

Title:

SOFA score performs worse than age for predicting mortality in patients with COVID-19

Short title:

SOFA score performs worse than age for predicting COVID mortality

Raphael A. G. Sherak<sup>1</sup>, Hoomaan Sajjadi<sup>2</sup>, Naveed Khimani<sup>2</sup>, Benjamin Tolchin<sup>3,4</sup>, Karen Jubanyik<sup>1</sup>, R. Andrew Taylor<sup>1</sup>, Wade Schulz<sup>5,6</sup>, Bobak J. Mortazavi<sup>2</sup>, Adrian D Haimovich<sup>1</sup>

<sup>1</sup> Yale Department of Emergency Medicine, Yale School of Medicine, 333 Cedar St, New Haven, CT, 06511

<sup>2</sup> Department of Computer Science and Engineering, Texas A&M Univ, College Station, TX, 77843

<sup>3</sup> Department of Neurology, Yale School of Medicine, 15 York Street St, New Haven, CT, 06510

<sup>4</sup> Yale New Haven Health Center for Clinical Ethics, 15 York Street, New Haven, CT, 06510

<sup>5</sup> Center for Outcomes Research and Evaluation, Yale–New Haven Hospital, 195 Church Street, New Haven, CT, 06510

<sup>6</sup> Department of Laboratory Medicine, Yale School of Medicine, 330 Cedar Street, New Haven, CT 06510

Abstract

The use of the Sequential Organ Failure Assessment (SOFA) score, originally developed to describe disease morbidity, is commonly used to predict in-hospital mortality. During the COVID-19 pandemic, many protocols for crisis standards of care used the SOFA score to select patients to be deprioritized due to a low likelihood of survival. A prior study found that age outperformed the SOFA score for mortality prediction in patients with COVID-19, but was limited to a small cohort of intensive care unit (ICU) patients and did not address whether their findings were unique to patients with COVID-19. Moreover, it is not known how well these measures perform across races.

In this retrospective study, we compare the performance of age and SOFA scores in predicting in-hospital mortality across two cohorts: a cohort of 2,648 consecutive adult

patients diagnosed with COVID-19 who were admitted to a large academic health system in the northeastern United States over a 4-month period in 2020 and a cohort of 75,601 patients admitted to one of 335 ICUs in the eICU database between 2014 and 2015.

Among the COVID-19 cohort, age (area under receiver-operating characteristic curve (AU-ROC) 0.795, 95% CI 0.762, 0.828) had a significantly better discrimination than SOFA score (AU-ROC 0.679, 95% CI 0.638, 0.721) for mortality prediction. Conversely, age (AU-ROC 0.628 95% CI 0.608, 0.628) underperformed compared to SOFA score (AU-ROC 0.735, 95% CI 0.726, 0.745) in non-COVID-19 ICU patients in the eICU database. There was no difference between Black and White COVID-19 patients in performance of either age or SOFA Score. Our findings bring into question the utility of SOFA score-based resource allocation in COVID-19 crisis standards of care.

## Author Summary

The COVID-19 pandemic has prompted hospitals to develop protocols for allocating resources if the number of patients exceed their capacity in order to save as many lives as possible. Many of these protocols use the Sequential Organ Failure Assessment (SOFA) score to identify patients who are unlikely to survive and thus should be deprioritized for care. There are concerns that the SOFA score may not accurately predict mortality in patients with COVID-19 or perform better in one racial group over another. We asked whether a simple measure, patient age, could better predict mortality than SOFA score in a group of adult patients admitted to a large academic health system in 2020. To see if any findings are unique to patients with COVID-19, we performed the same analysis in a group of adult patients taken from the eICU database, a large publicly available dataset that was collected prior to the COVID-19 pandemic. We found that age was better than SOFA score at predicting patient mortality in patients with COVID-19, but not in patients without COVID. For COVID-19, neither age or SOFA score performed better in one racial group over another. Caution is needed when applying an established disease severity index model to a new illness.

