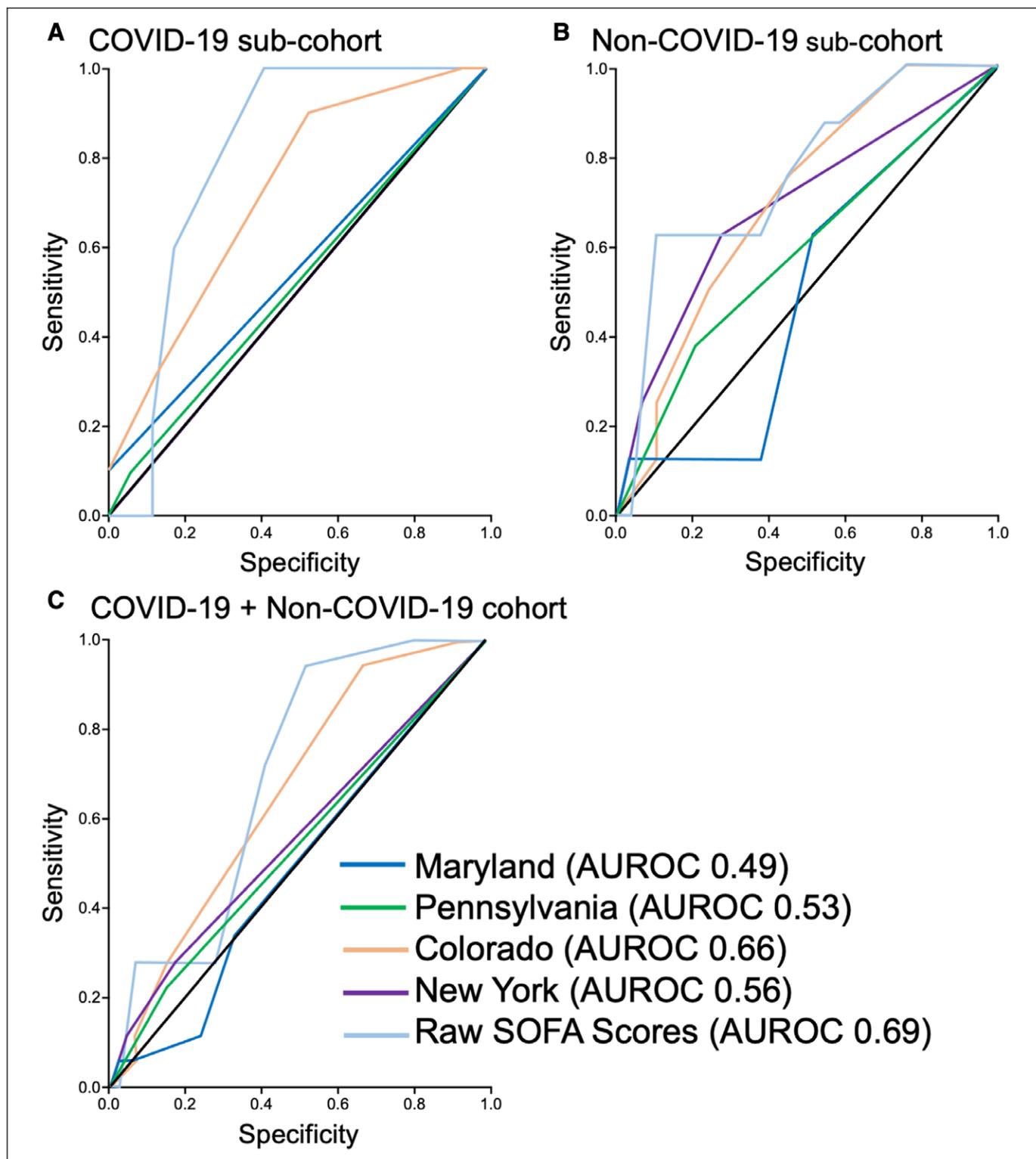


Figure 1. Area under the receiver operating characteristic (AUROC) curve analyses of prediction of 28 d mortality by priority scores. Shows the receiver operating characteristic curves for each of the state algorithms and each curve's associated AUROC, representing each algorithm's discriminatory capacity for 28 d survival. **A**, Shows the coronavirus disease 2019 (COVID-19) cohort, **B** shows the non-COVID-19 cohort, and **C** shows the entire cohort which includes both COVID-19 and non-COVID-19 patients. SOFA = Sequential Organ Failure Assessment.

