YALE JOURNAL OF BIOLOGY AND MEDICINE 96 (2023), pp.293-312.

ORIGINAL CONTRIBUTION

YALE JOURNAL OF BIOLOGY AND MEDICINE
YJBM
CELEBRATING 95 YEARS 1928–2023

Racial Disparities in Invasive ICU Treatments Among Septic Patients: High Resolution Electronic Health Records Analysis from MIMIC-IV

Sara Mohammed[a,b], João Matos[a,c], Matthieu Doutreligne[d,e], Leo Anthony Celi[a,f,g], and Tristan Struja[a,h,*]

[a]*Laboratory for Computational Physiology, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA;* [b]*Harvard Medical School, Boston, MA, USA;* [c]*Faculty of Engineering of University of Porto, Porto, Portugal;* [d]*Mission Data, Haute Autorité de Santé, Saint-Denis, France;* [e]*Soda Research Team, Institut National de Recherche en Informatique et en Automatique, Palaiseau, France;* [f]*Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA;* [g]*Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA;* [h]*Medical University Clinic, Kantonsspital Aarau, Aarau, Switzerland*

Background: Low-resolution administrative databases can give biased results, whereas high-resolution, time-stamped variables from clinical databases like MIMIC-IV might provide nuanced insights. We evaluated racial-ethnic disparities in life-sustaining ICU-treatments (Invasive Mechanical Ventilation (IMV), Renal Replacement Therapy (RRT), and Vasopressors (VP)) among patients with sepsis. **Methods**: In this observational retrospective cohort study, patients fulfilling sepsis-3 criteria were categorized by treatment assignment within the first 4 days. The outcomes were treatment allocations. The likelihood of receiving treatment was calculated by race-ethnicity (Racial-ethnic group (REG) or White group (WG)) using 5-fold sub-sampling nested logistic regression and XGBoost. **Results**: In 23,914 admissions, 82% were White, 42% were women. REG were less likely to receive IMV across all eligibility days (day 1 odds ratio (OR) 0.87, 95% confidence interval (CI) 0.83-0.94, day 4 OR 0.80, 95% CI 0.72 - 0.87). There were no differences in RRT (day 1 OR 1.00, 95% CI 0.96-1.09, day 4 OR 1.00, 95% CI 0.94-1.06). REG were also less likely to be treated with VP at days 1 to 3 (day 1 OR 0.87, 95% CI 0.76-0.94), but not at day 4 (OR 0.95, 95% CI 0.87-1.01). These findings remained robust when relaxing eligibility criteria for treatment allocation. **Conclusion**: Our findings reveal significant disparities in the use of invasive life-saving ICU treatments among septic patients from racial and ethnic minority backgrounds, particularly with respect to IMV and VP use. These disparities underscore not only the need to address inequality in critical care settings, but also highlight the importance of high-resolution data.

*To whom all correspondence should be addressed: Tristan Struja, MD, MSc, MPH, Laboratory for Computational Physiology, Massachusetts Institute of Technology, Cambridge, MA; Email: tristan.struja@gmail.com; ORCID: 0000-0003-0199-0184.

Abbreviations: ICU, Intensive care unit; WG, White group; ESRD, End-stage renal disease; CKD, Chronic kidney disease; SOFA, Sequential Organ Failure Assessment; RRT, Renal replacement therapy; VP, Vasopressors; IMV, Invasive mechanical ventilation; REG, Racial-ethnic group; LOS, Length of stay; COPD, Chronic obstructive pulmonary disease; OR, Odds ratios; CI, Confidence intervals; NIS, Nationwide Inpatient Sample; eICU, Collaborative Research Database; SDH, Social Determinants of Health.

Keywords: Sepsis, XGBoost, MIMIC-IV, Treatment allocation, Invasive Mechanical Ventilation, Renal Replacement Therapy, Vasopressors, Health Services, Critical Care, Health discrepancies

Author Contributions: Conception and design: JM, TS, SM; Statistics: JM, TS, MD; Interpretation: all authors; Drafting the manuscript: SM.

INTRODUCTION

Sepsis, a severe life-threatening systemic infection, impacts an estimated 1.7 million adults annually in the United States, contributing to approximately 270,000 deaths and accounting for significant healthcare expenditures estimated at $24 billion annually [1-4]. Despite advancements in sepsis management, disparities in access to critical care, quality of treatment, patient outcomes persist, with racial and ethnic minorities disproportionately affected [4-12].

In December 2022, the Center for Medicare and Medicaid Services (CMS) proposed a bold initiative to tie reimbursements to health equity [13]. Numerous studies have identified racial disparities in sepsis care such as differential time to admission, differences in treatment, and post-discharge care plans in the intensive care unit (ICU) setting [9,14]. However, most of these studies come from claims-based databases that are prone to variable misclassification or do not contain important timestamped information [15-17].

Our group has recently shown that observational studies using low-resolution claims data can introduce substantial bias in effect size estimation [18].

Our study aims to investigate racial-ethnic differences in the provision of life-sustaining treatments among ICU patients with sepsis using the Medical Information Mart for Intensive Care version 4, or MIMIC-IV database, a de-identified electronic health record database from the Beth Israel Deaconess Medical Center in Boston, Massachusetts [19-21]. The high-resolution MIMIC-IV database, which includes time-stamped features, like treatments that enable adjustments for relevant time-varying confounders to investigate disparities in the administration of ICU interventions in a large academic center.

METHODS

This observational retrospective study is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [22]. The health equity language, narrative, and concepts of this paper follows the American Medical Association's recommendations [23]. Data were extracted from the publicly available MIMIC-IV database using SQL via Google's BigQuery using due diligence as suggested by the literature [24,25]. The MIMIC-IV database is maintained by the Laboratory for Computational Physiology at the Massachusetts Institute of Technology and shared via the PhysioNet platform [19]. The dataset has been de-identified, and the institutional review boards of the Massachusetts Institute of Technology (No. 0403000206) and Beth Israel Deaconess Medical Center (2001-P-001699/14) both approved the use of the database for research. The MIMIC-IV database includes physiologic data collected from bedside monitors, laboratory test results, medications, medical images, and clinical progress notes captured in the electronic health record from patients admitted to the ICU between 2008-2019. Approximately 60,000 de-identified medical records are archived in the MIMIC-IV database.

We hypothesized that treatment allocation of ICU interventions, specifically invasive mechanical ventilation (IMV), renal replacement therapy (RRT), and vasopressor use (VP), is not equally distributed across race-ethnicity.

Exposure and Outcome

The race-ethnicity as the exposure variable was self-reported on admission or assigned by hospital personnel. Since race-ethnicity is heavily imbalanced in MIMIC-IV (approximately 70% White patients), patients identified as White (eg, White, White – Brazilian, White – Portuguese, etc.) were grouped as a White group (WG). Asian, Black, and Hispanic patients were grouped as Racial-ethnic group (REG).

The primary outcome was the likelihood of treatment initiation for IMV, RRT, and VP.

Cohort Selection

Patients older than 18 years of age who had sepsis as defined by the sepsis-3 criteria were included in the analyses. We excluded patients with a "do not resuscitate and do not intubate" code status upon admission, patients with missing race or other race information, and advanced chronic kidney disease (CKD) > 3. This created the primary cohort of the study as defined at baseline (Table 1 and Figure 1).

To mitigate immortal time bias [26,27], four treatment eligibility windows were considered: 0 – 1 day (cohort 1), 0 – 2 days (cohort 2), 0 – 3 days (cohort 3), and 0 – 4 days (cohort 4). Patients were excluded in the respective cohorts if they had a length of stay (LOS) in the ICU of less than the treatment eligibility window plus 1 day (eg, for a treatment eligibility window of 0 – 3 days, a minimum LOS of 4 days was required) to ensure a minimum follow-up time. Subsequent admissions to the ICU within the database were excluded from the analysis to avoid repeated measures.

We then applied strict treatment eligibility criteria creating three sub-cohorts each, eg, IMV 1, RRT 1, VP 1 for cohort 1 as follows:

For IMV, the criteria were adapted from the American Thoracic Society [28], and the American College of Chest Physicians [29]

- p_aO_2/FiO_2 ratio ≤ 300 (average of the day)
- $p_aCO_2 \geq 60$ mmHg (minimum of the day)

Table 1. Baseline Information on the Final Study Cohort on Day 1, Before Applying Length of Stay Criteria

