SOFA Produces Uncertain Mortality Predictions: Reconsidering the Use of SOFA in Triage

Katarina Pejcinovic MS[1], Brenna Park-Egan MS[1], Patrick Lyons MD[1]
Oregon Health & Science University

Introduction

During crises such as pandemics, there can be shortages of life-saving resources. Many hospitals in these situations triage patients for scarce resources based on guidelines prioritizing expected survivability.[1] These guidelines frequently use models that calculate survival probability, such as the Sequential Organ Failure Assessment (SOFA).[1]

For example, a SOFA score of 11 has been associated with a 90% chance of mortality,[2] and patients with scores above these thresholds are often deprioritized. This triaging is done in hopes of saving the greatest number of lives. See Table 1 for SOFA score calculation in practice.

But what if the model is uncertain in its predictions?

If SOFA is highly uncertain in predicting survival, it could mean that patients are incorrectly ranked for triage. This could ultimately lead to fewer patients surviving.

Our goal was to quantify the uncertainty in SOFA-based model predictions at the patient level

Score	0	1	2	3	4
Respiratory	PaO2/FiO2 > 400 SpO2/FiO2 > 302	PaO2/FiO2 < 400 SpO2/FiO2 < 302	PaO2/FiO2 < 300 SpO2/FiO2 < 221	PaO2/FiO2 < 200 SpO2/FiO2 < 142	PaO2/FiO2 < 100 SpO2/FiO2 < 67
Cardiovascular (dose in mcg/kg/min)	MAP ≥ 70 mm Hg	MAP ≥ 70 mm Hg	Dopamine ≤ 5 or ANY dobutamine	Dopamine > 5 or Norepinephrine ≤ 0.1 or Phenylephrine ≤ 0.8	Dopamine > 15 or Norepinephrine > 0.1 or Phenylephrine > 0.8
Liver (bilirubin, mg/dL)	< 1.2	1.2–1.9	2.0–5.9	6.0–11.9	> 12
Renal (creatinine, mg/dL)	< 1.2	1.2–1.9	2.0–3.4	3.5–4.9	> 5.0
Coagulation (platelets × 10^3/mm^3)	≥ 150	< 150	< 100	< 50	< 20
Neurologic (GCS score)	15	13–14	10–12	6–9	< 6

Table 1.[3] Table for SOFA score calculation. Points are added for each organ system based on clinical features. Scores range from 0-24.

The Data

This study involved 12,405 ICU patient encounters from a single academic hospital between 2020 and 2022. All encounters had circulatory or respiratory failure, 9864 of which ended in an outcome of survival. Each patient encounter had the clinical features required to calculate the SOFA score (e.g., platelets, creatinine), as well as their final calculated SOFA scores, based on data accrued within 48 hours before critical illness.

Methodology

Conformal Prediction is an uncertainty quantification method that outputs prediction sets guaranteed to contain the correct answer with a specified confidence level. The contents of this prediction set help determine our confidence in the model prediction.

For example, if a patient receives an output prediction set at 90% confidence (α =0.1) of:

- {survival} = 90% confidence that survival is the correct answer
- {non-survival} = 90% confidence that non-survival is the correct answer
- {survival, non-survival} = cannot determine the correct answer with 90% confidence