57% (48–67%), and 83% (80–86%) in the COVID-19 subcohort and 68% (65–71%), 84% (82–87%), and 90% (88–92.3%) in the non-COVID-19 subcohort,

respectively, with a similar trend in groups of two patients. The state algorithms had variable trends in performance between the two subcohorts in the

TABLE 3.
Crisis Standards of Care Guidelines Performance Inpatient Groups

Group Size and CSC Algorithm	Percent of Decisions Made Without Lottery	CI	Percent of Nonlottery Decisions Where Algorithm Chose Individual Who Survived	CI
Groups of two patients				
New York	37.8	34.8–40.4	64.3	59.7–68.4
Maryland	84.3	82.7–86.3	57.8	55.3–60.7
Pennsylvania	92.1	90.5–93.9	64.5	61.5–67.2
Colorado	95.1	93.8–96.4	66.5	63.6–69.1
Raw SOFA	92.3	90.9–93.7	70.8	68.0–72.8
Raw SOFA with age	98.5	97.8–99.1	69.7	66.8–72.0
Age	98.5	97.9–99.3	60.2	57.3–63.1
Groups of five patients				
New York	0.6	0.1–1.1	71.9	24.0–100
Maryland	57.0	53.9–59.5	87.3	84.8–90.3
Pennsylvania	80.4	78.4–82.6	85.9	83.4–88.3
Colorado	87.6	85.8–89.3	89.1	87.2–91.2
Raw SOFA	72.6	70.3–74.8	93.6	91.9–95.0
Raw SOFA with age	94.6	93.2–95.8	91.7	89.9–93.1
Age	97.6	96.7–98.4	81.8	79.3–84.0
Groups of 10 patients				
New York	0	0–0	NA	NA
Maryland	41.9	39.6–44.2	98.8	97.9–99.5
Pennsylvania	68.2	65.7–70.8	96.7	95.5–97.9
Colorado	78.5	76.5–80.5	97.8	96.8–98.5
Raw SOFA	63.4	61.2–65.7	99.5	99.1–100
Raw SOFA with age	92.4	91.0–93.9	98.3	97.6–98.9
Age	97.9	97.2–98.8	94.0	92.8–94.9

NA = not applicable, SOFA = Sequential Organ Failure Assessment.

Shows the percentages and CIs for percentage of decisions made without lottery, and percentage of nonlottery decisions where the algorithms chose the survivor.

simulation of groups of 10 patients. For New York's, Maryland's, Pennsylvania's, and Colorado's algorithms, the percent of decisions made without lottery (95% CI) were 0% (0–0%), 49% (46–52%), 50% (48–53%), and 84% (81–86%) in the COVID-19 subcohort and 0% (0–0%), 38% (35–40%), 70% (67–73%), and 79% (78–81%) in the non-COVID-19 subcohort, respectively.

The performance of state algorithms in “correctly” selecting surviving patients in nonlottery decisions

was similar in the COVID-19 and non-COVID-19 cohorts. For the simulations of groups of five patients, Maryland's, Pennsylvania's, and Colorado's algorithms selected a surviving patient in 94% (86–100%), 87% (77–96%), and 94% (87–100%) of nonlottery decisions in the COVID-19 subcohort and 86% (75–97%), 87% (83–90%), and 89% (86–91%) in the non-COVID-19 subcohort, respectively. All decisions in the New York algorithm went to lottery, so this metric could not be

assessed. The comparison of performance in COVID-19 and non-COVID-19 subcohorts in selecting a survivor was similar in the simulations of groups of two or 10 patients.