	Asian	Black	Hispanic	White
N (%) from 23,914	791 (3.3)	2520 (10.5)	1038 (4.3)	19565 (81.8)
Demographics				
Age, years, median [IQR]	67.00 [54.00,78.00]	65.50 [54.00,76.00]	58.00 [46.00,70.00]	68.00 [57.00,78.00]
Sex female, n (%)	311 (39.3)	1373 (54.5)	414 (39.9)	8014 (41.0)
English proficient, n (%)	294 (37.2)	2228 (88.4)	415 (40.0)	18651 (95.3)
Health Insurance, n (%)				
Medicaid	148 (18.7)	295 (11.7)	221 (21.3)	1117 (5.7)
Medicare	235 (29.7)	1071 (42.5)	320 (30.8)	9814 (50.2)
Other	408 (51.6)	1154 (45.8)	497 (47.9)	8634 (44.1)
Year of Admission, n (%)				
2008 - 2010	245 (31.0)	1242 (49.3)	402 (38.7)	7738 (39.6)
2011 - 2013	222 (28.1)	571 (22.7)	306 (29.5)	4854 (24.8)
2014 - 2016	195 (24.7)	436 (17.3)	209 (20.1)	4282 (21.9)
2017 - 2019	129 (16.3)	271 (10.8)	121 (11.7)	2691 (13.8)
Elective Admission, n (%)	102 (12.9)	167 (6.6)	118 (11.4)	3108 (15.9)
Major Surgery, n (%)	289 (36.5)	639 (25.4)	357 (34.4)	8033 (41.1)
Chronic co-morbidities				
Hypertension, n (%)	452 (57.1)	1823 (72.3)	581 (56.0)	12416 (63.5)
Congestive Heart failure, n (%)	225 (28.4)	975 (975)	281 (27.1)	6518 (33.3)
COPD, n (%)	131 (16.6)	747 (29.6)	242 (23.3)	4796 (24.5)
Asthma, n (%)	14 (1.8)	33 (1.3)	7 (0.7)	259 (1.3)
Coronary Artery Disease, n (%)	217 (27.4)	720 (28.6)	287 (27.6)	7393 (37.8)
CKD Stage, n (%)				
No CKD	741 (93.7)	2242 (89.0)	963 (92.8)	18335 (93.7)
1	2 (0.3)	2 (0.1)	2 (0.2)	8 (0.0)
2	3 (0.4)	38 (1.5)	5 (0.5)	130 (0.7)
3	45 (5.7)	238 (9.4)	68 (6.6)	1092 (5.6)
Diabetes mellitus type, n (%)				
No Diabetes	537 (67.9)	1411 (56.0)	615 (59.2)	13800 (70.5)

Type 1	12 (1.5)	88 (3.5)	30 (2.9)	525 (2.7)
Type 2	242 (30.6)	1021 (40.5)	393 (37.9)	5240 (26.8)
Connective Tissue Disease, n (%)	27 (3.4)	142 (5.6)	37 (3.6)	820 (4.2)
Charlson Comorbidity Index, median [IQR]	6.0 [4.0,8.0]	6.0 [4.0,8.0]	5.0 [3.0,7.0]	6.0 [4.0,8.0]
Admission SOFA median [IQR]	5.0 [3.0,8.0]	5.0 [3.0,8.0]	5.0 [3.0,7.0]	5.00 [3.0,7.0]
Invasive ICU treatments				
Received IMV, n (%)	372 (47.0)	1155 (45.8)	485 (46.7)	9386 (48.0)
Received RRT, n (%)	35 (4.4)	133 (5.3)	58 (5.6)	1006 (5.1)
Received VP, n (%)	397 (50.2)	996 (39.5)	459 (44.2)	9681 (49.5)
Outcomes				
Mortality, n (%)	142 (18.0)	413 (16.4)	138 (13.3)	3038 (15.5)
ICU LOS, days, median [IQR]	2.71 [1.48,5.56]	2.92 [1.58,5.96]	2.79 [1.50,5.41]	2.79 [1.50,5.75]

Legend: Racial-ethnic patient group includes Asian, Black, and Hispanic. Initiation offset denotes the time between ICU admission and start of the treatment. Abbreviations: IMV, invasive mechanical ventilation; RRT, renal replacement therapy; ICU, intensive care unit; LOS, length of stay; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; IQR, interquartile range; SOFA, sequential organ failure assessment.

- Respiratory rate ≥ 20 bpm (average of the day)

For RRT, the criteria were adapted from the Kidney Disease Improving Global Outcomes (KDIGO) [30,31]

- Urine output / Weight ≤ 12 L/kg (total of the day)
- Potassium ≥ 6.5 mEq/L (minimum of the day)
- pH ≤ 7.2 and bicarbonate ≤ 12 mEq/L (both maximum of the day)

And for VP, the criteria were adapted from the Surviving Sepsis Campaign [31]

- Mean blood pressure ≤ 65 mmHg (average of the day)
- Lactate ≥ 2 mmol/L (minimum of the day)

This was done to reflect decision making by the clinicians. If treatment was started after any eligibility period, the patient was assigned to the control group.

Covariates

A total of 50 patient-level variables were extracted, including:

- Baseline variables: demographics including English proficiency and insurance type to account for Social Determinants of Health (SDH), comorbidities, and admission information such as infection source, year, and type of ICU.
- Time-varying variables: Sequential Organ Failure Assessment (SOFA) score [32] and its components, vital signs, and laboratory values.

To best represent the patient's clinical status, time-varying variables were selected either from the day of treatment initiation if the patient received the treatment, or the day with the worst SOFA score if the patient did not receive the treatment. For variables with repeated measures, the maximum, minimum, or average daily value was used, as clinically appropriate (Appendix A: Supplementary Table 1). ICD-10 codes for comorbidities were also extracted, including hypertension, COPD, asthma, congestive heart failure, diabetes, stages of chronic kidney disease (CKD), coronary artery disease, and connective tissue disease. Additionally, the Charlson Comorbidity Index was calculated [33].

Statistical Analysis

Statistical analysis was performed using Python 3.10.9. For summary statistics, we used the *tableone* Python package [34]. For modelling, penalized linear models – multivariate ridge-penalized logistic regression (LogReg) [35] – and non-linear models – XGBoost classifier [36] – were used, adjusted for confounders to estimate the likelihood to receive each of the three interventions stratified by race-ethnicity. We report our findings as ORs with 95% confidence intervals (OR 95% CI). White patients were considered as the reference group.

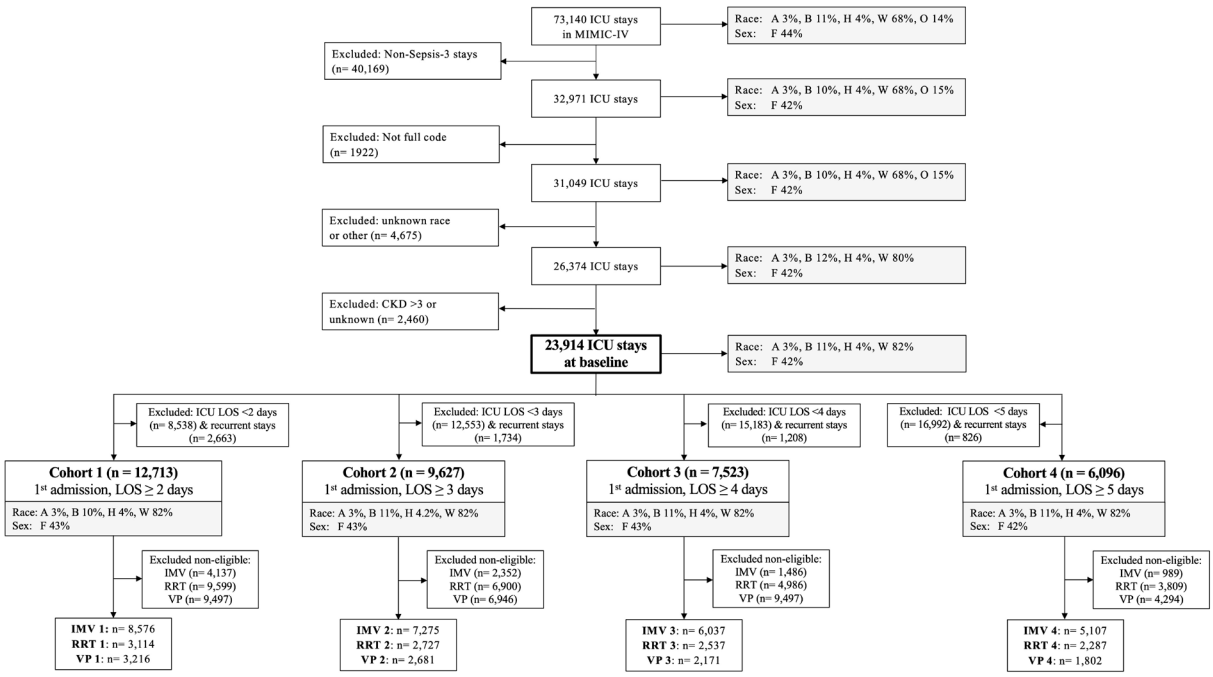


Figure 1. **Study cohort selection flow chart**. Abbreviations: ICU, intensive care unit; A, Asian; B, Black; H, Hispanic; W, White; O, other; F, female; CKD, chronic kidney disease; LOS, length of stay; d, day; IMV, mechanical ventilation; RRT, renal replacement therapy; VP, vasopressor.

Different groupings were considered for the racial-ethnic groups: A) with all REG collapsed into one category, and B) with REG patients disaggregated. The choice of ridge logistic regression was influenced by its ability to handle multicollinearity, thereby stabilizing the model.

ORs from LogReg were computed with the exponential of the race-ethnicity variable's coefficient, in line with previous literature [8,37].