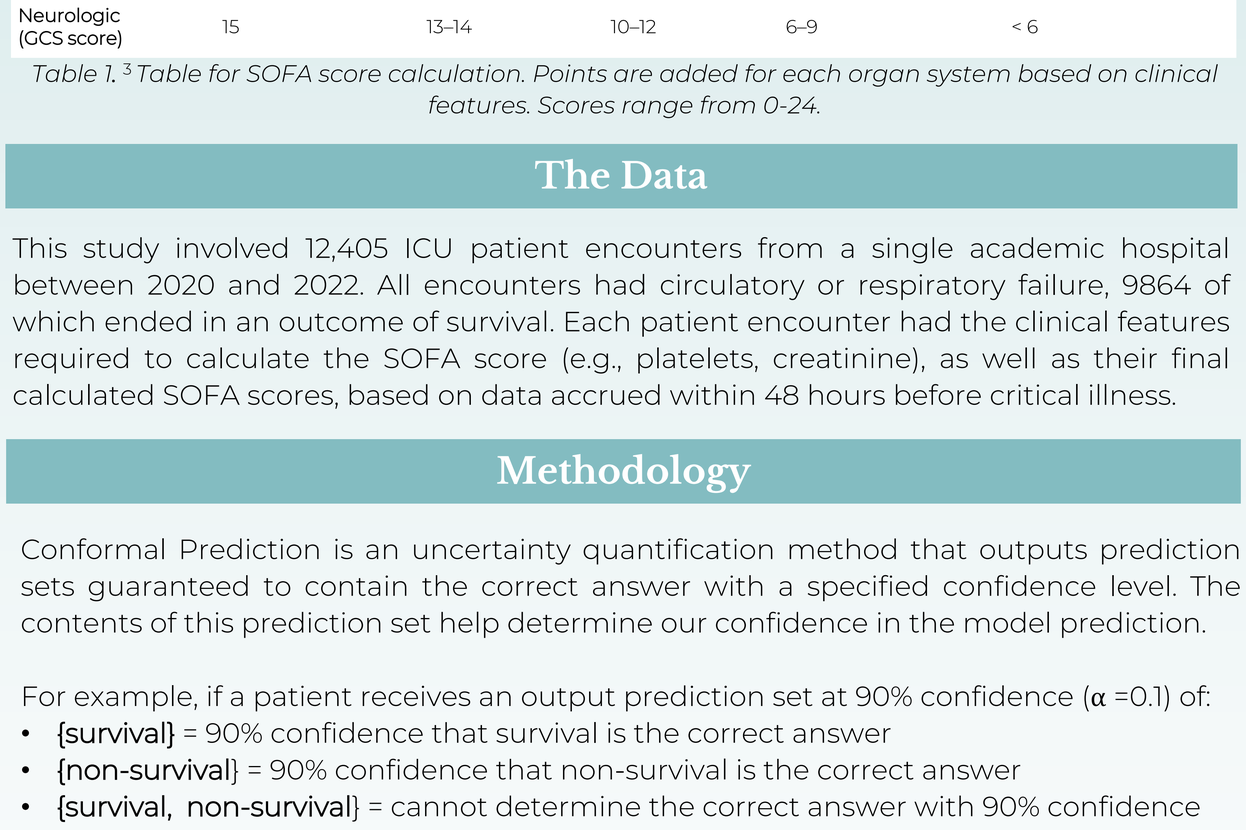

Figure 1. Overview of Study Pipeline. The data, represented at the top, were used to train two separate ML algorithms. SOFA scores were used to train the Logistic Regression, whereas underlying lab values were used to train the Random Forest. Conformal Prediction was applied to both methods, and a prediction set was produced for each patient encounter

Results

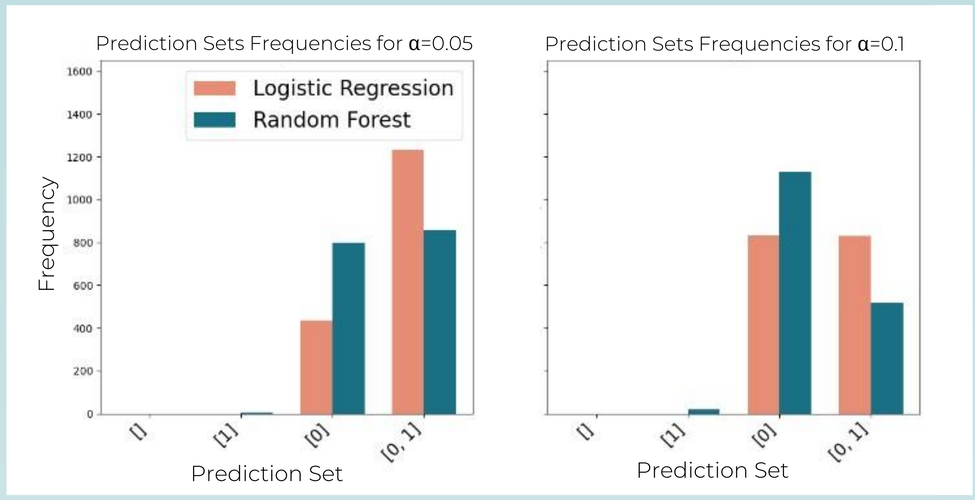

Figure 2. Number of patient encounters receiving each type of prediction set, where 0=survival and 1=non-survival