DISCUSSION

Our study presents a proof-of-concept framework for empirically testing CSC state guidelines to assess how well they meet their stated goals of maximizing the number of lives saved. Our analysis of ROC curves showed that, for the initial CSC algorithms from New York, Maryland, and Pennsylvania, priority scores assigned at ICU admission had poor predictive value for 28-day survival, with better performance for Colorado and raw SOFA score. Application of state guidelines, particularly New York and Maryland, frequently resulted in identical priority scores requiring tiebreakers based on age or lottery. In simulations applying CSC algorithms to groups of two, five, or 10 patients, the ability of CSC algorithms to avoid tiebreakers and correctly select patients who survived to 28-days depended to a large extent on the simulation method, specifically the number of patients (two, five, or 10) in the triage group.

One state algorithm that included comorbidities (Colorado) had slightly improved performance (by AUROC curve) when predicting 28-day survival compared with New York's algorithm, which relied on SOFA score ranges alone. The inclusion of comorbidities, however, raises several concerns. One worry is the risk of exacerbating racial and socioeconomic disparities since those who are most disadvantaged are most likely to have multiple comorbidities (19–21). The use of age as a tiebreaker raises further concerns about age-based discrimination, although proponents often appeal to the goal of giving people equal opportunity to experience each stage of life.

Our hypothetical algorithm using raw SOFA scores with age as a tiebreaker performed better than other algorithms on several fronts. It relied the least on a lottery system, making decisions without lottery greater than 90% of the time across all analyses. Of those nonlottery decisions, this algorithm consistently chose a surviving individual in nearly 70% of decisions in groups of two analyses and greater than 90% of decisions in groups of five and 10 across all cohorts. However, the SOFA score itself has faced criticism for racial bias and limited ability to predict mortality

(8, 22–24). Furthermore, using raw SOFA scores may place undue emphasis on a tool that was never designed for triage.

A strength of our study was the inclusion of both COVID-19 and non-COVID-19 patients. The frequency of tied priority scores was higher in the COVID-19 only subcohort compared with the non-COVID-19 subcohort. The high rate of ties in the COVID-19 subcohort is consistent with another recent study of New York's and Maryland's CSC algorithms (9). One potential reason for these frequent ties is that COVID-19 often affects only one organ system, the lungs. Patients with COVID-19 frequently have SOFA scores with a respiratory component of 2+ (since a higher score requires mechanical ventilation) while receiving no points for other organ systems. The grouping of raw SOFA scores into ranges further increases the likelihood of ties. Ties are not necessarily a negative feature of CSC algorithms. Some ethicists have favored lotteries as a way to promote fair allocation. However, it is important to understand how frequently an algorithm results in ties since many algorithms, including those examined here, seek to go beyond random allocation when allocating resources.

Our study is limited by its small sample size, recruitment at a single center, and limited follow-up time. The CSC algorithms continue to evolve, and there are other CSC algorithms distinct from the approaches examined in this study (25, 26). In addition, we excluded certain variables (e.g., essential worker status) upon which several algorithms rely. Our focus was on triage to the ICU, and future work can examine other clinical situations, such as triage of noncritically ill patients and other potentially scarce resources, such as supplemental oxygen, hemodialysis, or extracorporeal membrane oxygenation.

Further work is needed to empirically test CSC algorithms. We have demonstrated that computational simulations of triage in groups of two to 10 patients can differentiate among state algorithms more effectively than AUROC analyses. The relevant group size for simulations depends on the extent to which demand exceeds supply. If 10 people need a ventilator for every one that is available, then simulations with groups of 10 are most relevant; by contrast, when demand is lower and/or supply is higher, then simulations with smaller groups may be more relevant. Understanding how each algorithm performs with different group sizes may

help decision-makers choose the right approach for a given health crisis. Further work is needed to apply the framework that we have developed to larger datasets and to identify and validate other types of prognostic information that could improve these algorithms such as imaging or biomarkers. The ethical defensibility of these algorithms depends, in part, on empirical analyses of how they function in practice.

