

Prediction of in-hospital mortality among intensive care unit patients using modified daily Laboratory-based Acute Physiology Scores, version 2 (LAPS2)

Rachel Kohn, MD, MSCE,¹⁻³ Gary E. Weissman, MD, MSHP,¹⁻³ Wei Wang, PhD,² Nicholas E. Ingraham, MD,⁴ Stefania Scott, MS,² Brian Bayes, MS,² George L. Anesi, MD, MSCE,¹⁻³ Scott D. Halpern, MD, PhD,^{1-3,5,6} Patricia Kipnis, PhD,⁷ Vincent X. Liu, MD, MS,⁷ R. Adams Dudley, MD, MBA,⁴ Meeta Prasad Kerlin, MD, MSCE¹⁻³

¹Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, ²Palliative and Advanced Illness Research (PAIR) Center at the University of Pennsylvania, Philadelphia, Pennsylvania, ³Leonard Davis Institute of Health Economics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, ⁴Department of Medicine, University of Minnesota, ⁵Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine at the University of Pennsylvania, ⁶Department of Medical Ethics and Health Policy, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, ⁷Division of Research, Kaiser Permanente, Oakland, California

Corresponding author:

Rachel Kohn, MD, MSCE
Hospital of the University of Pennsylvania
836 Gates Building
3400 Spruce Street
Philadelphia, PA 19104
Phone: 215-908-1037
rachel.kohn2@penmedicine.upenn.edu

Authors' contributions: All of the authors made substantial contributions to the conception or design of the study, or the acquisition, analysis, or interpretation of data, and drafting or revising the manuscript.

Research funding source: Dr. Kohn is supported by NIH/NHLBI K23 HL146894. Dr. Ingraham is supported by NIH/NHLBI T32 HL07741. Dr. Anesi is supported by NIH/NHLBI K23 HL161353. Dr. Kerlin is supported by NIH/NHLBI R01 HL146386.

Author disclaimers: none

Keywords: prediction models, intensive care units, severity of illness, in-hospital mortality

Abstract word count: 242

Total word count: 1,993

ABSTRACT

Background: Mortality prediction for intensive care unit (ICU) patients frequently relies on single acuity measures based on ICU admission physiology without accounting for subsequent clinical changes.

Objectives: Evaluate novel models incorporating modified admission and daily, time-updating Laboratory-based Acute Physiology Scores, version 2 (LAPS2) to predict in-hospital mortality among ICU patients.

Research design: Retrospective cohort study.

Subjects: All ICU patients in five hospitals from October 2017 through September 2019.

Measures: We used logistic regression, penalized logistic regression, and random forest models to predict in-hospital mortality within 30 days of ICU admission using admission LAPS2 alone in patient-level and patient-day-level models, or admission and daily LAPS2 at the patient-day level. Multivariable models included patient and admission characteristics. We performed internal-external validation using four hospitals for training and the fifth for validation, repeating analyses for each hospital as the validation set. We assessed performance using scaled Brier scores (SBS), c-statistics, and calibration plots.

Results: The cohort included 13,993 patients and 120,101 ICU days. The patient-level model including the modified admission LAPS2 without daily LAPS2 had an SBS of 0.175 (95% CI 0.148-0.201) and c-statistic of 0.824 (95% CI 0.808-0.840). Patient-day-level models including daily LAPS2 consistently outperformed models with modified admission LAPS2 alone. Among patients with <50% predicted mortality, daily models were better calibrated than models with modified admission LAPS2 alone.

Conclusions: Models incorporating daily, time-updating LAPS2 to predict mortality among an ICU population perform as well or better than models incorporating modified admission LAPS2 alone.

INTRODUCTION

Time-updating acuity assessment is critical for clinical decision-making, quality and safety initiatives, hospital performance, and clinical research. Multiple mortality prediction scores incorporating acute physiologic data have been developed and validated among intensive care unit (ICU) patients. However, these scores do not capture changes in acuity over time, which may add predictive value, and those that do tend to have narrow ranges, potentially limiting discrimination.¹⁻⁵

Escobar et al. developed the Laboratory-based Acute Physiology Score, version 2 (LAPS2), a score assigned at hospital admission ranging from 0-414 to predict mortality among hospitalized adults.^{6,7} LAPS2 uniquely includes administrative and physiologic variables from the 72 hours preceding admission, data readily available within the integrated health system in which it was developed. Additionally, LAPS2 utilizes a preliminary model to subdivide the population into low and high mortality risk in order to impute missing data accordingly, rather than assuming normal values as is common in other mortality prediction scores.⁷

LAPS2 performs better than many other mortality prediction models. Furthermore, recent work comparing LAPS2 to the commonly used Sequential Organ Failure Assessment (SOFA) score demonstrated that LAPS2 is more racially equitable,⁸ making it an exemplar model for further expansion and refinement. However, several features limit its utility in ICU populations. First, most hospitals lack integration with outpatient sites to capture the required data from 72 hours prior to hospital presentation. Second, admission acuity does not account for informative changes in patients' physiology. Finally, LAPS2 has not been comprehensively studied among ICU patients, for whom mortality risk is highest and updated prognostication is most critical to decision-making. Therefore, our objective in this study was to develop and compare novel prediction models using daily LAPS2 based exclusively on data collected after hospital presentation to predict in-hospital mortality among ICU patients.