## Introduction

The Sequential Organ Failure Assessment (SOFA) score was developed in 1994 by a European Society of Intensive Care Medicine working group to objectively quantify the degree of organ dysfunction and failure in intensive care unit (ICU) patients with sepsis (1). The SOFA score assigns a value of 0 to 4 to the dysfunction of 6 organ systems (respiratory, coagulation, hepatic, cardiovascular, neurologic, and renal), with higher numbers indicating more dysfunction. The score is calculated using the worst clinical values observed in the previous 24 hours. It may be used to describe each organ system or as a summative measure.

Although the initial intention of the SOFA score was to describe morbidity in ICU patients with sepsis, subsequent studies have used the SOFA score or SOFA score-based models to predict mortality (2–4). While the SOFA score on admission performed well for mortality prediction in sepsis in some studies (5,6), other studies have suggested only intermediate discriminatory accuracy (7,8). The use of SOFA score has also expanded beyond patients with sepsis (9,10) and is used outside of the ICU (6,11).

With COVID-19 positive patients occupying up to 90% of ICU beds during the global SARS-CoV-2 pandemic (12), the SOFA score gained new applications. One study reported that 20 of the 26 COVID-19 ventilator triage policies surveyed used the SOFA score (13). Many of these guidelines for crisis standards of care involve withholding resources from patients with an expected low likelihood of survival, with the rationale of saving the most lives possible with the limited resources available. One way this is done is by assuming patients with a SOFA score at or above a predetermined threshold are unlikely to survive the hospitalization. However, there is a dearth of data about the use of the SOFA score to predict mortality in COVID-19, potentially leading to improper allocation of resources (14,15).

One prominent study of 675 COVID-19 positive patients from a single healthcare system found that SOFA score had inferior discriminant prognostic accuracy for in-hospital mortality compared to patient age (16). However, the study was limited by its small sample size with missing data for approximately 25% of its cohort and restriction to patients requiring mechanical ventilation. Moreover, the study population had higher and less variable SOFA scores than the average ICU population (6,7).

Here, we expand on prior work of SOFA score mortality prediction in COVID-19 by assessing SOFA score performance for an undifferentiated population of admitted patients. The primary aim of this retrospective study was to assess how well SOFA score predicts in-hospital mortality in a consecutive cohort of patients with COVID-19 admitted to a quaternary medical center in the Northeast United States (US). We compare the discriminative performance of SOFA to age alone. To examine whether any findings are specific to patients with COVID-19, we contrast these findings to a large cohort of general ICU patients from the publicly accessible eICU database. As a secondary objective, given concerns that prediction models can perpetuate systemic inequities (17–22), we compared the prognostic value of SOFA across race in both cohorts.

## Methods

This was a retrospective study comprising two separate patient cohorts.

### Data Acquisition and Preprocessing:

#### COVID-19 Cohort

Patients in the COVID-19 cohort consisted of all patients with an age  $\geq 18$  admitted to any of the 5 hospitals in the Yale-New Haven Health System (YNHH) from March 29th, 2020 to August 1, 2020 with a diagnosis of COVID-19, defined as either a positive PCR test for COVID-19 or designated as a COVID-19 patient by an attending physician. Data was obtained retrospectively from the YNHH electronic medical record (EMR, Epic Systems Corporation, Verona, WI). SOFA score was automatically calculated for all

admitted patients and recorded every 4 hours by the YNHH EMR. Patients were excluded if they did not have a SOFA score recorded in the first 24 hours of admission or if they had a prior admission with a COVID-19 diagnosis. This study was exempted by the Yale University Institutional Review Board (2000027747). The patient cohort is described extensively in a prior, publicly available manuscript (23).

### eICU Cohort

The eICU Collaborative Research Database is a publicly available database containing data from over 139,000 patients hospitalized between 2014 and 2015 at one of the 335 American ICUs participating Philips Healthcare eICU program (24).

Patients with an ICU stay of at least 24 hours were included. Also, for the patients having more than one hospitalization, only the most recent hospitalization was used.