CONCLUSIONS

This proof-of-concept study demonstrates that the performance of CSC algorithms can be quantitatively tested in cohorts of patients with critical illness due to COVID-19 or non-COVID-19 diagnoses. Simulation of triage in smaller groups of patients demonstrated that CSC algorithms have significant differences in percentages of decisions requiring a tiebreaking lottery and performance in selecting patients with a better clinical outcome. A hypothetical algorithm of raw SOFA score plus age outperformed representative U.S. state CSC algorithms.

- 1 Harvard Medical School, Boston, MA.
- 2 Department of Medicine, Brigham and Women's Hospital, Boston, MA.
- 3 Department of Health System Design & Global Health, Icahn School of Medicine at Mount Sinai, New York, NY.
- 4 Division of Infectious Diseases, Department of Medicine, Brigham and Women's Hospital, Boston, MA.
- 5 Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA.
- 6 Program on Regulation, Therapeutics, and Law, Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, Boston, MA.

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/ccejournal>).

Drs. Bharadwaj, Jezmir, and Kishore equal contributors as cofirst authors. Drs. Feldman and Kim are equal contributors as cose-nior authors.

Dr. Kishore serves as a consultant for Resolve to Save Lives for hypertension control and has led a partnership on mul-tiple chronic conditions supported by the Arnhold Institute for Global Health at the Icahn School of Medicine at Mount Sinai and Teva Pharmaceuticals. Dr. Feldman receives funding from the National Institutes of Health (5T32HL007633-35). He serves as a consultant for Alosa Health, served as a con-sultant for Aetion and received an honorarium for a presen-tation to Blue Cross/Blue Shield of Massachusetts. Dr. Kim is a coinvestigator in NCT04389671 (Windtree Therapeutics)

testing lucinactant (surfactant-like treatment) in coronavirus di-sease 2019 patients and received research funding from Bayer AG, the U.S. National Institutes of Health, the American Heart Association, and the American Thoracic Society. The remaining authors have disclosed that they do not have any conflicts of interest.

For information regarding this article, E-mail: wbfeldman@bwh.harvard.edu; ekim11@bwh.harvard.edu

REFERENCES

1. Emanuel EJ, Persad G, Upshur R, et al: Fair allocation of scarce medical resources in the time of Covid-19. *N Engl J Med* 2020;382:2049–2055
2. White DB, Lo B: A framework for rationing ventilators and critical care beds during the COVID-19 pandemic. *JAMA* 2020;146:e61S–e74S
3. Baker M, Fink S: At the Top of the Covid-19 Curve, How Do Hospitals Decide Who Gets Treatment? Available at: <https://www.nytimes.com/2020/03/31/us/coronavirus-covid-triage-rationing-ventilators.html>. Accessed June 24, 2020
4. Colorado State: Colorado Guidelines for Scarce Resources. Available at: <https://www.colorado.gov/pacific/sites/default/files/Crisis%20Standards%20of%20Care%20Triage%20Standards-April%202020.pdf>. Accessed April 30, 2020
5. Daugherty-Biddison L, Gwon H, Regenberg A, et al: Maryland Framework for the Allocation of Scarce Life-sustaining Medical Resources in a Catastrophic Public Health Emergency. Available at: https://www.law.umaryland.edu/media/SOL/pdfs/Programs/Health-Law/MHECN/ASR%20Framework_Final.pdf. Accessed June 30, 2020
6. Pennsylvania Department of Public Health: Interim Pennsylvania Crisis Standards of Care for Pandemic Guidelines. Available at: <https://www.haponline.org/Portals/1/docs/COVID-19/COVID-19-Interim-Crisis-Standards-of-Care-040320-DRAFT-FOR-REVIEW.pdf?ver=2020-04-03-154613-847>. Accessed June 30, 2020
7. 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
8. Christian MD, Hamielec C, Lazar NM, et al: A retrospective cohort pilot study to evaluate a triage tool for use in a pandemic. *Crit Care* 2009; 13:R170
9. Rubin EB, Knipe RS, Israel RA, et al: Existing crisis standards of care triage protocols may not significantly differentiate between patients with coronavirus disease 2019 who require intensive care. *Crit Care Explor* 2021; 3:e0412
10. Wunsch H, Hill AD, Bosch N, et al: Comparison of 2 triage scoring guidelines for allocation of mechanical ventilators. *JAMA Netw Open* 2020; 3:e2029250
11. NY State: NY Guidelines for Ventilators. Available at: https://www.health.ny.gov/regulations/task_force/reports_publications/docs/ventilator_guidelines.pdf. Accessed June 30, 2020
12. Dolinay T, Kim YS, Howrylak J, et al: Inflammasome-regulated cytokines are critical mediators of acute lung injury. *Am J Respir Crit Care Med* 2012; 185:1225–1234