XGBoost, on the other hand, offers an advantage in handling nonlinear relationships and was chosen for its ability to compute feature importance through SHAP values [36], offering insights into the relative influence of variables. After obtaining the SHAP values relative to race-ethnicity, we applied equation 1 to obtain a measure that is comparable to OR [38].

$$OR_{XGBoost} = \frac{\exp\left(\frac{\sum SHAP|_{race=REG}}{n|_{race=REG}}\right)}{\exp\left(\frac{\sum SHAP|_{race=WG}}{n|_{race=WG}}\right)} \approx \frac{Odds\ of\ Treatment\ in\ REG}{Odds\ of\ Treatment\ in\ WG} \quad (1)$$

For both the linear and nonlinear models, 95% CI were computed with 100 iterations of estimation with a 5-fold train/estimation procedure. At each iteration, for each fold, 20% of the data is used to train the model and compute the before averaging on the 5 folds. The final OR is the median of all iterations. This procedure tries to mimic the computation of OR with LogReg.

Our choice of statistical analysis techniques was guided by both the complexity of our data and the imperative for interpretability of results. Therefore, we employed XGBoost and LogReg in our study, both of which were adept at meeting our research requirements. While Neural Networks have been shown to perform well in unstructured data such as images, there is evidence that their performance in structured tabular data like ours is subpar [39]. XGBoost emerged as a particularly favorable tool for this investigation. A key strength of XGBoost over other techniques like Random Forests lies in its inherent capacity for model tuning. This attribute of XGBoost results in improved model performance, yielding more robust and reliable findings, while eliminating the need for complex and time-consuming hyperparameter tuning [40].

RESULTS

Baseline Study Cohort

The MIMIC-IV database comprised of 73,140 ICU stays, of which 32,971 met the Sepsis-3 criteria. Following the inclusion and exclusion criteria, the final cohort consists of 23,914 ICU admissions (Figure 1). The race-ethnicity distribution is as follows: 81.8% White, 10.5% Black, 4.3% Hispanic, and 3.3% Asian. The groups differed in age, with White patients being the oldest (68

Table 2. Unadjusted Treatment Allocation for Cohort 1 for Each Treatment, Stratified by Race-Ethnicity Category

	IMV 1		RRT 1		VP 1	
	REG	**WG**	**REG**	**WG**	**REG**	**WG**
N	1503	7073	521	2593	637	2579
Age, median (IQR)	64 [51,76]	68 [58,79]	69 [57,78]	70 [60,80]	65 [51,77]	69 [58,79]
Sex female, n (%)	716 (47.6)	2932 (41.5)	244 (46.8)	1101 (42.5)	291 (45.7)	1097 (42.5)
Health Insurance, n (%)						
Medicare	545 (36.3)	3586 (50.7)	214 (41.1)	1358 (52.4)	238 (37.4)	1330 (51.6)
Medicaid	205 (13.6)	399 (5.6)	43 (8.3)	139 (5.4)	92 (14.4)	130 (5.0)
Other	753 (50.1)	3088 (43.7)	264 (50.7)	1096 (42.3)	307 (48.2)	1119 (43.4)
English Proficiency limited, n (%)	480 (31.9)	331 (4.7)	133 (25.5)	114 (4.4)	206 (32.3)	118 (4.6)
Elective Admission, n (%)	114 (7.6)	986 (13.9)	51 (9.8)	419 (16.2)	33 (5.2)	271 (10.5)
Major Surgery, n (%)	432 (28.7)	2806 (39.7)	160 (30.7)	1050 (40.5)	162 (25.4)	924 (35.8)
In-Hospital Mortality, n (%)	313 (20.8)	1420 (20.1)	125 (24.0)	585 (22.6)	181 (28.4)	689 (26.7)
COPD, n (%)	405 (26.9)	1930 (27.3)	138 (26.5)	644 (24.8)	139 (21.8)	590 (22.9)
Asthma, n (%)	22 (1.5)	137 (1.9)	2 (0.4)	32 (1.2)	4 (0.6)	38 (1.5)
Coronary Artery Disease, n (%)	445 (29.6)	2814 (39.8)	174 (33.4)	972 (37.5)	184 (28.9)	973 (37.7)
CKD Stage, n (%)						
0-2	1379 (91.7)	6680 (94.4)	476 (91.4)	2432 (93.8)	593 (93.1)	2426 (94.1)
3	124 (8.3)	393 (5.6)	45 (8.6)	161 (6.2)	44 (6.9)	153 (5.9)
Pneumonia, n (%)	103 (6.9)	384 (5.4)	28 (5.4)	122 (4.7)	40 (6.3)	130 (5.0)
Urinary Tract Infection, n (%)	4 (0.3)	33 (0.5)	5 (1.0)	29 (1.1)	3 (0.5)	13 (0.5)
Biliary Tract Infection, n (%)	1 (0.1)	14 (0.2)	0 (0.0)	9 (0.3)	0 (0.0)	7 (0.3)
Skin Infection, n (%)	2 (0.1)	9 (0.1)	0 (0.0)	1 (0.0)	1 (0.2)	3 (0.1)
ICU LOS (days, if deceased), median (IQR)	5.83 [3.5,10.2]	5.90 [3.6,10.5]	5.50 [3.1,10.1]	5.79 [3.4,10.7]	5.00 [3.4,9.6]	5.50 [3.3,10.2]
ICU LOS (days, if survived), median (IQR)	4.54 [3.0,8.8]	4.42 [2.9,8.2]	4.83 [3.0,8.5]	4.38 [2.9,8.00]	4.58 [3.0,8.9]	4.21 [2.9,8.2]
Charlson Comorbidity Index, median (IQR)	6 [4,8]	6 [4,8]	6 [5,9]	6 [4,8]	6 [4,9]	6 [4,8]
SOFA Score (admission), median (IQR)	6 [4,9]	6 [4,9]	7 [4,10]	6 [4,9]	7 [5,10]	7 [4,10]
IMV initiated until the cohort day, n (%)	850 (56.6)	4061 (57.4)	284 (54.5)	1420 (54.8)	346 (54.3)	1317 (51.1)
RRT initiated until the cohort day, n (%)	55 (3.7)	214 (3.0)	31 (6.0)	143 (5.5)	31 (4.9)	142 (5.5)
Vasopressor initiated until the cohort day, n (%)	684 (45.5)	3665 (51.8)	254 (48.8)	1384 (53.4)	319 (50.1)	1519 (58.9)

IMV initiation offset in hours, median (IQR)	2.0 [1.0,7.0]	3.0 [1.0,8.0]	3.0 [1.0,8.0]	3.0 [1.0,7.0]	2.0 [1.0,8.0]	3.0 [1.0,9.0]	
RRT initiation offset in hours, median (IQR)	41.0 [10.0,91.0]	45.0 [12.0,100.2]	34.0 [12.0,74.0]	36.0 [11.0,76.2]	31.0 [12.5,65.5]	34.0 [11.0,76.0]	
VP initiation offset in hours, median (IQR)	3.0 [1.0,11.0]	3.0 [1.0,10.0]	3.0 [1.0,11.0]	3.0 [1.0,8.0]	3.0 [1.0,11.5]	3.0 [1.0,9.0]	

Legend: Racial-ethnic patient groups includes Asian, Black, and Hispanic. Abbreviations: IMV, invasive mechanical ventilation; RRT, renal replacement therapy; VP, vasopressors; REG, Racial-ethnic patient groups; WG, White group; IQR; interquartile range; COPD, chronic obstructive pulmonary disease; OASIS, Oxford Acute Severity of Illness Score; SOFA, sequential organ failure assessment.

years [IQR 57,78]) and Hispanic patients the youngest (58 years [IQR 46,70]). Sex distribution also differed, with the Black cohort having the highest proportion of females (54.5%) compared to 41% in White patients and 40% in Asian and Hispanic patients.

Insurance coverage also differed between race-ethnicities with White patients having the highest Medicare coverage (50.2%), and Hispanic patients having the highest Medicaid coverage (21.3%).

Variations in chronic medical conditions such as hypertension, congestive heart failure, COPD, and coronary artery disease were observed. For example, hypertension was predominantly prevalent among Black patients (72.3%), whereas White patients showed a higher prevalence of coronary artery disease (37.8%). The occurrence of type 2 diabetes was notably higher in Black (40.5%) and Hispanic patients (37.9%) in contrast to Asian (30.6%) and White patients (26.8%).

Upon admission, the SOFA scores across all groups were comparable, reflecting similar illness severity. Mortality varied across race-ethnicities, with Asian patients having the highest death rate (18.0%). However, the median ICU LOS showed no difference across the groups.

Life-sustaining Treatment Eligible Cohorts

After the application of treatment-specific eligibility criteria, we computed baseline characteristics for each invasive treatment for cohort 1 (ie, IMV1, RRT1, and VP1) (Table 2). The median age of the REG was lower for all treatment sub-cohorts except for RRT. REG had similar Charlson co-morbidity Index and SOFA scores compared to WG.

Main Analysis

For IMV, the results showed a significant disadvantage for REG across all cohorts IMV1 to IMV4 using both XGBoost and LogReg models, except in the case of IMV1 LogReg OR 1.01, 95% CI 0.97 - 1.06) (Appendix A: Supplementary Table 2). Breaking down the results to distinct ethnicities, the Hispanic group had the lowest probability of receiving IMV with OR 0.71 (95% CI 0.61 - 0.94) in the IMV4 sub-cohort using LogReg. In contrast, the Black group had an increased likelihood of receiving IMV in the IMV1 LogReg model OR 1.16, CI 1.09 - 1.22, but the OR flipped in the IMV4 sub-cohort (XGBoost OR 0.91, CI 0.82 - 0.97; LogReg OR 0.86, CI 0.78 - 0.98) (Figures 2b and 3b). Lastly, for the Asian group, no significant difference was observed in the likelihood of receiving IMV across all sub-cohorts.