METHODS

Study overview, sites, and population

We created a retrospective cohort of patients ≥ 18 years old admitted to an ICU of four hospitals of one health system between October 1, 2017, and September 30, 2019; and a fifth hospital between June 1, 2018, and September 30, 2019 (when that hospital integrated into the health system). The five hospitals vary in size (approximately 250 to 700 beds), location (two suburban and three urban), and academic affiliation (two primary sites for multiple large residency programs, one primary site for a few small residency programs, and two without trainees). The study was approved by the University of Pennsylvania Institutional Review Board.

Specification of admission and daily LAPS2 scores

For each patient, we estimated a modified LAPS2 for hospital admission adapted from the original specification to include data obtained exclusively after hospital admission (**Supplement**). To align with shift schedules, we defined hospital and ICU days as 7:00am to

6:59am the following calendar day. We estimated admission LAPS2 in two steps. First, we built a logistic regression model (pre-LAPS2 model) including patient age, gender, hospital admission source, and values at the end of the first hospital day of serum sodium, blood urea nitrogen (BUN) to creatinine ratio, and anion gap to serum bicarbonate ratio to predict in-hospital mortality, stratifying patients into high ($\geq 6\%$) or low ($< 6\%$) mortality risk. Second, we used the worst values of all LAPS2 variables from the first hospital day to assign points as in the original LAPS2 specification, imputing a normal value for low mortality risk patients and an abnormal value for high mortality risk patients when values were missing (**Supplement**).⁷ The sum of all assigned points was the *admission LAPS2 score*. We then similarly estimated *daily LAPS2 scores* for each ICU admission day by first building daily pre-LAPS2 models to stratify patients into high or low mortality risk daily, and then using the worst values of all LAPS2 variables from the prior ICU day to assign points to calculate the daily LAPS2 score.

Statistical analysis

To contextualize our results within the existing evidence (i.e., performance of LAPS2 as a single score at hospital admission),⁷ we first built patient-level models using multivariable logistic regression, including hospital admission LAPS2 and covariates similar to the original specification (**Supplement**) to predict in-hospital mortality within 30 days of ICU admission. Next, we built four sets of patient-day-level models including all ICU days. First, we built logistic regression models including LAPS2 for ICU day 1 assigned to all ICU days, to simulate clinical practice where daily decisions are made with knowledge of only the admission LAPS2. Then we built logistic regression, penalized logistic regression using least absolute shrinkage and selection operator (LASSO), and random forest models that included updated daily LAPS2 for each ICU day, in order to evaluate the predictive value of adding updated daily scores and to optimize predictive ability. Hyperparameters for the penalized logistic regression using LASSO were determined by maximizing the 5-times repeated 10-fold cross validated scaled Brier score (SBS) in the training set. The SBS is a composite measure of discrimination and calibration that incorporates the difference between observed and expected outcomes and number of observations and accounts for the event rate, providing a score that is comparable across models, with higher positive values indicating better performance.⁹

For all sets of models, we used internal-external validation, such that four hospitals were included in a training set and the fifth hospital was included in a validation set, and then repeating all analyses with each hospital as the validation set. All models included the hospital admission LAPS2 and covariates similarly to the original specification of LAPS2 (**Supplement**).

We evaluated performance from fitting predicted estimates in the validation sample. The primary performance measure was the SBS. We evaluated discrimination using c-statistics. We additionally evaluated calibration by visual inspection of calibration plots.

We performed three secondary analyses. Under a hypothesis that daily LAPS2 predictive performance deteriorates over time, as physiological changes transition from acute to chronic and contribute less to short-term mortality,¹⁰ we evaluated the predictive performance by

estimating SBS and c-statistics separately for each ICU day. To assess predictive equity, we stratified analyses by gender and race.¹¹

We constructed 95% confidence intervals (CIs) using the standard deviation of estimates from bootstrapping 500 randomly selected samples. All statistical analyses were conducted using Stata version 17 (StataCorp LP, College Station, TX) and R language for statistical computing (version 4.1.2).

RESULTS

The final study population included 13,993 patients and 120,101 ICU days. **Table 1** summarizes patient characteristics and outcomes, and **Supplementary Table 1** summarizes missingness.

