JAMA | Original Investigation | CARING FOR THE CRITICALLY ILL PATIENT

Individualized Treatment Effects of Oxygen Targets in Mechanically Ventilated Critically Ill Adults

Kevin G. Buell, MBBS; Alexandra B. Spicer, MS; Jonathan D. Casey, MD, MSc; Kevin P. Seitz, MD, MSc; Edward T. Qian, MD, MSc; Emma J. Graham Linck, MS; Wesley H. Self, MD, MPH; Todd W. Rice, MD, MSc; Pratik Sinha, MBChB, PhD; Paul J. Young, MD, PhD; Matthew W. Semler, MD, MSc; Matthew M. Churpek, MD, MPH, PhD

IMPORTANCE Among critically ill adults, randomized trials have not found oxygenation targets to affect outcomes overall. Whether the effects of oxygenation targets differ based on an individual's characteristics is unknown.

OBJECTIVE To determine whether an individual's characteristics modify the effect of lower vs higher peripheral oxygenation-saturation (Spo_2) targets on mortality.

DESIGN, SETTING, AND PARTICIPANTS A machine learning model to predict the effect of treatment with a lower vs higher Spo_2 target on mortality for individual patients was derived in the Pragmatic Investigation of Optimal Oxygen Targets (PILOT) trial and externally validated in the Intensive Care Unit Randomized Trial Comparing Two Approaches to Oxygen Therapy (ICU-ROX) trial. Critically ill adults received invasive mechanical ventilation in an intensive care unit (ICU) in the United States between July 2018 and August 2021 for PILOT (n = 1682) and in 21 ICUs in Australia and New Zealand between September 2015 and May 2018 for ICU-ROX (n = 965).

EXPOSURES Randomization to a lower vs higher Spo_2 target group.

MAIN OUTCOME AND MEASURE 28-Day mortality.

RESULTS In the ICU-ROX validation cohort, the predicted effect of treatment with a lower vs higher Spo_2 target for individual patients ranged from a 27.2% absolute reduction to a 34.4% absolute increase in 28-day mortality. For example, patients predicted to benefit from a lower Spo_2 target had a higher prevalence of acute brain injury, whereas patients predicted to benefit from a higher Spo_2 target had a higher prevalence of sepsis and abnormally elevated vital signs. Patients predicted to benefit from a lower Spo_2 target experienced lower mortality when randomized to the lower Spo_2 group, whereas patients predicted to benefit from a higher Spo_2 target experienced lower mortality when randomized to the higher Spo_2 group (likelihood ratio test for effect modification $P = .02$). The use of a Spo_2 target predicted to be best for each patient, instead of the randomized Spo_2 target, would have reduced the absolute overall mortality by 6.4% (95% CI, 1.9%-10.9%).

CONCLUSION AND RELEVANCE Oxygenation targets that are individualized using machine learning analyses of randomized trials may reduce mortality for critically ill adults. A prospective trial evaluating the use of individualized oxygenation targets is needed.

Editorial page 1179

Related articles pages 1185 and 1225

Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Kevin G. Buell, MBBS, Division of Pulmonary and Critical Care, University of Chicago, 5841 S Maryland Ave, MC 6076, Chicago, IL 60637 (kevin.buell@uchicagomedicine.org).

Section Editor: Christopher Seymour, MD, Associate Editor, *JAMA* (christopher.seymour@jamanetwork.org).

JAMA. 2024;331(14):1195-1204. doi:10.1001/jama.2024.2933
Published online March 19, 2024.

Mechanical ventilation involves titrating the fraction of inspired oxygen to maintain arterial oxygenation, but the oxygenation target that results in the best outcomes for critically ill adults is unknown. Multiple randomized trials reported similar outcomes for patients allocated to lower oxygenation targets that prioritize avoiding hyperoxia or to higher oxygenation targets that prioritize avoiding hypoxemia.[1-7]

Randomized trials report the average effect of treatment on outcomes for the overall trial population. Retrospective studies and post hoc analyses of randomized trials suggest that the optimal oxygenation target may differ for patients with hypoxemic brain injury, traumatic brain injury, shock, or sepsis.[8-13] Such nonrandom variation in the magnitude or direction of treatment effect, referred to as *heterogeneity of treatment effect*, is not adequately addressed by traditional one-at-a-time subgroup analyses because each patient has multiple characteristics that may each influence the effect of the intervention on outcomes.[14,15] Data from randomized trials can be analyzed using machine learning methods to predict individualized treatment effects, defined as the *predicted difference* in outcome between 2 treatments for an individual patient based on his or her unique characteristics.[16-19]

To test the hypothesis that the effect of peripheral oxygenation-saturation (Spo_2) targets on mortality would differ based on individual patient characteristics, this study derived and externally validated predicted individualized treatment effects using 2 temporally and geographically distinct randomized trials of lower vs higher Spo_2 targets in critically ill patients receiving mechanical ventilation.

Key Points

Question Among critically ill adults, do the effects of peripheral oxygen-saturation (Spo_2) targets on outcomes differ based on an individual's characteristics?

Findings A machine learning model derived in one randomized trial and validated in another found that the predicted effect of lower vs higher Spo_2 targets ranged from a 27% absolute reduction to a 34% absolute increase in 28-day mortality and differed significantly based on an individual's characteristics.

Meaning The effect of oxygen-saturation targets on mortality varied by patients' individual characteristics.