The SOFA scores on eICU were calculated and extracted using Python's Pandas library, based on its standard definition on the values of creatinine, bilirubin, platelets, fraction of inspired oxygen, partial pressure of inspired oxygen, Glasgow Coma Scale, mean arterial pressure, and mechanical ventilation status. As part of the eICU patient de-identification process, all patients with age >89 are grouped together and were assigned an age of 90 for the analysis. Then, the corresponding tables and columns were pre-processed to calculate the final scores. The overall steps taken in this procedure are described in the supplement.

In both cohorts, gender was dichotomized and race was classified as a single category based on predefined fields in the EMR. Depending on a patient's clinical status, gender and race were either self-selected by patients or assigned by hospital registration staff.

### Logistic Regression Model Development:

Age in years at time of admission and the maximum SOFA score recorded in the first 24 hours after admission were each used as predictor variables in separate univariate logistic regression models for the binary outcome of in-hospital mortality. Each model was fit to a sample of 60% of the respective cohorts.

## Model Assessment:

The remaining 40% of each cohort were used as a validation cohort to calculate area under the receiver operator characteristic curves (AU-ROCs). AU-ROC intervals were estimated using DeLong's method (25). This method was then repeated to calculate area under precision-recall curves (AU-PRCs). All analyses were performed in Python (Version 3.7.7) and R (Version 1.4.1717).

We performed several secondary analyses. We stratified each cohort by race including only patients identifying as Black/African American or White/Caucasian as we had insufficient sample sizes for further groups. Because the eICU cohort is restricted to patients admitted to the ICU, we performed an additional analysis of the COVID-19 cohort restricted to ICU patients. We also conducted an exploratory analysis of the COVID-19 cohort of survival rates at and above a given SOFA score. We then stratified by race to explore whether there was a difference in survival rates at various SOFA score thresholds. The binomial exact test was used to calculate confidence intervals.

## Results

### Patient characteristics:

A total of 2,648 patients were included in the COVID-19 cohort with an average age of 63 (SD 19.3), mean SOFA score of 2.1 (SD 2.7) and an in-hospital mortality rate of 17.1%. The eICU cohort consisted of 75,601 patients with an average age of 63 (SD 17.6), mean SOFA score of 7.7 (SD 2.7), and an in-hospital mortality rate of 10.8%. The Yale and eICU cohorts respectively were 52.4% and 46.1% female and 25.4% and 10.7% Black/African American (Table 1).

Table 1 (Demographics for COVID-19 and eICU Cohorts):

COVID-19 Cohort				eICU Cohort				
		Missing	Total				Missing	Total
n			2648	n				75601
Survival to Hospital Discharge, n (%)		0	2196 (82.9)	Survival to Hospital Discharge, n (%)			0	67463 (89.2)
SOFA Score, mean (SD)		0	2.1 (2.7)	SOFA Score, mean (SD)			0	7.7 (2.7)
Gender, n (%)	Female	0	1387 (52.4)	Gender, n (%)	Female		12	34849 (46.1)
	Male		1261 (47.6)		Male			40740 (53.9)
Age, mean (SD)		0	63.4 (19.3)	Age, mean (SD)				62.9 (17.6)
ICU, n (%)		0	642 (24.2)	ICU, n (%)			0	75601 (100)
Race (%)	Asian	2	50 (1.9)	Ethnicity, n (%)	Asian		0	1444 (1.9)
	Black/African American		673 (25.4)		Black/African American			8100 (10.7)

	Native American or Alaska Native		3 (0.1)		Native American or Alaska Native			561 (0.7)
	Native Hawaiian		1 (0.0)		Hispanic			3116 (4.1)
	Other Pacific Islander		8 (0.3)					
	Other/ Not Listed		603 (22.7)					
	Patient Refused		16 (0.6)					
	Unknown		25 (0.9)		Other/ Unknown			3990 (5.3)
	White/ Caucasian		1267 (47.9)		White/ Caucasian			58390 (77.2)
Insurance, n (%)	Medicaid	0	492 (18.6)					
	Medicare		891 (33.6)					
	Private		1082 (40.9)					
	Uninsured		183 (6.9)					