The patient-level model including admission LAPS2 without daily LAPS2 had an SBS of 0.175 (95% CI 0.148-0.201) and c-statistic of 0.824 (95% CI 0.808-0.840) with Hospital A as the validation set, with modest variability when validated in the other hospitals (**Supplementary Table 2**).

All models including daily LAPS2 scores performed better than the model that carried forward the single LAPS2 score from the first ICU day (with Hospital A as the validation set: SBS 0.163 [95% CI 0.151-0.174] and c-statistic 0.805 [95% CI 0.800-0.810] compared to SBS 0.096 [95% CI 0.088-0.105] and c-statistic 0.749 [95% CI 0.743-0.755]) (**Figure 1, Supplementary Table 2**). The logistic regression, penalized logistic regression using LASSO, and random forest models had similar point estimates and overlapping 95% CIs for each validation set, though performance differed among hospitals. All models were well-calibrated among patients with <50% mortality risk, representing most ICU patients (**Figure 2, Supplementary Table 2, Supplementary Figure**).

We performed secondary analyses using logistic regression for computational efficiency, given the findings of similar performance of the main models across modeling approaches. Models performed similarly with SBS \geq 0.13 and c-statistics $>$ 0.75 for most or all days through ICU day 15, with subsequent declining performance (**Figure 3, Supplementary Table 3**). Calibration differed between male and female patients in some validation sets, and between Black and White patients in most validation sets, but SBSs were not statistically different for most hospitals (**Figure 4, Supplementary Tables 4 and 5**).

DISCUSSION

This study adds to the literature on ICU mortality prediction by adapting LAPS2 from a single hospital admission score incorporating data from the 72 hours preceding admission in an integrated health system to an admission score incorporating only data from the first 24 hours of admission plus a daily updating score. Additionally, we are not aware of prior attempts to evaluate the performance of models incorporating these LAPS2 modifications to predict in-hospital mortality among an exclusively ICU population. Our patient-level model demonstrated lower discrimination than Escobar et al.'s model.⁷ Our patient-day-level models including daily

LAPS2 score improved upon this discrimination to a maximum c-statistic of 0.878. Thus, our results validate the use of modified LAPS2 scores, broadening the generalizability. Furthermore, having high-performing, time-updating mortality prediction models are critical to clinical outcome prediction, risk adjustment, quality and safety initiatives, and hospital performance.

Performance of the original LAPS2 is excellent and has been externally validated,¹²⁻¹⁴ however, it has several important limitations. Many health systems do not have ready access to physiologic data preceding hospital admission. Furthermore, a single score on hospital admission cannot account for physiologic changes that occur after admission and may add predictive value. Expanding the scope of LAPS2 to calculate time-updating scores to estimate mortality risk may mitigate these limitations. Indeed, one study modifying the prior version of LAPS2 by adding daily updating covariates yielded a model with excellent performance to estimate the hazard of death on each hospital day among ward patients.¹⁵ Another study of time-updating severity of illness scores including SOFA to predict ICU mortality similarly demonstrated high discrimination.¹⁶ Our results add to the evidence by demonstrating the value of daily LAPS2 scores to predict mortality on a time-updating basis among ICU patients.

While the overall performance of our time-updating LAPS2 models was strong, there was performance variability across hospitals, and calibration differed by gender and race. In particular, performance was poor among Black patients in hospitals D and E with wide CIs, likely due to data sparsity. Such variability could be problematic for clinical outcome prediction or limited resource allocation at the institutional level. Similar to Ashana et al.'s findings, overprediction of mortality among Black patients could erroneously result in deprioritization of Black patients to receive ICU beds or mechanical ventilation during mass casualty events, for example.⁸ However, other use cases such as risk adjustment or outcome assessment for ICU staffing and organizational studies may not be as adversely influenced by the race-based performance differences we observed. Future work is needed to further evaluate for algorithmic equity in similar models across health systems and to improve model equity performance.¹⁷⁻²⁴

Additionally, while prior studies vary in “chronic critical illness” definitions from 72 hours of mechanical ventilation to 21 days of critical illness,²⁵ our time-updating LAPS2 model performance highlights a decrement around ICU day 15.¹⁰ Although external validation is needed, we hypothesize that this may define an important transition point from acute to chronic critical illness, where persistent abnormal physiology may have lower predictive value. Having a physiologic definition of chronic critical illness could transform both clinical care and critical illness survivorship research.

Our study has several important limitations. First, we performed our study within a single health system, potentially limiting generalizability. However, our health system includes five diverse hospitals, and differs from the integrated health system where LAPS2 was originally developed. Second, we only included covariates in our prediction models that were utilized in Escobar et al.'s original manuscript for comparative purposes, but additional critical illness measures could enhance prediction. Third, LAPS2 is computationally more complex than other

frequently used mortality prediction models such as SOFA. Fourth, while time-updating LAPS2 scores improved individual mortality prediction, time-updating scores may not be appropriate for other applications such as overall ICU performance. Finally, our study was not powered to assess for algorithmic equity. Larger studies are needed prior to implementing time-updating LAPS2 scores among ICU patients for clinical or research use.