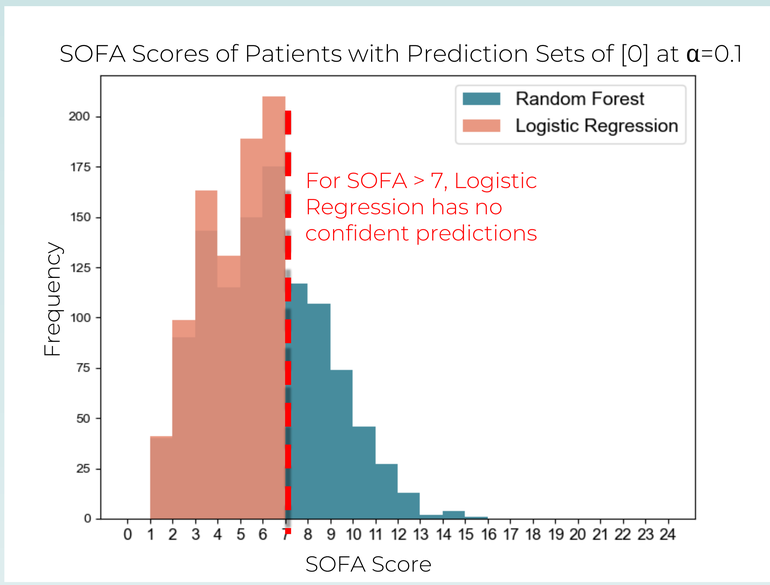

Figure 3. Underlying SOFA scores for patient encounters with 90% certain survival predictions for each model

Analysis

- The SOFA-based model consistently had more patients be assigned the uncertain prediction set of {survival, non-survival} when compared to the Random Forest (Figure 2).
- The SOFA-based model produced no prediction sets that only contained {non-survival} (Figure 2). This means not a single patient encounter could be confidently classified as non-survival.
- The comparator model can confidently classify over a wider range of SOFA values compared to the SOFA model (Figure 3). The SOFA-based model is unable to definitively assign survival or non-survival at an α=0.1 for SOFA scores above 7. A similar trend is observed for α=0.05 (cutoff score of 5), and α=0.2 (cutoff score of 13).

Conclusions and Future Directions

Our findings suggest that SOFA is an uncertain predictor of mortality, especially at currently accepted SOFA score thresholds.

- The SOFA-based model consistently produces a higher number of uncertain predictions and is unable to confidently classify survival for patients with a SOFA score above 7 with 90% confidence.
- We show that a simple, Random Forest model that requires no additional information can outperform SOFA scores in terms of uncertainty.

These findings, along with other shortcomings of the SOFA Score, call into question the use of SOFA for triage

Future directions

- Stratifying by sensitive groups (e.g., age, race, sex)
- Comparison with other academic hospital datasets

References and Acknowledgements

1. Piscitello GM, Kapania EM, Miller WD, Rojas JC, Siegler M, Parker WF. Variation in ventilator allocation guidelines by US state during the coronavirus disease 2019 pandemic: a systematic review. JAMA Netw Open. 2020 Jun 1;3(6):e2012606.
2. 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 Jul;22(7):707–10.
3. Updates and controversies in the early management of sepsis and septic shock | eb medicine [Internet]. [cited 2025 Apr 24]. Available from: https://www.ebmedicine.net/topics.php?paction=showTopic&topic_id=587

Special thanks to the National Library of Medicine (grant T15LM007088) for funding this project.

Communications

Email: Pejcinovic@ohsu.edu