### Primary Outcome:

Among the COVID-19 cohort, age (AU-ROC 0.795, 95% CI 0.762, 0.828) had a significantly better discrimination than SOFA score (AU-ROC 0.679, 95% CI 0.638, 0.721) for mortality prediction. When restricted to patients admitted to the ICU, age (AU-ROC 0.698, 95% CI 0.631, 0.765) still generally performed better than SOFA score (AU-ROC 0.562, 95% CI 0.486, 0.639), for mortality prediction in COVID-19 patients. Conversely, age (AU-ROC 0.628 95% CI 0.608, 0.628) significantly underperformed compared to SOFA score (AU-ROC 0.735, 95% CI 0.726, 0.745) in non-COVID-19 eICU patients (Table 2).

Table 2 (Model Results):

	Predictor	AU-ROC (95% CI)	AU-PRC
COVID-19 Cohort	Age	0.795 (0.762, 0.828)	0.387
	SOFA Score	0.679 (0.638, 0.721)	0.289
COVID-19 ICU Sub-Cohort	Age	0.698 (0.631, 0.765)	0.455
	SOFA Score	0.562 (0.486, 0.639)	0.375
eICU Cohort	Age	0.628 (0.608, 0.628)	0.177
	SOFA Score	0.735 (0.726, 0.745)	0.320

Secondary Analyses:

When stratified by race, age performed similarly between Black and White patients in both cohorts (Figure 3 and Table 3). There also was no significant difference in the performance of the SOFA score between Black and White patients with COVID-19. In the eICU cohort, SOFA score was better than age at discriminating between survivors and non-survivors both in Black and White patients. Additionally, the SOFA score performed better in Black patients compared to White patients in the eICU cohort.

Table 3: (AU-ROC of Age and SOFA Score for Mortality Prediction Stratified by Cohort and Race)

	AU-ROC (95% CI)	
	Black Patients	White Patients
<b>COVID-19 Cohort</b>		
Age	0.82 (0.734, 0.886)	0.782 (0.737, 0.826)
SOFA score	0.727 (0.645, 0.808)	0.651 (0.594, 0.708)
<b>eICU Cohort</b>		
Age	0.604 (0.571, 0.638)	0.619 (0.608, 0.630)
SOFA score	0.808 (0.781, 0.834)	0.726 (0.715, 0.738)

In the COVID-19 cohort, there was no significant difference in survival rates between Black and White patients at and above SOFA scores of 3 (Supplement Figure 2). The survival rate of COVID-19 patients only dropped below 50% at SOFA scores greater than or equal to 12 (Supplement Figure 3).

#### Discussion

Age significantly outperformed SOFA score for predicting mortality in hospitalized COVID-19 patients, including those in the ICU. This phenomenon may be unique to COVID-19 as SOFA score was significantly better at predicting mortality in the eICU cohort. The finding that a simple metric such as age outperformed SOFA score for mortality prediction in COVID-19 patients suggests that caution should be taken when applying established prediction models to a completely novel disease process. This is especially prudent when using mortality prediction models to guide treatment decisions

and resource allocation. Many guidelines for crisis standards of care suggested a SOFA score threshold of  $\geq 6$  to identify patients with low likelihood of survival (26,27). However, there was only a 33% mortality rate using that threshold in our COVID-19 cohort.

Although the SOFA score was originally created for ICU patients (1), the inclusion of non-ICU patients in the COVID-19 cohort does not explain our findings. When restricted to COVID-19 cohort to ICU patients, a similar trend was noted. Age performed better than SOFA score for mortality prediction, with minimally overlapping confidence intervals. However, these results were limited by the relatively small sample size of COVID-19 ICU patients (642). While applying SOFA score to non-ICU patients is outside of the original intent of SOFA score, so is using it for mortality prediction. In the original study describing the SOFA score, Vincent et al. state, "it is important to realize that the SOFA score is designed not to *predict* outcome but to *describe* a sequence of complications"(1).

Prior studies have theorized that SOFA score underperforms for mortality prediction in COVID-19 patients because the illness affects fewer organ systems resulting in lower variability in scores (16). Contradictory to that theory, patients in the COVID-19 cohort had the same standard deviation in scores (2.7) as the eICU cohort. Since age  $>65$  is one of the characteristics with the strongest association with increased mortality in COVID-19 positive patients (28,29), it may partially explain why age outperforms SOFA score only in COVID-19 patients.