CONCLUSION

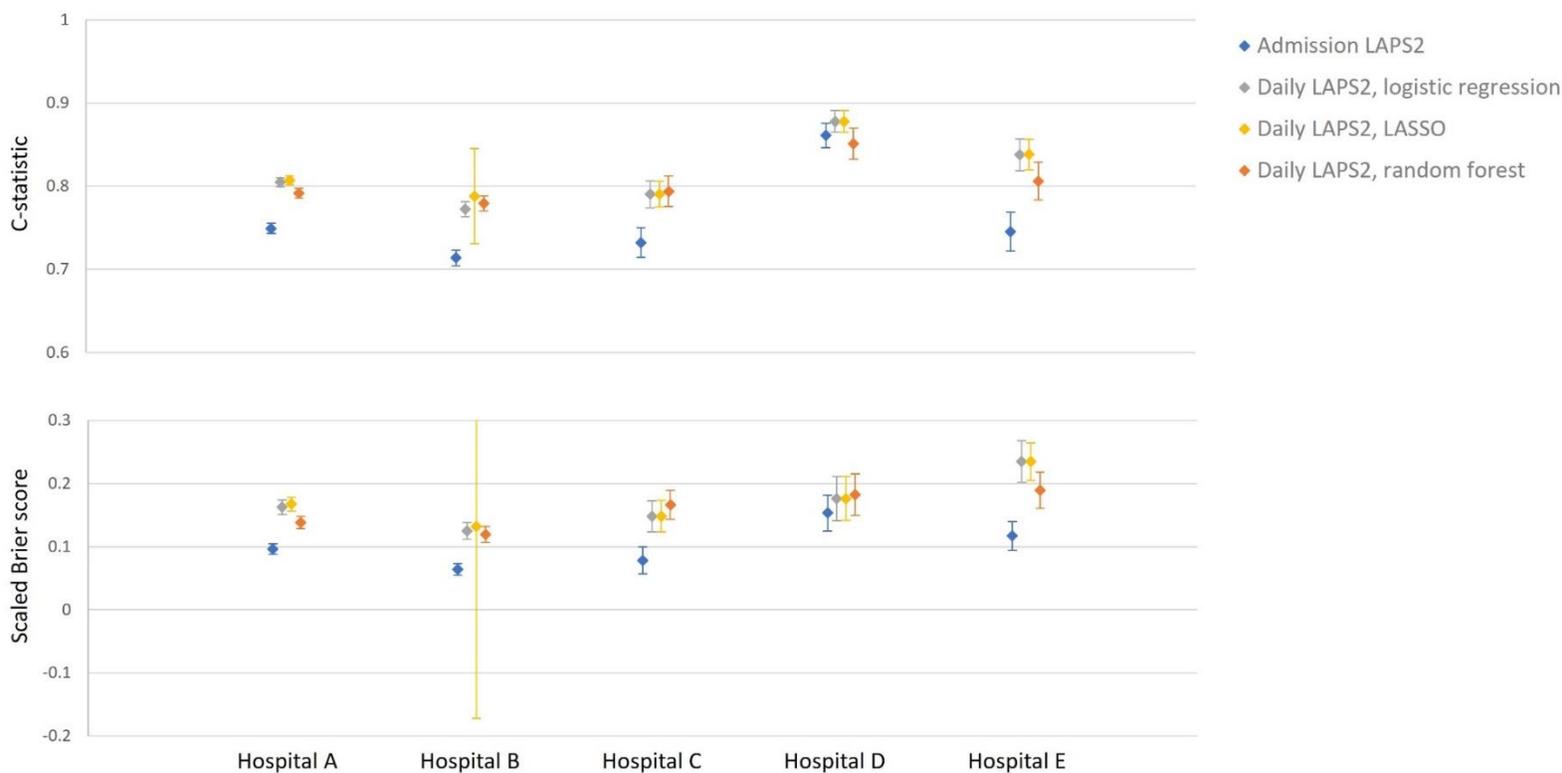
In summary, novel prediction models incorporating daily, time-updating, modified LAPS2 scores to predict mortality among an ICU population perform better than the original LAPS2. Individual daily scores used to predict 30-day in-hospital mortality perform consistently well for the first 15 days of hospitalization. Future studies should evaluate these findings in large, multicenter cohorts.