For RRT, we observed no significant difference between REG and WG across all four sub-cohorts. But again, differences emerged within specific REG groups. Specifically, the Asian group showed an increased like-

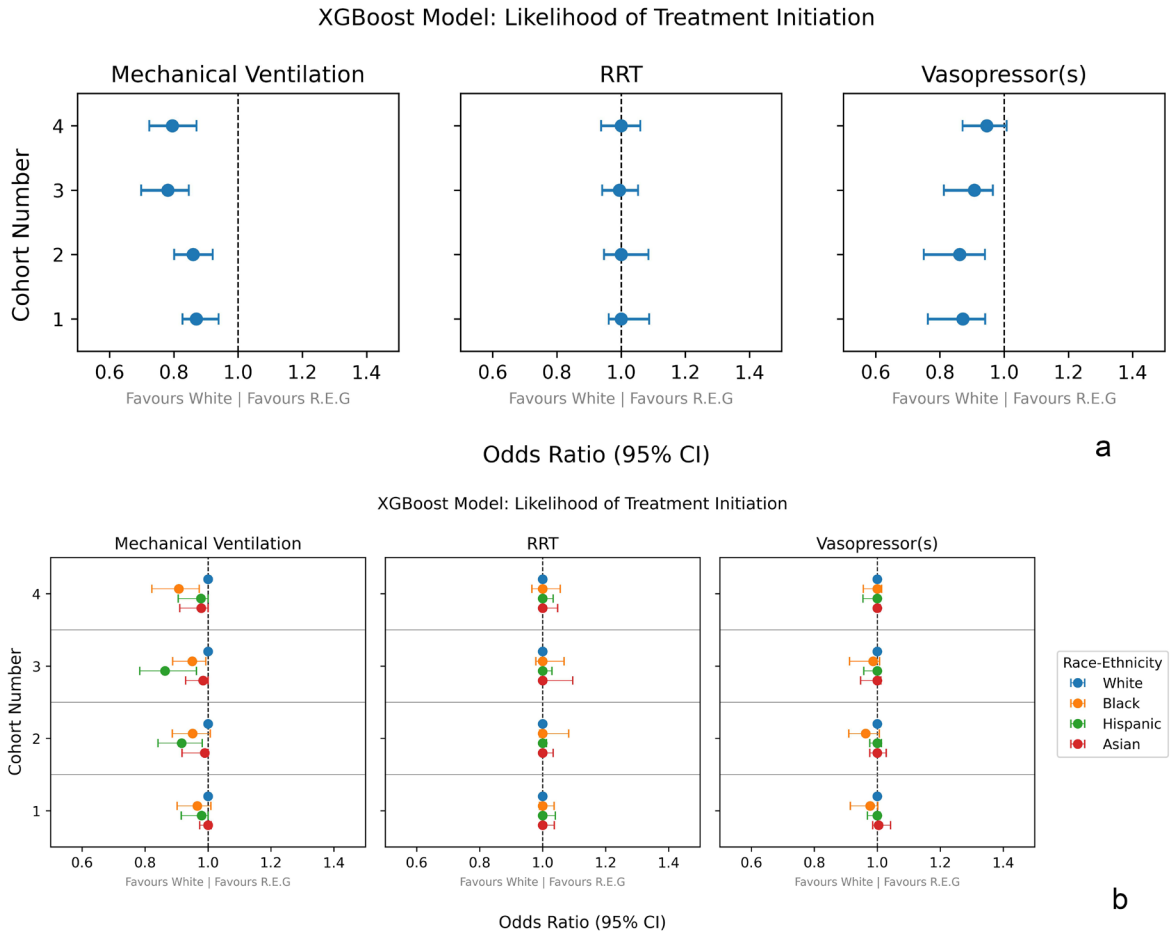


Figure 2a. **XGBoost results for likelihood of treatment initiation, White group versus Racial-ethnic group in the strict inclusion criteria cohorts**. **b**. **XGBoost results for likelihood of treatment initiation, White group versus Racial-ethnic groups in the strict inclusion criteria cohorts**. Abbreviations: RRT, renal replacement therapy; REG, racial-ethnic group.

lihood of RRT initiation in the RRT4 sub-cohort, which was significant according to the LogReg model (OR: 1.2, 95% CI 1.01 - 1.45). No significant differences in RRT initiation were observed for the Black and Hispanic groups.

For VP treatment, we found that REG was generally less likely to receive this treatment with diminishing effect size over the sub-cohorts in both the XGBoost and LogReg analyses. We did not discover relevant differences in VP treatment initiation for the Asian, Black, and Hispanic groups.

Models' Performance

Supplementary Table 3 reports models' performances for the main analysis, when fitting to 80% of the dataset and testing within the remaining 20%. AUC varies between approximately 91% and 97%; Brier scores between 4% and 9%. The nonlinear XGBoost models tend to achieve better discrimination than the linear LogReg but with slightly lower calibration. Overall, the performance metrics are high, indicating that residual clinical confounding is negligible.

Sensitivity Analysis

The strict inclusion criteria for treatment were removed in the sensitivity analyses. The results were unchanged. However, we observed a tendency for widening disparities in treatment provision between the WG and REG. Especially in IMV and VP, the REG was less likely to be treated in all sub-cohorts. There were no significant disparities between REG and WG on RRT initiation across all four sub-cohorts (Supplementary Figures 2a and 3a).

DISCUSSION

Using the high-resolution MIMIC-IV database, we demonstrated racial-ethnic disparities in the provision of

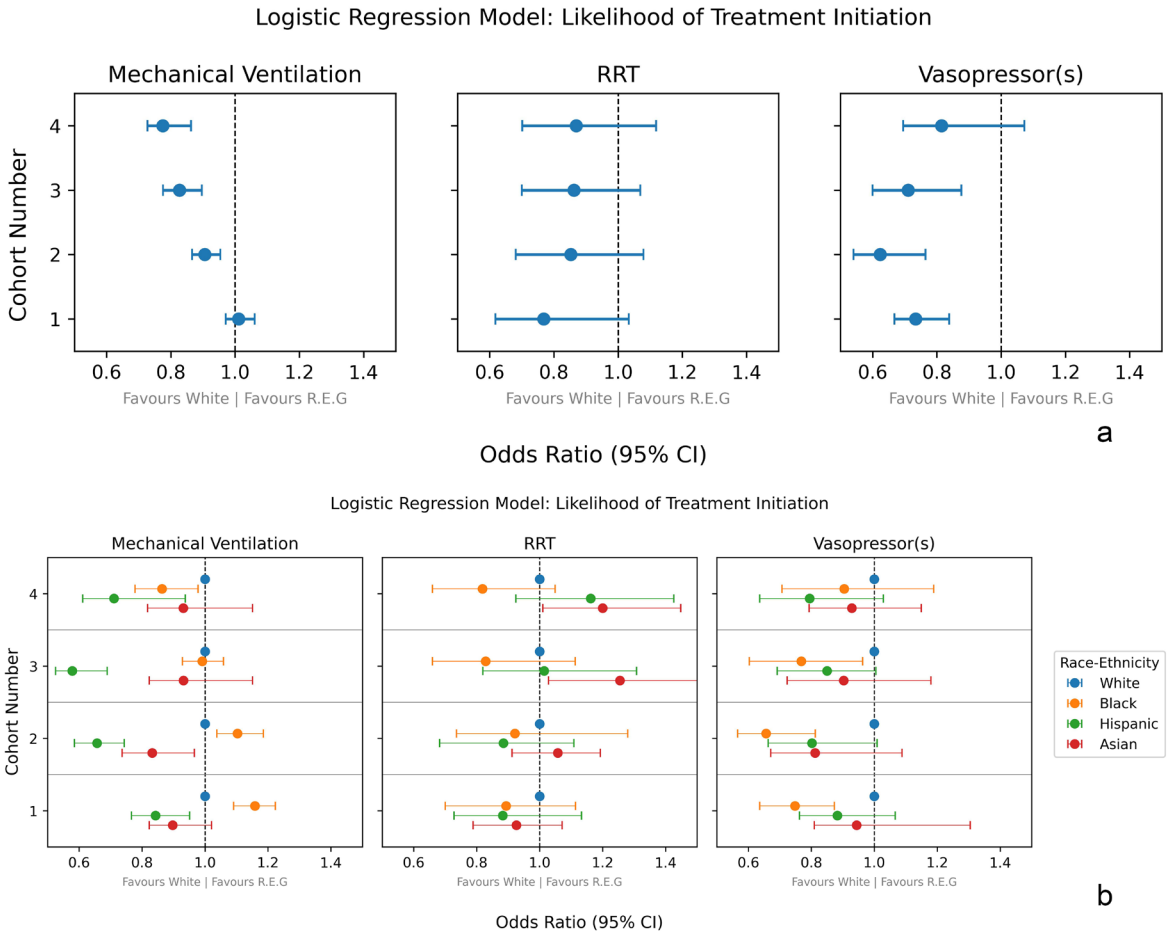


Figure 3a. **Logistic regression results for likelihood of treatment initiation, White group versus Racial-ethnic group in the strict inclusion criteria cohorts**. **b**. **Logistic regression results for likelihood of treatment initiation, White group versus Racial-ethnic groups in the strict inclusion criteria cohorts**. Abbreviations: RRT, renal replacement therapy; REG, racial-ethnic group.

life-sustaining ICU interventions among patients with sepsis. Our findings unveiled notable disparities for REG in the allocation of IMV and VP to patients admitted to the ICU for sepsis, but not RRT. These findings stand in contrast to existing literature suggesting that racial and ethnic minority groups in the US are more likely to receive these therapies.