13. Schumski A, Ortega-Gómez A, Wichapong K, et al: Endotoxemia accelerates atherosclerosis through electrostatic charge-mediated monocyte adhesion. *Circulation* 2021; 143:254–266
14. Fajnzylber J, Regan J, Coxen K, et al; Massachusetts Consortium for Pathogen Readiness: SARS-CoV-2 viral load is associated with increased disease severity and mortality. *Nat Commun* 2020; 11:5493
15. Reyes M, Filbin MR, Bhattacharyya RP, et al: An immune-cell signature of bacterial sepsis. *Nat Med* 2020; 26:333–340
16. Kim EY, Ner-Gaon H, Varon J, et al: Post-sepsis immunosuppression depends on NKT cell regulation of mTOR/IFN- γ in NK cells. *J Clin Invest* 2020; 130:3238–3252
17. Rogers AJ, McGeachie M, Baron RM, et al: Metabolomic derangements are associated with mortality in critically ill adult patients. *PLoS One* 2014; 9:e87538
18. Yeming W, Fan G, Horby P, et al: Comparative outcomes of adults hospitalized with seasonal influenza A or B virus infection: Application of the 7-category ordinal scale. *Open Forum Infect Dis* 2019; 6:ofz053
19. Zou KH, O'Malley AJ, Mauri L: Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 2007; 115:654–657
20. Gupta S, Hayek SS, Wang W, et al; STOP-COVID Investigators: Factors associated with death in critically ill patients with coronavirus disease 2019 in the US. *JAMA Intern Med* 2020; 180:1436–1447
21. Chomilo N, Heard-Garris N, DeSilva M, et al: "The Harm of a Colorblind Allocation of Scarce Resources". *Health Affairs Blog*, 2020. Available at: <https://www.healthaffairs.org/doi/10.1377/hblog20200428.904804/full/>. Accessed June 30, 2020
22. Christian MD, Hamielec C, Lazar NM, et al: A retrospective cohort pilot study to evaluate a triage tool for use in a pandemic. *Crit Care* 2009; 13:R170
23. Khan Z, Hulme J, Sherwood N: An assessment of the validity of SOFA score based triage in H1N1 critically ill patients during an influenza pandemic. *Anaesthesia* 2009; 64:1283–1288
24. Miller WD, Peek ME, Parker WF: Scarce resource allocation scores threaten to exacerbate racial disparities in health care. *Chest* 2020; 158:1332–1334
25. Grasselli G, Zangrillo A, Zanella A, et al; COVID-19 Lombardy ICU Network: Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region, Italy. *JAMA* 2020; 323:1574–1581
26. 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