TABLES AND FIGURES

Table 1: Characteristics and outcomes of included patients

Characteristic or outcome	N=13,993
Age, median (IQR)	65 (55-74)
Male, n (%)	8,511 (60.8%)
ICU admission source, n (%)	
Operating room	6,314 (45.1%)
Interhospital transfer	3,130 (22.4%)
Emergency department	2,884 (20.6%)
Ward transfer	1,608 (11.5%)
Direct admission	57 (0.4%)
Hospital LOS prior to ICU admission in days, median (IQR)	0.3 (0.1-0.9)
Surgical service type, n (%)	9,268 (66.2%)
Elixhauser comorbidity index, ³ median (IQR)	11 (4-18)
Mortality, n (%)	1,657 (11.8%)
Hospital LOS in days, median (IQR)	10.4 (6.09-19.8)
Abbreviations: IQR, interquartile range; ICU, intensive care unit; LOS, length of stay	

Figure 1: Performance of models to predict in-hospital mortality within 30 days of intensive care unit admission

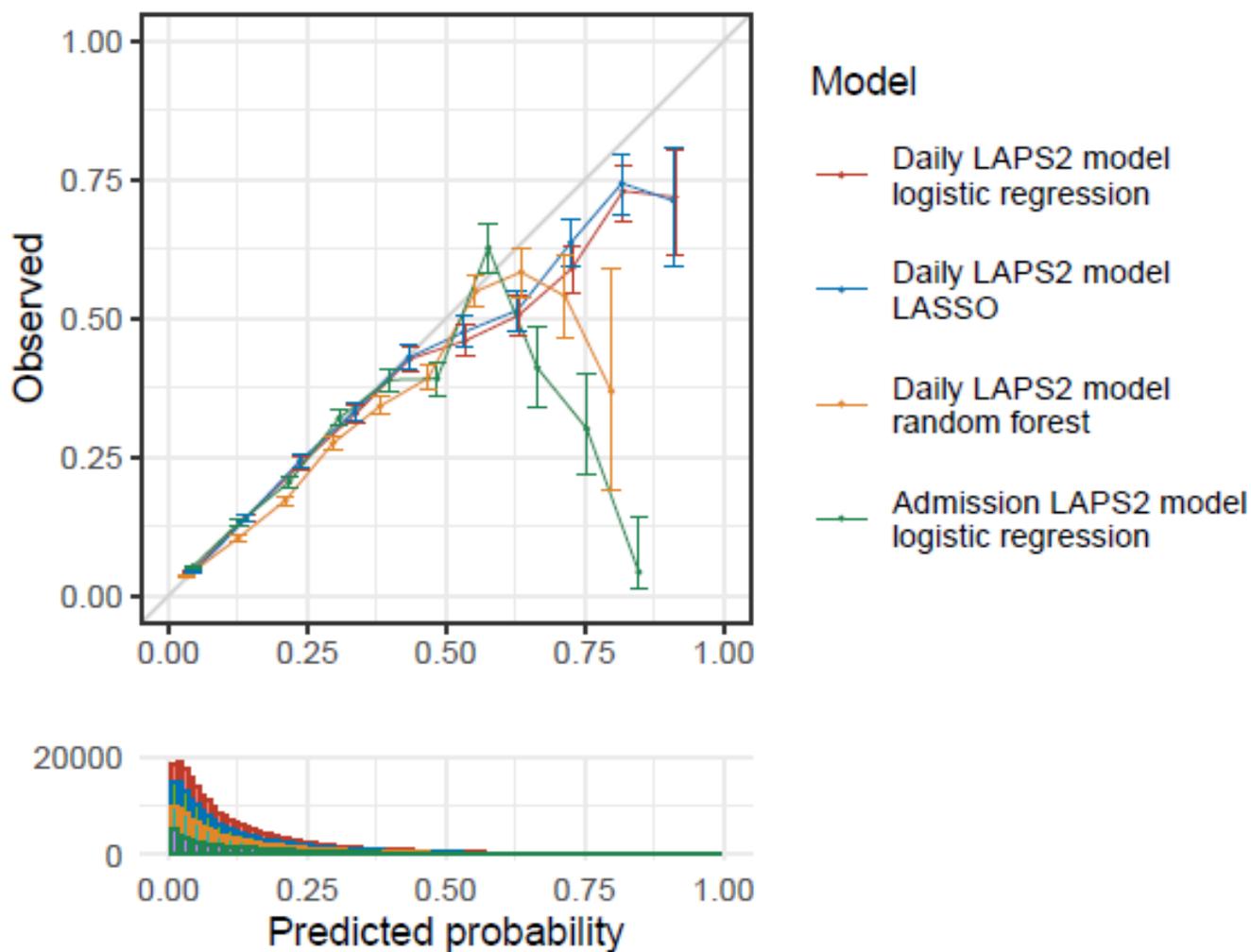


Abbreviations: LAPS2, Laboratory-based Acute Physiology Score, version 2; LASSO, least absolute shrinkage and selection operator

Closed diamonds represent point estimates for c-statistics and scaled Brier scores, as indicated.