There was no difference between White and Black COVID-19 patients in the performance of either age or SOFA score for mortality prediction. Additionally, there was no significant difference in mortality rate between White and Black COVID-19 patients at SOFA scores greater than 2. Notably, a prior study of patients from the same COVID-19 cohort, Black patients had 1.5 times the odds of a SOFA score  $\geq 6$  than white patients, even when adjusting for age, sex, insurance status, BMI, liver and renal diseases (23). This suggests that Black patients with COVID-19 may be more likely than

White patients to be assigned higher SOFA scores but will have similar mortality rates at those higher SOFA scores.

This study builds on a recent analysis from a single health system of 675 COVID-19 positive patients requiring mechanical ventilation in which SOFA score had inferior prognostic accuracy for in-hospital mortality compared to simply using patient age (16). However, the study was limited by its cohort having higher and less variable SOFA scores than the average ICU population as well as missing data for approximately 25% of the cohort. Our study had over 2,600 COVID-19 positive patients with an average SOFA score similar to prior analyses on non-COVID-19 patients (6,7) and a mortality rate comparable to other cohorts of COVID-19 patients that were admitted to US hospitals during a similar time period (28). Moreover, the SOFA scores used in the COVID-19 cohort were automatically calculated by the electronic health record, a pragmatic approach that would be used in triage scenarios.

Our study has several limitations. The cohort of COVID-19 patients was restricted to a single health system in the northeast US which may not have a comparable population to COVID-19 patients at other academic health systems or those in the eICU database. Due to sample size, we also restricted our analysis to two racial groups and did not consider patients that identified as multi-racial. We did not consider sex or its potential interaction with race or age in our study. Furthermore, an unknown proportion of patients in the COVID-19 cohort were too critically ill to answer demographic questions and had their race and sex recorded by a hospital clerk based on assumption. Additionally, this study only describes the performance of max SOFA score within 24 hours of hospital admission. Many COVID-19 positive patients present to the hospital with respiratory complaints and develop multisystem organ dysfunction later in their disease course (29). However, information on SOFA subscores in the cohort of COVID-19 patients were not available so we were unable to test this hypothesis. SOFA score may have greater utility for predicting mortality with serial measurements or later in disease course (2).

As the COVID-19 pandemic continues well into 2022, SOFA score continues to feature prominently in guidelines for crisis standards of care (30). This study suggests caution should be used when considering SOFA score as a prognostic tool, as it has limited prognostic performance.

## References

1. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, 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* [Internet]. 1996 Jul;22(7):707–10. Available from: <http://dx.doi.org/10.1007/BF01709751>
2. Ferreira FL, Bota DP, Bross A, Mélot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA* [Internet]. 2001 Oct 10;286(14):1754–8. Available from: <http://dx.doi.org/10.1001/jama.286.14.1754>
3. Vincent J-L, de Mendonca A, Cantraine F, Moreno R, Takala J, Suter PM, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: Results of a multicenter, prospective study. *Crit Care Med* [Internet]. 1998 Nov [cited 2021 Mar 30];26(11):1793. Available from: [https://journals.lww.com/ccmjournal/Fulltext/1998/11000/Use\\_of\\_the\\_SOFA\\_score\\_to\\_assess\\_the\\_incidence\\_of.16.aspx](https://journals.lww.com/ccmjournal/Fulltext/1998/11000/Use_of_the_SOFA_score_to_assess_the_incidence_of.16.aspx)
4. Cárdenas-Turanzas M, Ensor J, Wakefield C, Zhang K, Wallace SK, Price KJ, et al. Cross-validation of a Sequential Organ Failure Assessment score-based model to predict mortality in patients with cancer admitted to the intensive care unit. *J Crit Care* [Internet]. 2012 Dec;27(6):673–80. Available from: <http://dx.doi.org/10.1016/j.jcrc.2012.04.018>
5. Jones AE, Trzeciak S, Kline JA. The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation. *Crit Care Med* [Internet]. 2009 May;37(5):1649–54. Available from: <http://dx.doi.org/10.1097/CCM.0b013e31819def97>
6. Kovach CP, Fletcher GS, Rudd KE, Grant RM, Carlbom DJ. Comparative prognostic accuracy of sepsis scores for hospital mortality in adults with suspected infection in non-ICU and ICU at an academic public hospital. *PLoS One* [Internet]. 2019 Sep 16;14(9):e0222563. Available from: <http://dx.doi.org/10.1371/journal.pone.0222563>
7. Raith EP, Udy AA, Bailey M, McGloughlin S, Maclsaac C, Bellomo R, et al. 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* [Internet]. 2017 Jan 17;317(3):290–300. Available from: <http://dx.doi.org/10.1001/jama.2016.20328>
8. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* [Internet]. 2016 Feb 23;315(8):762–74. Available from: <http://dx.doi.org/10.1001/jama.2016.0288>
9. Minne L, Abu-Hanna A, de Jonge E. Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. *Crit Care* [Internet]. 2008 Dec 17;12(6):R161. Available from: <http://dx.doi.org/10.1186/cc7160>
10. Kashyap R, Sherani KM, Dutt T, Gnanapandithan K, Sagar M, Vallabhajosyula S, et al. Current Utility of Sequential Organ Failure Assessment Score: A Literature Review and