Contrary to our work, many studies analyzed claims data or disease registries. These data sources do not provide adequate information when it comes to important clinical covariates [18], and thus have intrinsic limitations and mis-classification biases [41], which can be further amplified by model misspecifications [42].

For instance, Frei et al. analyzed data from 35,706 patients on general wards with community-acquired pneumonia from the Veterans Health Administration database [43]. They found that African American patients were equally likely to receive guideline-concordant antibiotics and have the same 30-day mortality as White patients. When looking at a subset of 5,172 patients being admitted to an ICU, they found that African Americans patients were still as likely to receive guideline-concordant antibiotics as White patients but had lower 30-day mortality (OR 0.82, 95% CI 0.68 - 0.99). But their analysis only adjusted for 18 covariates which were all collected for billing purposes and had no information on clinical covariates such as laboratory values or vital signs. It is important to note that high-resolution databases like MIMIC-IV are prone to errors in data extraction or data capture, but these errors should be random and equally distributed across patient groups in theory [44].

The importance of including relevant covariates is highlighted by the finding that the XGBoost model did not outperform penalized LogReg. Despite XGBoost being able to learn non-linear correlations, time-stamped high-resolution clinical information reflecting provider decision-making process seems to mute the theoretical advantages of machine learning models.

While we found different treatment likelihoods for IMV and VP, the absence of disparity in RRT administration warrants further investigation. It may be attributed to better established guidelines for RRT initiation [30,31], reducing the influence of personal biases. In comparison to RRT and VP, there are even fewer clear-cut recommendations on the initiation of IMV [28,29]. Given the low adherence of physicians to the surviving sepsis campaign's recommendations [45], it is not surprising that this leaves much room for subjective decisions. Interestingly, we saw an increasing tendency for bias in MV from day 1 to day 4 for the REG which is contrary to the current literature [8,46], and might be explained by local policies or our high-resolution data. In the case of treatment with VP, we saw diminishing differences between the WG and the REG over time which might show that health care professionals might be rethinking their treatment strategies despites subconscious biases. When looking at the results for the specific Racial-Ethnic groups, there are some discrepancies between XGBoost and LogReg which might be due to the handling of non-linear associations by XGBoost. In general, the overall differences seem to be driven by Black and Hispanic patients while Asian patients were more likely to be treated similarly to White patients. While this finding is hard to explain, it might be related to better SDH and emphasis on healthcare among Asian groups compared to the other Racial-Ethnic groups when looking at the higher rates of health insurance among Asians serving as a proxy [47].

The reasons for these disparities are likely complex and multifactorial, including socio-economic factors, systemic issues, and implicit biases in medical decision-making. Beyond the differences in treatment allocation demonstrated in our study, we do not fully understand how these disparities impact clinical outcomes.

Limitations

While our study brings significant insights into the discussion of racial-ethnic disparities in critical care, we recognize limitations of our analysis. Firstly, the MIMIC-IV database does not include patients with sepsis who were admitted to the regular ward. Selection bias is introduced if there is difference in triaging decisions as regards who are admitted to the ICU [48], and the disparities may be more pronounced than the estimates in our models.

Secondly, while MIMIC-IV is highly granular on clinical data, it lacks specific information on certain SDH, mostly to protect the confidentiality of the patients. Numerous studies as highlighted in a review by Sheikh et al. [49], have affirmed the significant role of SDH as confounding factors in healthcare outcomes. Our study adjusted for insurance type and English language proficiency as proxies for SDH. However, the absence of more specific data on health literacy and accessibility to healthcare facilities limited our ability to fully control for all potential SDH confounders. This constraint may impact the interpretation of observed disparities, underscoring the importance of considering these factors in future research.

Moreover, the scope of our data, derived from a single academic center in the US, limits the generalizability of the results. It is important to approach our findings with caution, acknowledging that they might not mirror experiences in other healthcare settings.

Despite these limitations, our study contributes valuable insights into the understanding of racial-ethnic disparities in ICU care. It provides avenues for subsequent research to build upon, with the ultimate goal of delivering equitable healthcare for all patients.

CONCLUSIONS

Our study demonstrates racial-ethnic disparities in the delivery of life-sustaining ICU interventions using high-resolution data, specifically IMV and VP, among patients who are admitted with sepsis. Despite clinicians' commitment to individualized care, our study elucidates that differences in treatment patterns persist, not being justified by clinical covariates alone, highlighting that cognitive biases are still subconsciously influencing treatment decisions currently. However, we did not observe such disparities in the administration of RRT, suggesting the beneficial role of stringent guidelines and standardized practices. Besides standardizing treatment practices, we need to thoroughly incorporate the detection of subconscious biases and ways to mitigate them into the curriculum of every of health care profession. Moreover, our study underscores the potential pitfalls of relying on low-resolution clinical data for answering complex research questions. These types of data might introduce substantial statistical noise, skewing our understanding and interpretation of healthcare realities. Thus, we advocate for the broader adoption of high-resolution clinical data in future healthcare research to ensure the production of accurate, reliable, and meaningful insights.

Conclusively, our goal is to usher in an era of healthcare that is fair and of high-quality, a system that serves all patients equally, irrespective of their racial or ethnic backgrounds. This study forms part of that ongoing journey towards healthcare equity.

Acknowledgements: JM, SM, and TS conceived the study. TS and JM had full access to all data and take responsibility for its integrity and the accuracy of the data analysis. Data analysis was performed by SM, TS, MD, and JM. All authors contributed to the review of results. SM, TS, and JM drafted the first version of the manuscript, which was reviewed by all authors. TS takes ultimate responsibility for the content of the manuscript,

including the data and analysis.

Funding: LAC (NIBIB, R01 EB001659); TS (Swiss National Science Foundation, P400PM_194497 / 1); JM (Fulbright / FLAD Grant, Portugal, AY 2022/2023). The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data availability: The data that support the findings of this study are available in MIMIC-IV with the identifier doi.org/10.1093/jamia/ocx084. The database is publicly available on PhysioNet.

Code availability: The code that produces the results in this manuscript can be accessed at https://github.com/joamats/mit-sepsis-tx.