Error bars represent 95% confidence intervals, estimated by bootstrapping 500 randomly selected samples

Figure 2: Calibration plots of models to predict in-hospital mortality within 30 days of intensive care unit admission



Abbreviations: LAPS2, Laboratory-based Acute Physiology Score, version 2; LASSO, least absolute shrinkage and selection operator

Figure 3: Performance of daily Laboratory-based Acute Physiology Score, version 2 (LAPS2) logistic regression model by day of intensive care unit (ICU) admission

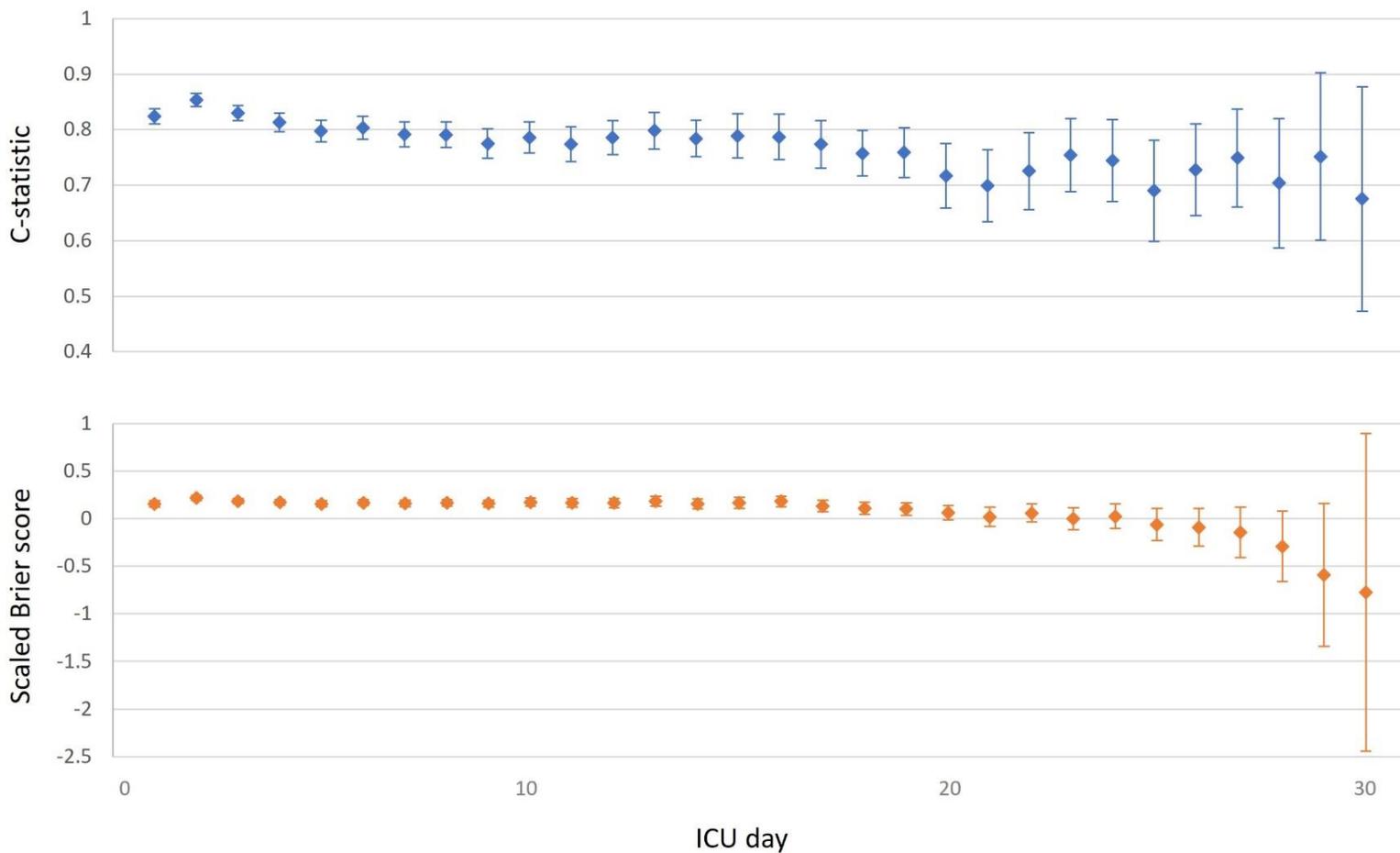
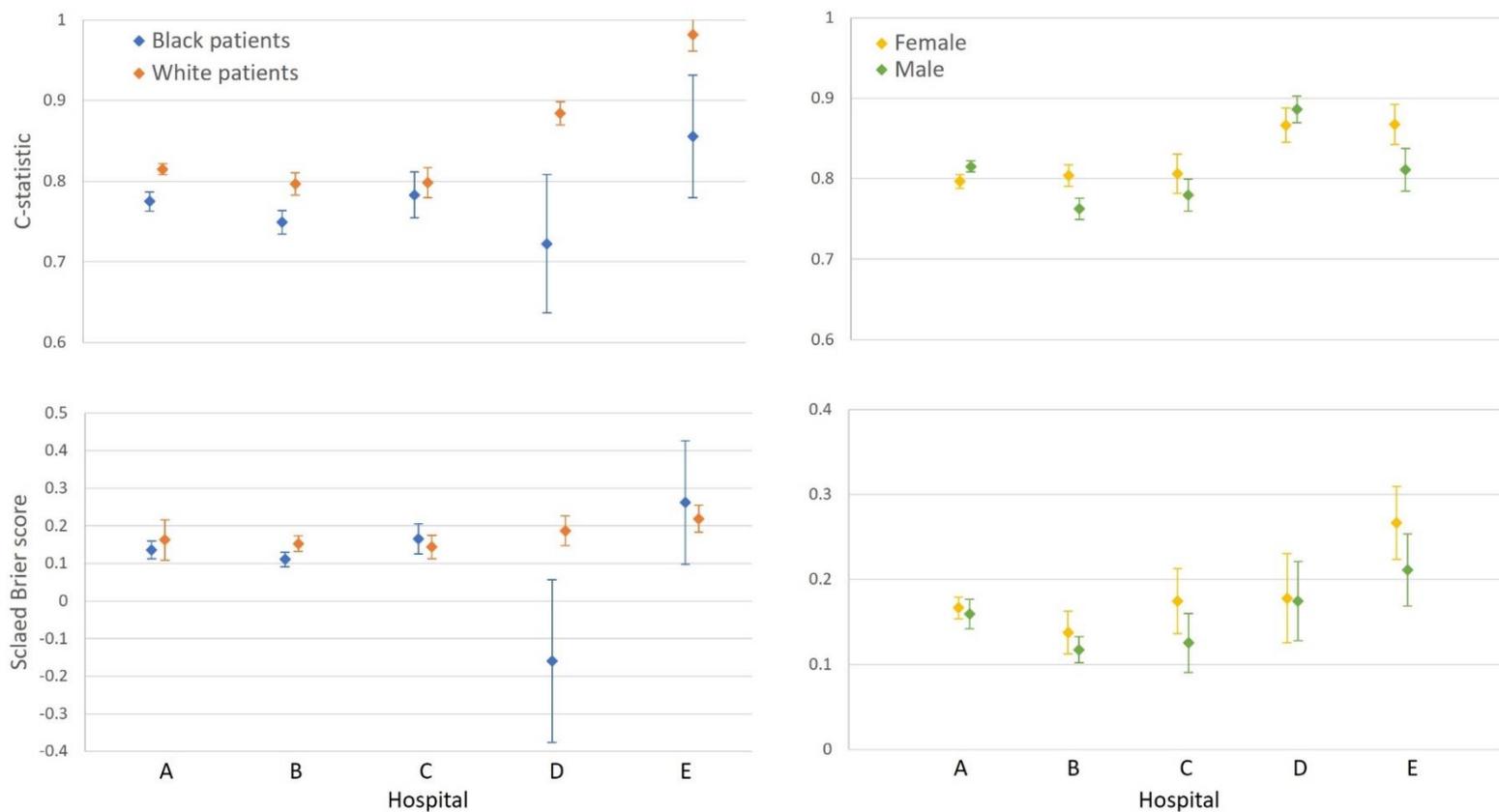