- Future Directions. *Open Respir Med J* [Internet]. 2021 Apr 13;15:1–6. Available from: <http://dx.doi.org/10.2174/1874306402115010001>
11. Yu S, Leung S, Heo M, Soto GJ, Shah RT, Gunda S, et al. Comparison of risk prediction scoring systems for ward patients: a retrospective nested case-control study. *Crit Care* [Internet]. 2014 Jun 26;18(3):R132. Available from: <http://dx.doi.org/10.1186/cc13947>
  12. Douin DJ, Ward MJ, Lindsell CJ, Howell MP, Hough CL, Exline MC, et al. ICU Bed Utilization During the Coronavirus Disease 2019 Pandemic in a Multistate Analysis-March to June 2020. *Crit Care Explor* [Internet]. 2021 Mar;3(3):e0361. Available from: <http://dx.doi.org/10.1097/CCE.0000000000000361>
  13. Antommaria AHM, Gibb TS, McGuire AL, Wolpe PR, Wynia MK, Applewhite MK, et al. 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* [Internet]. 2020 Aug 4;173(3):188–94. Available from: <http://dx.doi.org/10.7326/M20-1738>
  14. Shahpori R, Stelfox HT, Doig CJ, Boiteau PJE, Zygun DA. Sequential Organ Failure Assessment in H1N1 pandemic planning. *Crit Care Med* [Internet]. 2011 Apr;39(4):827–32. Available from: <http://dx.doi.org/10.1097/CCM.0b013e318206d548>
  15. 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* [Internet]. 2009 Dec;64(12):1283–8. Available from: <http://dx.doi.org/10.1111/j.1365-2044.2009.06135.x>
  16. Raschke RA, Agarwal S, Rangan P, Heise CW, Curry SC. Discriminant Accuracy of the SOFA Score for Determining the Probable Mortality of Patients With COVID-19 Pneumonia Requiring Mechanical Ventilation. *JAMA* [Internet]. 2021 Feb 17; Available from: <http://dx.doi.org/10.1001/jama.2021.1545>
  17. Tolchin B, Hull SC, Kraschel K. Triage and justice in an unjust pandemic: ethical allocation of scarce medical resources in the setting of racial and socioeconomic disparities. *J Med Ethics* [Internet]. 2020 Oct 16; Available from: <http://dx.doi.org/10.1136/medethics-2020-106457>
  18. Egede LE, Walker RJ. Structural Racism, Social Risk Factors, and Covid-19 - A Dangerous Convergence for Black Americans. *N Engl J Med* [Internet]. 2020 Sep 17;383(12):e77. Available from: <http://dx.doi.org/10.1056/NEJMp2023616>
  19. Miller WD, Han X, Peek ME, Charan Ashana D, Parker WF. Accuracy of the Sequential Organ Failure Assessment Score for In-Hospital Mortality by Race and Relevance to Crisis Standards of Care. *JAMA Netw Open* [Internet]. 2021 Jun 1;4(6):e2113891. Available from: <http://dx.doi.org/10.1001/jamanetworkopen.2021.13891>
  20. Ashana DC, Anesi GL, Liu VX, Escobar GJ, Chesley C, Eneanya ND, et al. Equitably Allocating Resources During Crises: Racial Differences in Mortality Prediction Models. *Am J Respir Crit Care Med* [Internet]. 2021 Mar 22; Available from: <http://dx.doi.org/10.1164/rccm.202012-4383OC>
  21. Gershengorn HB, Holt GE, Rezk A, Delgado S, Shah N, Arora A, et al. Assessment of Disparities Associated With a Crisis Standards of Care Resource Allocation Algorithm for Patients in 2 US Hospitals During the COVID-19 Pandemic. *JAMA Netw Open* [Internet].