REFERENCES

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):801-10. Epub 2016/02/24. doi: 10.1001/jama.2016.0287.
2. Rhee C, Dantes R, Epstein L, Murphy DJ, Seymour CW, Iwashyna TJ, et al. Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014. JAMA. 2017;318(13):1241-9. Epub 2017/09/14. doi: 10.1001/jama.2017.13836.
3. Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. N Engl J Med. 2003;348(16):1546-54. Epub 2003/04/18. doi: 10.1056/NEJMoa022139.
4. Seymour CW, Gesten F, Prescott HC, Friedrich ME, Iwashyna TJ, Phillips GS, et al. Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. N Engl J Med. 2017;376(23):2235-44. Epub 2017/05/23. doi: 10.1056/NEJMoa1703058.
5. Williams JF, Zimmerman JE, Wagner DP, Hawkins M, Knaus WA. African-American and white patients admitted to the intensive care unit: is there a difference in therapy and outcome? Crit Care Med. 1995;23(4):626-36. Epub 1995/04/01. doi: 10.1097/00003246-199504000-00009.
6. Esper AM, Moss M, Lewis CA, Nisbet R, Mannino DM, Martin GS. The role of infection and comorbidity: Factors that influence disparities in sepsis. Crit Care Med. 2006;34(10):2576-82. Epub 2006/08/18. doi: 10.1097/01.CCM.0000239114.50519.0E.
7. Mayr FB, Yende S, Linde-Zwirble WT, Peck-Palmer OM, Barnato AE, Weissfeld LA, et al. Infection rate and acute organ dysfunction risk as explanations for racial differences in severe sepsis. JAMA. 2010;303(24):2495-503. Epub 2010/06/24. doi: 10.1001/jama.2010.851.
8. Mayr FB, Yende S, D'Angelo G, Barnato AE, Kellum JA, Weissfeld L, et al. Do hospitals provide lower quality of care to black patients for pneumonia? Crit Care Med. 2010;38(3):759-65. Epub 2009/12/17. doi: 10.1097/CCM.0b013e3181c8fd58.
9. Pines JM, Russell Localio A, Hollander JE. Racial disparities in emergency department length of stay for admitted patients in the United States. Acad Emerg Med. 2009;16(5):403-10. Epub 2009/02/28. doi: 10.1111/j.1553-2712.2009.00381.x.
10. Soto GJ, Martin GS, Gong MN. Healthcare disparities in critical illness. Crit Care Med. 2013;41(12):2784-93. Epub 2013/10/15. doi: 10.1097/CCM.0b013e3182a84a43.
11. Galiatsatos P, Sun J, Welsh J, Suffredini A. Health Disparities and Sepsis: a Systematic Review and Meta-Analysis on the Influence of Race on Sepsis-Related Mortality. J Racial Ethn Health Disparities. 2019;6(5):900-8. Epub 2019/05/31. doi: 10.1007/s40615-019-00590-z.
12. Prest J, Sathananthan M, Jeganathan N. Current Trends in Sepsis-Related Mortality in the United States. Crit Care Med. 2021;49(8):1276-84. Epub 2021/07/16. doi: 10.1097/CCM.0000000000005017.
13. Centers for Medicare & Medicaid Services. CMS Proposes Policies to Improve Patient Safety and Promote Health Equity: U.S. Department of Health & Human Services; 2023 [cited 2023 May 29, 2023]. Available from: https://www.cms.gov/newsroom/press-releases/cms-proposes-policies-improve-patient-safety-and-promote-health-equity.
14. Barnato AE, Alexander SL, Linde-Zwirble WT, Angus DC. Racial variation in the incidence, care, and outcomes of severe sepsis: analysis of population, patient, and hospital characteristics. Am J Respir Crit Care Med. 2008;177(3):279-84. Epub 2007/11/03. doi: 10.1164/rccm.200703-480OC.
15. Motheral BR, Fairman KA. The use of claims databases for outcomes research: rationale, challenges, and strategies. Clin Ther. 1997;19(2):346-66. Epub 1997/03/01. doi: 10.1016/s0149-2918(97)80122-1.
16. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care. 2013;51(8 Suppl 3):S30-7. Epub 2013/06/19. doi: 10.1097/MLR.0b013e31829b1dbd.
17. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013;20(1):144-51. Epub 2012/06/27. doi: 10.1136/amiajnl-2011-000681.
18. Ziegler J, Rush BNM, Gottlieb ER, Celi LA, Armengol de la Hoz MA. High resolution data modifies intensive care unit dialysis outcome predictions as compared with low resolution administrative data set. PLOS Digit Health. 2022;1(10):e0000124. Epub 2023/02/23. doi: 10.1371/journal.pdig.0000124.
19. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):E215-20. Epub 2000/06/14. doi: 10.1161/01.cir.101.23.e215.
20. Johnson AE, Stone DJ, Celi LA, Pollard TJ. The MIMIC Code Repository: enabling reproducibility in critical care research. J Am Med Inform Assoc. 2018;25(1):32-9. Epub 2017/10/17. doi: 10.1093/jamia/ocx084.

21. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10(1):1. Epub 2023/01/04. doi: 10.1038/s41597-022-01899-x.
22. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med. 2007;4(10):e296. Epub 2007/10/19. doi: 10.1371/journal.pmed.0040296.
23. Colleges AMAaAoAM. Advancing Health Equity: Guide on Language, Narrative and Concepts. American Medical Association; 2021. p. 54.
24. Yang J, Li Y, Liu Q, Li L, Feng A, Wang T, et al. Brief introduction of medical database and data mining technology in big data era. J Evid Based Med. 2020;13(1):57-69. Epub 2020/02/23. doi: 10.1111/jebm.12373.
25. Wu WT, Li YJ, Feng AZ, Li L, Huang T, Xu AD, et al. Data mining in clinical big data: the frequently used databases, steps, and methodological models. Mil Med Res. 2021;8(1):44. Epub 2021/08/13. doi: 10.1186/s40779-021-00338-z.
26. Vail EA, Gershengorn HB, Wunsch H, Walkey AJ. Attention to Immortal Time Bias in Critical Care Research. Am J Respir Crit Care Med. 2021;203(10):1222-9. Epub 2021/03/25. doi: 10.1164/rccm.202008-3238CP.
27. Yadav K, Lewis RJ. Immortal Time Bias in Observational Studies. JAMA. 2021;325(7):686-7. Epub 2021/02/17. doi: 10.1001/jama.2020.9151.
28. Fan E, Del Sorbo L, Goligher EC, Hodgson CL, Munshi L, Walkey AJ, et al. An Official American Thoracic Society/European Society of Intensive Care Medicine/Society of Critical Care Medicine Clinical Practice Guideline: Mechanical Ventilation in Adult Patients with Acute Respiratory Distress Syndrome. Am J Respir Crit Care Med. 2017;195(9):1253-63. Epub 2017/05/02. doi: 10.1164/rccm.201703-0548ST.
29. Ervin JN, Rentes VC, Dibble ER, Sjoding MW, Iwashyna TJ, Hough CL, et al. Evidence-Based Practices for Acute Respiratory Failure and Acute Respiratory Distress Syndrome: A Systematic Review of Reviews. Chest. 2020;158(6):2381-93. Epub 2020/07/20. doi: 10.1016/j.chest.2020.06.080.
30. Section 5: Dialysis Interventions for Treatment of AKI. Kidney Int Suppl (2011). 2012;2(1):89-115. Epub 2012/03/01. doi: 10.1038/kisup.2011.35.
31. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. Crit Care Med. 2021;49(11):e1063-e143. Epub 2021/10/05. doi: 10.1097/CCM.0000000000005337.
32. 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. 2009;37(5):1649-54. Epub 2009/03/28. doi: 10.1097/CCM.0b013e31819def97.
33. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. Journal of Chronic Diseases. 1987;40(5):373-83. doi: 10.1016/0021-9681(87)90171-8.
34. Pollard TJ, Johnson AEW, Raffa JD, Mark RG. tableone: An open source Python package for producing summary statistics for research papers. JAMIA Open. 2018;1(1):26-31. Epub 2018/05/23. doi: 10.1093/jamiaopen/ooy012.
35. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12(null):2825–30.
36. Chen T, Guestrin C, editors. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016.
37. Slim MAM, Lala HM, Barnes N, Martynoga RA. Maori health outcomes in an intensive care unit in Aotearoa New Zealand. Anaesth Intensive Care. 2021;49(4):292-300. Epub 2021/06/23. doi: 10.1177/0310057X21989715.
38. Orsini N, Moore A, Wolk A. Interaction Analysis Based on Shapley Values and Extreme Gradient Boosting: A Realistic Simulation and Application to a Large Epidemiological Prospective Study. Front Nutr. 2022;9:871768. Epub 2022/08/05. doi: 10.3389/fnut.2022.871768.
39. Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? in Advances in Neural Information Processing Systems 35 (NeurIPS 2022). Koyejo, S. and Mohamed, S. et al., eds. 2022.
40. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Francisco, California, USA: Association for Computing Machinery; 2016. p. 785–94.
41. van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. J Clin Epidemiol. 2012;65(2):126-31. Epub 2011/11/15. doi: 10.1016/j.jclinepi.2011.08.002.
42. Khera R, Angraal S, Couch T, Welsh JW, Nallamothu BK, Girotra S, et al. Adherence to Methodological Standards in Research Using the National Inpatient Sample. JAMA. 2017;318(20):2011-8. Epub 2017/11/29. doi: 10.1001/jama.2017.17653.
43. Frei CR, Mortensen EM, Copeland LA, Attridge RT, Pugh MJ, Restrepo MI, et al. Disparities of care for African-Americans and Caucasians with community-acquired pneumonia: a retrospective cohort study. BMC Health Serv Res. 2010;10:143. Epub 2010/05/29. doi: 10.1186/1472-6963-10-143.
44. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and Opportunities of Big Data in Health Care: A Systematic Review. JMIR Med Inform. 2016;4(4):e38. Epub 2016/11/23. doi: 10.2196/medinform.5359.
45. Kollef MH, Micek ST. Using protocols to improve patient outcomes in the intensive care unit: focus on mechanical ventilation and sepsis. Semin Respir Crit Care Med. 2010;31(1):19-30. Epub 2010/01/27. doi: 10.1055/s-0029-1246286.
46. Engoren M, Arslanian-Engoren C. Race and sex based disparities in sepsis. Heart Lung. 2022;52:37-41. Epub 2021/11/28. doi: 10.1016/j.hrtlng.2021.11.001.
47. Sohn H. Racial and Ethnic Disparities in Health Insurance

Coverage: Dynamics of Gaining and Losing Coverage over the Life-Course. Popul Res Policy Rev. 2017;36(2):181-201. Epub 2017/04/04. doi: 10.1007/s11113-016-9416-y.
48. Valley TS, Admon AJ, Zahuranec DB, Garland A, Fagerlin A, Iwashyna TJ. Estimating ICU Benefit: A Randomized Study of Physicians. Crit Care Med. 2019;47(1):62-8. Epub 2018/10/12. doi: 10.1097/CCM.0000000000003473.
49. Sheikh F, Douglas W, Catenacci V, Machon C, Fox-Robichaud AE. Social Determinants of Health Associated With the Development of Sepsis in Adults: A Scoping Review. Crit Care Explor. 2022;4(7):e0731. Epub 2023/02/24. doi: 10.1097/CCE.0000000000000731.