Figure 4: Performance of daily Laboratory-based Acute Physiology Score, version 2 (LAPS2) logistic regression model by self-reported race and gender



References

1. Keegan MT, Gajic O, Afessa B. Severity of illness scoring systems in the intensive care unit. *Crit Care Med*. Jan 2011;39(1):163-9. doi:10.1097/CCM.0b013e3181f96f81
2. Keegan MT, Gajic O, Afessa B. Comparison of APACHE III, APACHE IV, SAPS 3, and MPM0III and influence of resuscitation status on model performance. *Chest*. Oct 2012;142(4):851-858. doi:10.1378/chest.11-2164
3. Kim EK, Kwon YD, Hwang JH. [Comparing the performance of three severity scoring systems for ICU patients: APACHE III, SAPS II, MPM II]. *J Prev Med Public Health*. Aug 2005;38(3):276-82.
4. Kramer AA, Higgins TL, Zimmerman JE. Comparison of the Mortality Probability Admission Model III, National Quality Forum, and Acute Physiology and Chronic Health Evaluation IV hospital mortality models: implications for national benchmarking*. *Crit Care Med*. Mar 2014;42(3):544-53. doi:10.1097/CCM.0b013e3182a66a49
5. Lambden S, Laterre PF, Levy MM, Francois B. The SOFA score-development, utility and challenges of accurate assessment in clinical trials. *Crit Care*. Nov 2019;23(1):374. doi:10.1186/s13054-019-2663-7
6. Escobar GJ, Greene JD, Scheirer P, Gardner MN, Draper D, Kipnis P. Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. *Med Care*. Mar 2008;46(3):232-9. doi:10.1097/MLR.0b013e3181589bb6
7. Escobar GJ, Gardner MN, Greene JD, Draper D, Kipnis P. Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. *Med Care*. May 2013;51(5):446-53. doi:10.1097/MLR.0b013e3182881c8e
8. 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*. 07 15 2021;204(2):178-186. doi:10.1164/rccm.202012-4383OC
9. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. Jan 2010;21(1):128-38. doi:10.1097/EDE.0b013e3181c30fb2
10. Iwashyna TJ, Hodgson CL, Pilcher D, et al. Timing of onset and burden of persistent critical illness in Australia and New Zealand: a retrospective, population-based, observational study. *Lancet Respir Med*. 07 2016;4(7):566-573. doi:10.1016/S2213-2600(16)30098-4
11. Ingraham NE, Jones EK, King S, et al. Re-Aiming Equity Evaluation in Clinical Decision Support: A Scoping Review of Equity Assessments in Surgical Decision Support Systems. *Ann Surg*. Aug 09 2022;doi:10.1097/SLA.0000000000005661
12. van Walraven C, Escobar GJ, Greene JD, Forster AJ. The Kaiser Permanente inpatient risk adjustment methodology was valid in an external patient population. *J Clin Epidemiol*. Jul 2010;63(7):798-803. doi:10.1016/j.jclinepi.2009.08.020
13. Wong J, Taljaard M, Forster AJ, Escobar GJ, van Walraven C. Addition of time-dependent covariates to a survival model significantly improved predictions for daily risk of hospital death. *J Eval Clin Pract*. Apr 2013;19(2):351-7. doi:10.1111/j.1365-2753.2012.01832.x
14. Lagu T, Pekow PS, Shieh MS, et al. Validation and Comparison of Seven Mortality Prediction Models for Hospitalized Patients With Acute Decompensated Heart Failure. *Circ Heart Fail*. 08 2016;9(8)doi:10.1161/CIRCHEARTFAILURE.115.002912

15. Wong J, Taljaard M, Forster AJ, Escobar GJ, van Walraven C. Derivation and validation of a model to predict daily risk of death in hospital. *Med Care*. Aug 2011;49(8):734-43. doi:10.1097/MLR.0b013e318215d266
16. Badawi O, Liu X, Hassan E, Amelung PJ, Swami S. Evaluation of ICU Risk Models Adapted for Use as Continuous Markers of Severity of Illness Throughout the ICU Stay. *Crit Care Med*. 03 2018;46(3):361-367. doi:10.1097/CCM.0000000000002904
17. Parikh RB, Teeple S, Navathe AS. Addressing Bias in Artificial Intelligence in Health Care. *JAMA*. 12 24 2019;322(24):2377-2378. doi:10.1001/jama.2019.18058
18. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. *J Glob Health*. 12 2019;9(2):010318. doi:10.7189/jogh.09.020318
19. Gurupur V, Wan TTH. Inherent Bias in Artificial Intelligence-Based Decision Support Systems for Healthcare. *Medicina (Kaunas)*. Mar 20 2020;56(3)doi:10.3390/medicina56030141
20. DeCamp M, Lindvall C. Latent bias and the implementation of artificial intelligence in medicine. *J Am Med Inform Assoc*. 12 09 2020;27(12):2020-2023. doi:10.1093/jamia/ocaa094
21. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med*. 01 01 2019;170(1):W1-W33. doi:10.7326/M18-1377
22. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med*. 01 01 2019;170(1):51-58. doi:10.7326/M18-1376
23. Allen A, Mataraso S, Siefkas A, et al. A Racially Unbiased, Machine Learning Approach to Prediction of Mortality: Algorithm Development Study. *JMIR Public Health Surveill*. 10 22 2020;6(4):e22400. doi:10.2196/22400
24. Rojas JC, Fahrenbach J, Makhni S, et al. Framework for Integrating Equity Into Machine Learning Models: A Case Study. *Chest*. Jun 2022;161(6):1621-1627. doi:10.1016/j.chest.2022.02.001
25. Loss SH, Nunes DSL, Franzosi OS, Salazar GS, Teixeira C, Vieira SRR. Chronic critical illness: are we saving patients or creating victims? *Rev Bras Ter Intensiva*. 2017 Jan-Mar 2017;29(1):87-95. doi:10.5935/0103-507X.20170013