- 2021 Mar 1;4(3):e214149. Available from: <http://dx.doi.org/10.1001/jamanetworkopen.2021.4149>
22. Sarkar R, Martin C, Mattie H, Gichoya JW, Stone DJ, Celi LA. Performance of intensive care unit severity scoring systems across different ethnicities in the USA: a retrospective observational study. *Lancet Digit Health* [Internet]. 2021 Apr;3(4):e241–9. Available from: [http://dx.doi.org/10.1016/S2589-7500\(21\)00022-4](http://dx.doi.org/10.1016/S2589-7500(21)00022-4)
  23. Tolchin B, Oladele C, Galusha D, Kashyap N, Showstark M, Bonito J, et al. Racial disparities in the SOFA score among patients hospitalized with COVID-19. *PLoS One* [Internet]. 2021 Sep 17;16(9):e0257608. Available from: <http://dx.doi.org/10.1371/journal.pone.0257608>
  24. 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* [Internet]. 2018 Sep 11;5:180178. Available from: <http://dx.doi.org/10.1038/sdata.2018.178>
  25. Sun X, Xu W. Fast Implementation of DeLong's Algorithm for Comparing the Areas Under Correlated Receiver Operating Characteristic Curves. *IEEE Signal Process Lett* [Internet]. 2014 Nov;21(11):1389–93. Available from: <http://dx.doi.org/10.1109/LSP.2014.2337313>
  26. White DB, Lo B. A Framework for Rationing Ventilators and Critical Care Beds During the COVID-19 Pandemic. *JAMA* [Internet]. 2020 May 12;323(18):1773–4. Available from: <http://dx.doi.org/10.1001/jama.2020.5046>
  27. Tolchin B, Latham SR, Bruce L, Ferrante LE, Kraschel K, Jubanyik K, et al. Developing a Triage Protocol for the COVID-19 Pandemic: Allocating Scarce Medical Resources in a Public Health Emergency. *J Clin Ethics* [Internet]. 2020 Winter;31(4):303–17. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/32991327>
  28. Petrilli CM, Jones SA, Yang J, Rajagopalan H, O'Donnell L, Chernyak Y, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ* [Internet]. 2020 May 22;369:m1966. Available from: <http://dx.doi.org/10.1136/bmj.m1966>
  29. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *JAMA* [Internet]. 2020 Aug 25;324(8):782–93. Available from: <http://dx.doi.org/10.1001/jama.2020.12839>
  30. ID\_CSC\_Standards\_2021\_V3.pdf. Available from: [https://coronavirus.idaho.gov/wp-content/uploads/2021/02/ID\\_CSC\\_Standards\\_2021\\_V3.pdf](https://coronavirus.idaho.gov/wp-content/uploads/2021/02/ID_CSC_Standards_2021_V3.pdf)

# Figures

Figure 1 (Distribution of SOFA Scores):



medRxiv preprint doi: <https://doi.org/10.1101/2022.05.02.22274575>; this version posted May 3, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Figure 2 (AU-ROCs and AU-PRCs of Age and SOFA Score for Mortality Prediction):