Appendix A

Supplementary Table. Variables used in the analysis to adjust the models

Variable	Description	Handling of missing values
Treatment	Within the eligibility window	
IMV (any invasive ventilation)	p_aO_2/FiO_2 ratio ≤ 300 (average of the day) or Respiratory rate ≥ 20 bpm (average of the day) as indication for hypoxemic respiratory failure p_aCO_2 ≥ 60 mmHg (minimum of the day) indication for hypercapnic respiratory failure	N/A
RRT (any acute continuous renal replacement method)	Urine output / Weight ≤ 12 L/kg (total of the day) as indication for acute renal failure according to KDIGO (2012) AKIN 3 Potassium ≥ 6.5 mEq/L (minimum of the day) as indication for persistent hyperkalemia pH ≤ 7.2 and bicarbonate ≤ 12 mEq/L (both maximum of the day) as indication for persistent metabolic acidosis	N/A
VP ((nor-) epinephrine, phenylephrine, or vasopressin)	Mean blood pressure ≤ 65 mmHg (average of the day) or lactate ≥ 2 mmol/L (minimum of the day) indication according to Sepsis-3 guidelines	N/A
Controls	Patients not receiving IMV, RRT, or VP within the eligibility window	N/A
Outcomes		
Primary: Treatment initiation	As provided by dataset	N/A
Covariates		
Age	At admission	N/A
Sex	As provided by dataset	N/A
Race-Ethnicity	0 if Racial-ethnic group (self-reported), 1 if White	Exclusion of Group “Other”
Insurance	As provided by dataset, Medicare/Medicaid or other	N/A
English Proficiency	As provided by dataset, English proficiency or limited	N/A
Year group	As provided by dataset, bi-yearly bins	N/A

Elective admission	Admission categorized as 'ELECTIVE', or 'SURGICAL SAME DAY ADMISSION'	N/A
Surgery during admission	As provided by the OASIS score	N/A
Coding status	Binary, full code on admission and upon discharge	N/A
SOFA	SOFA score with each of its subcomponents: on admission and for the selected 24 hours, aggregated by the maximum value	Assumption of best possible value in case of missing variables as provided by MIMIC
Charlson comorbidity index	As provided by dataset	N/A
Fluids	Sum of the volume administered during the selected 24 hours	N/A
Vital signs	Respiratory rate, heart rate, mean blood pressure, temperature, and SpO_2; Mean during the first 24 hours	Missing values replaced with normal values: Respiratory rate 15/min; Heart rate 90/bpm; Mean blood pressure 85mmHg; Temperature 36.5°C; SpO2 95%
Laboratory values	Minimum value during the selected 24 hours: Sodium, pH, p_aO_2, fibrinogen, cortisol, hemoglobin Maximum value during the selected 24 hours: Glucose, potassium, INR, lactate, p_aCO_2	Missing values replaced with normal values: pO_2 90 mmHg; pCO_2 40mmHg; pH 7.35; Lactate 1.05 mmol/L; Glucose 95 mg/dL; Sodium 140 mEq/L; Potassium: 3.5 mEq/L; Cortisol: 20 µg/dL; Fibrinogen: 200 mg/dL; INR: 1.1
Hypertension	ICD-10 codes I11.X-I16X and I.70X	N/A
Congestive heart failure	ICD-10 codes I50.X, I11.0X, I27.X, I42.X, I43.X, I51.7X	N/A
COPD	ICD-10 codes J41.X-J47.X	N/A
Asthma	ICD-10 codes J84.1X	N/A
Coronary artery disease	ICD-10 codes I20.X-I25.X	N/A
Chronic kidney disease	ICD-10 codes N18.1X-N18.6X	N/A
Diabetes type	ICD-10 codes E08.X-E11.X, and E13.X	N/A
Connective tissue disease	ICD-10 codes L94.0X, L94.1X, L94.3X, M05.X, M06.X, M08.X, M12.0X, M12.3X, M30.X-M31.3X, M32.X-M35.X, M45, M46.1X, M46.8X, or M46.9X	N/A

Pneumonia on admission	ICD codes J09.X, J1X.X, J85.X, or J86.X if listed among top 3 diagnoses by billing department	N/A
Urinary tract infection on admission	ICD codes N30.0X, or N39.0X if listed among top 3 diagnoses by billing department	N/A
Biliary tract infection on admission	ICD codes K81.X, K83.0X, or K85.1X if listed among top 3 diagnoses by billing department	N/A
Skin infection on admission	ICD codes L0X.X if listed among top 3 diagnoses by billing department	N/A

Legend: "selected 24 hours" means the day of treatment initiation, if the patient was treated, or the day with worst SOFA score, if patient was in the control group

Abbreviations: N/A, not applicable; IMV, invasive mechanical ventilation; RRT, renal replacement therapy; VP, vasopressor; SOFA, Sequential Organ Failure Assessment; OASIS, Oxford Acute Severity of Illness Score; COPD, chronic obstructive pulmonary disease

Supplementary Table 2. Numerical result of the models' main analysis in the strict inclusion criteria cohorts.

		IMV		RRT		VP	
Cohort	**Comparison**	**XGBoost**	**LogReg**	**XGBoost**	**LogReg**	**XGBoost**	**LogReg**
1	**REG vs. WG (ref)**	0.87 (0.83 - 0.94)	1.01 (0.97 - 1.06)	1 (0.96 - 1.09)	0.77 (0.62 - 1.03)	0.87 (0.76 - 0.94)	0.73 (0.67 - 0.84)
	WG (ref)	1.00	1.00	1.00	1.00	1.00	1.00
	Black	0.97 (0.9 - 1.01)	1.16 (1.09 - 1.22)	1 (0.99 - 1.04)	0.89 (0.7 - 1.11)	0.98 (0.91 - 1)	0.75 (0.64 - 0.87)
	Hispanic	0.98 (0.91 - 1)	0.84 (0.77 - 0.95)	1 (1 - 1.04)	0.88 (0.73 - 1.13)	1 (0.97 - 1.01)	0.88 (0.76 - 1.07)
	Asian	1 (0.97 - 1.01)	0.9 (0.82 - 1.02)	1 (1 - 1.04)	0.93 (0.79 - 1.07)	1 (0.99 - 1.04)	0.94 (0.81 - 1.3)
2	**REG vs. WG (ref)**	0.86 (0.8 - 0.92)	0.91 (0.87 - 0.95)	1 (0.95 - 1.09)	0.85 (0.68 - 1.08)	0.86 (0.75 - 0.94)	0.62 (0.54 - 0.76)
	WG (ref)	1.00	1.00	1.00	1.00	1.00	1.00
	Black	0.95 (0.89 - 1.01)	1.1 (1.04 - 1.19)	1 (0.99 - 1.08)	0.92 (0.74 - 1.28)	0.96 (0.91 - 1.01)	0.66 (0.57 - 0.81)
	Hispanic	0.92 (0.84 - 0.98)	0.66 (0.59 - 0.74)	1 (1 - 1.01)	0.88 (0.68 - 1.11)	1 (0.98 - 1.01)	0.8 (0.66 - 1.01)
	Asian	0.99 (0.92 - 1)	0.83 (0.74 - 0.97)	1 (1 - 1.03)	1.06 (0.91 - 1.19)	1 (0.98 - 1.03)	0.81 (0.67 - 1.09)
3	**REG vs. WG (ref)**	0.78 (0.7 - 0.85)	0.83 (0.78 - 0.9)	0.99 (0.94 - 1.05)	0.86 (0.7 - 1.07)	0.91 (0.81 - 0.96)	0.71 (0.6 - 0.88)
	WG (ref)	1.00	1.00	1.00	1.00	1.00	1.00
							4
	Black	0.95 (0.89 - 0.99)	0.99 (0.93 - 1.06)	1 (0.98 - 1.07)	0.83 (0.66 - 1.11)	0.99 (0.91 - 1.01)	0.77 (0.6 - 0.96)
	Hispanic	0.86 (0.78 - 0.96)	0.58 (0.53 - 0.69)	1 (1 - 1.03)	1.01 (0.82 - 1.31)	1 (0.96 - 1.01)	0.85 (0.69 - 1.01)
	Asian	0.98 (0.93 - 1)	0.93 (0.82 - 1.15)	1 (1 - 1.1)	1.25 (1.03 - 1.52)	1 (0.95 - 1.01)	0.9 (0.72 - 1.18)
4	**REG vs. WG (ref)**	0.8 (0.72 - 0.87)	0.77 (0.73 - 0.86)	1 (0.94 - 1.06)	0.87 (0.7 - 1.12)	0.95 (0.87 - 1.01)	0.81 (0.7 - 1.07)
	WG (ref)	1.00	1.00	1.00	1.00	1.00	1.00
	Black	0.91 (0.82 - 0.97)	0.86 (0.78 - 0.98)	1 (0.97 - 1.06)	0.82 (0.66 - 1.05)	1 (0.96 - 1.01)	0.9 (0.71 - 1.19)
	Hispanic	0.98 (0.91 - 1)	0.71 (0.61 - 0.94)	1 (1 - 1.03)	1.16 (0.92 - 1.43)	1 (0.95 - 1)	0.79 (0.64 - 1.03)
	Asian	0.98 (0.91 - 1)	0.93 (0.82 - 1.15)	1 (1 - 1.05)	1.2 (1.01 - 1.45)	1 (1 - 1.01)	0.93 (0.79 - 1.15)