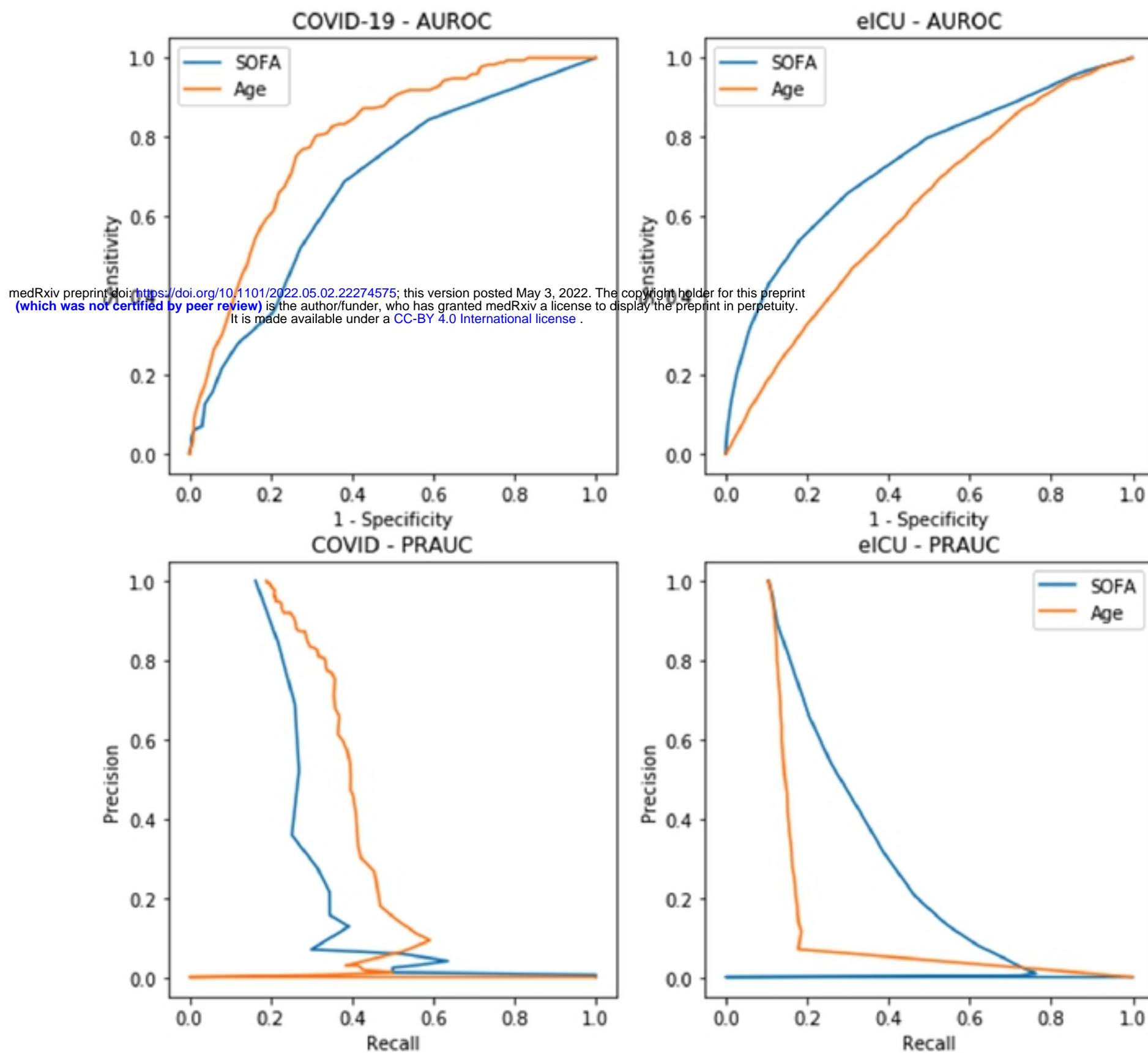


Figure 3 (Predictor Performance Stratified by Cohort and Race):