Abbreviations: REG, Racial-ethnic groups; WG, White group; LogReg, logistic regression; ref, reference; IMV, invasive mechanical ventilation; RRT, renal replacement therapy; VP, vasopressor; SOFA, Sequential Organ Failure Assessment; OASIS, Oxford Acute Severity of Illness Score; COPD, chronic obstructive pulmonary disease

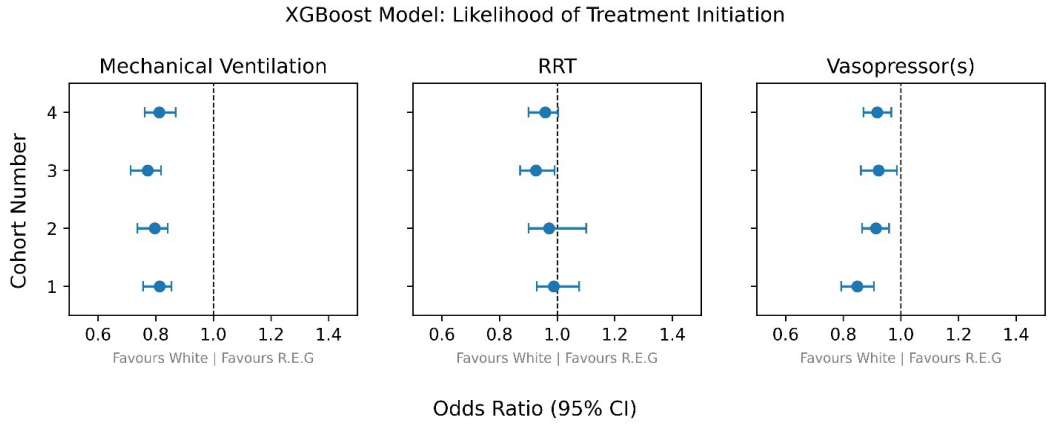
Supplementary Table 3. Performance measures of the models of the main analysis

Treatment		**IMV**		**RRT**		**VP**	
	Model	XGBoost	LogReg	XGBoost	LogReg	XGBoost	LogReg
Cohort	Metric						
1	AUC ↑	0.970	0.937	0.925	0.924	0.973	0.971
	Brier ↓	0.066	0.100	0.039	0.035	0.064	0.061
2	AUC ↑	0.954	0.918	0.914	0.925	0.965	0.969
	Brier ↓	0.085	0.110	0.067	0.053	0.071	0.069
3	AUC ↑	0.958	0.917	0.900	0.907	0.952	0.947
	Brier ↓	0.077	0.102	0.079	0.071	0.086	0.087
4	AUC ↑	0.962	0.934	0.915	0.922	0.945	0.945
	Brier ↓	0.068	0.084	0.086	0.079	0.093	0.088

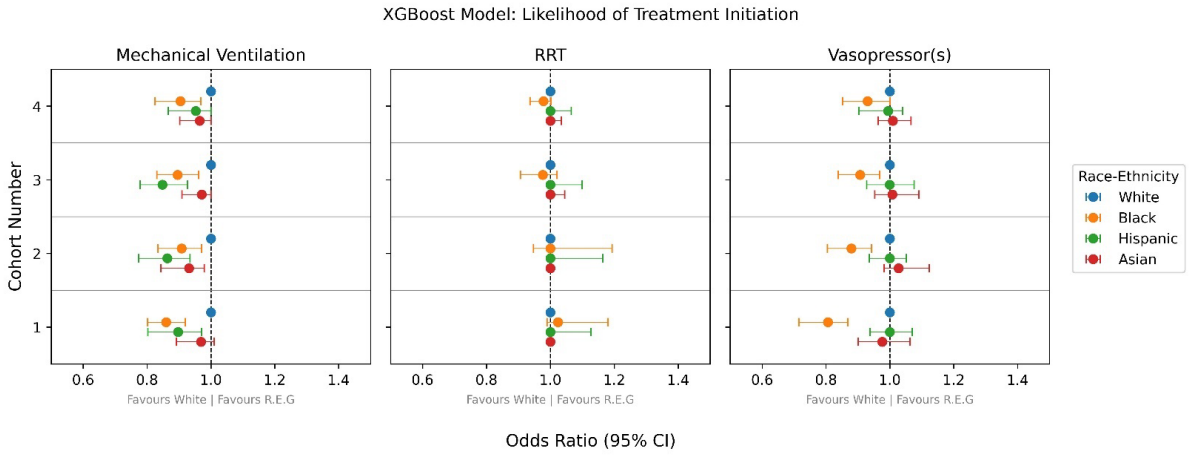
Legend: ↑ and ↓ indicate the more favorable direction of a metric, i.e., the higher the AUC, the better a model's discrimination, and the lower the Brier score, the better a model's calibration. Models trained with 80% of the dataset and tested in the remaining 20%.

Abbreviations: IMV, invasive mechanical ventilation; RRT, renal replacement therapy; VP, use of vasopressors; LogReg, Logistic Regression; AUC, area under the receiver operating characteristic (ROC) curve; Brier, Brier score; d, day

Supplementary Figure 1a. Sensitivity analysis with logistic regression results for likelihood of treatment initiation White group versus Racial-ethnic group in the lenient inclusion criteria cohorts.

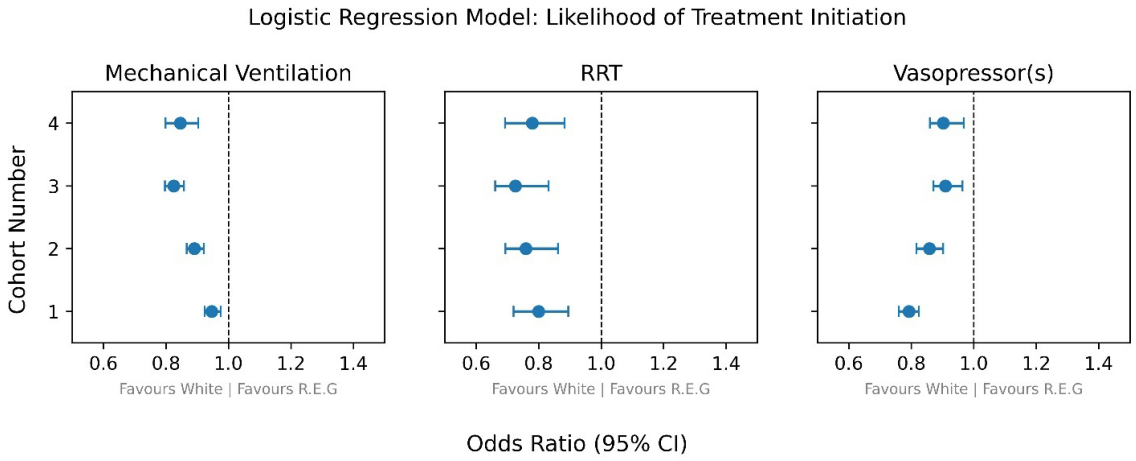


Supplementary Figure 1b. Sensitivity analysis with logistic regression results for likelihood of treatment initiation White group versus Racial-ethnic groups in the lenient inclusion criteria cohorts.

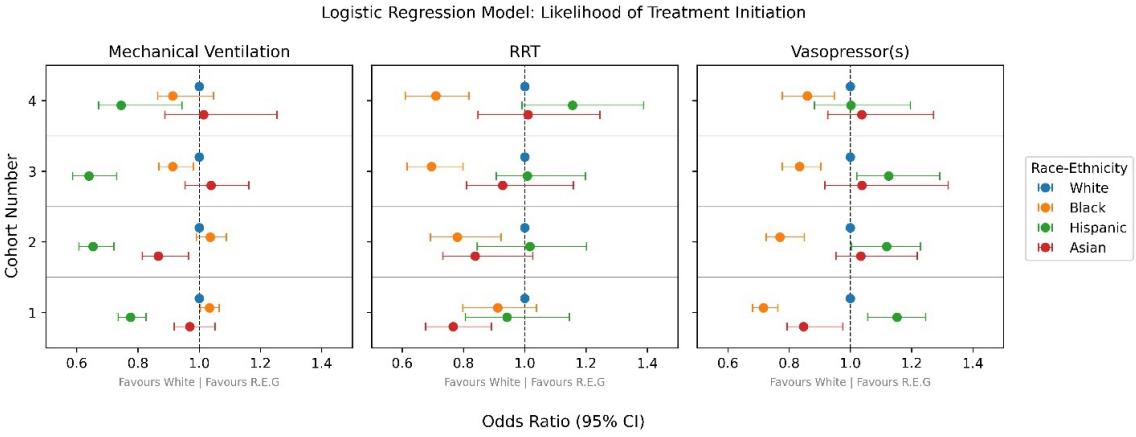


Abbreviations: RRT, renal replacement therapy; REG, racial-ethnic group; CI, confidence interval

Supplementary Figure 2a. Sensitivity analysis with XGBoost results for likelihood of treatment initiation White group versus Racial-ethnic group in the lenient inclusion criteria cohorts.



Supplementary Figure 2b. Sensitivity analysis with XGBoost results for likelihood of treatment initiation White group versus Racial-ethnic groups in the lenient inclusion criteria cohorts.



Abbreviations: RRT, renal replacement therapy; REG, racial-ethnic group; CI, confidence interval