Methods

Study Population and Outcomes

This secondary analysis used data from 2 randomized trials (**Table 1**; eMethods in Supplement 1). The Pragmatic Investigation of Optimal Oxygen Targets (PILOT) was a cluster-randomized, cluster-crossover trial that compared a lower Spo_2 target (90%; goal range, 88% to 92%), an intermediate Spo_2 target (94%; goal range, 92% to 96%), and a higher Spo_2 target (98%; goal range, 96% to 100%) among 2541 patients receiving mechanical ventilation in the emergency department and medical intensive care unit (ICU) at an academic medical center in the United States between July 1, 2018, and August 31, 2021.[2] Patients in the intermediate Spo_2 target group of the PILOT trial were not included in this study. The primary outcome of the PILOT trial was the number of days alive and free of mechanical ventilation (ventilator-free days) through day 28. The secondary outcome of the PILOT trial was in-hospital mortality by day 28. Patients in the PILOT trial who were discharged alive prior to 28 days were presumed to be alive on day 28.[2]

The Intensive Care Unit Randomized Trial Comparing Two Approaches to Oxygen Therapy (ICU-ROX) was a multicenter, parallel-group, randomized trial that compared conservative-oxygen therapy vs usual-oxygen therapy among 965 mechanically ventilated patients in 21 ICUs in Australia and New Zealand between September 2015 and May 2018.[3] Patients in the conservative-oxygen group received the lowest fraction of inspired oxygen that maintained a Spo_2 of more than 90% with an alarm set to sound if the Spo_2 was 96% or higher. Patients in the usual-oxygen group received oxygen therapy at the discretion of the treating clinicians targeting Spo_2 values between 91% to 100%. The primary outcome of the ICU-ROX trial was ventilator-free days through day 28; secondary outcomes included mortality at multiple time points.[3]

This secondary analysis used deidentified data from patients randomized to the lower or higher Spo_2 target groups in the PILOT trial and all patients in the ICU-ROX trial. The primary outcome of the current analysis was 28-day mortality, a secondary outcome in the original PILOT and ICU-ROX trials. Both trials were approved by institutional review boards in the countries in which the trials were conducted. This study was approved with a waiver of informed consent by the University of Wisconsin Institutional Review Board (IRB: 2019-1258).

Statistical Analysis

Study Design

This study used an effect-based analysis of heterogeneity of treatment effect, as described in the effect-modeling approach from the Predictive Approaches to Treatment Effect Heterogeneity (PATH) statement.[15] The study report followed the guidelines from the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement. The machine learning model for individualized treatment effect was derived in the PILOT derivation cohort and then externally validated in the ICU-ROX validation cohort. The PILOT and ICU-ROX trials were temporally distinct (2018-2021 vs 2015-2018) and geographically distinct (United States vs Australia and New Zealand).

Analyzing the datasets for each trial separately was designed to demonstrate the generalizability of the individualized treatment effect predictions beyond the original trial (see the eMethods in Supplement 1 for details). Within their respective trials, the higher Spo_2 target group of PILOT and the usual-oxygen group of ICU-ROX were treated as reference (control) groups and are henceforth referred to as the *higher Spo_2 group*. The lower Spo_2 target group of PILOT and the

Table 1. Characteristics of the PILOT and ICU-ROX Trials

Trial characteristics	PILOT trial (n = 2541)	ICU-ROX trial (n = 965)
Randomization	Cluster-randomized, cluster crossover	Individual patient, parallel group
Interventions and groups	Lower Spo_2 target: 90% (range, 88%-92%)	Lower Spo_2 target: 91%-96%
	Intermediate Spo_2 target: 94% (range, 92%-96%)	No intermediate group
	Higher Spo_2 target: 98% (range, 96%-100%)	Higher Spo_2 target: 91%-100%
No. of centers	1	21
Location	United States	Australia and New Zealand
Duration	2018-2021	2015-2018
Location of patient enrollment	Emergency department or ICU	ICU
Time to enrollment	At initiation of mechanical ventilation	A median of 3 h after initiation of mechanical ventilation
Primary outcome	Ventilator-free days	Ventilator-free days
Secondary outcome of mortality	28-d mortality	Mortality at multiple time points
Trial result	No between-group difference in outcomes	No between-group difference in outcomes

Abbreviations: ICU-ROX, Intensive Care Unit Randomized Trial Comparing Two Approaches to Oxygen Therapy; PILOT, Pragmatic Investigation of Optimal Oxygen Targets; Spo_2, peripheral oxygenation-saturation by pulse oximetry.

conservative-oxygen group of ICU-ROX were treated as comparator (intervention) groups and are henceforth referred to as the *lower Spo_2 group*. Because the intracluster correlation in the PILOT trial was only 0.001, no adjustments were made for the cluster-randomized, cluster-crossover design. The main analysis reports the effect of randomization to a lower Spo_2 group compared with randomization to a higher Spo_2 group on the primary outcome of 28-day mortality for individual patients in the ICU-ROX validation cohort.

Predictor Variables

Predictor variables included in the analyses were limited to baseline variables from the trials collected at the time of enrollment, prior to receipt of trial interventions. The following baseline variables were chosen a priori for inclusion as predictors in the model: patient demographics (age and sex), source of admission to the ICU (hospital ward, emergency department, operating room, or postanesthesia care unit or transfer from another ICU within the study hospital or from another hospital), ICU admitting diagnoses (acute hypoxic brain injury, acute nonhypoxic brain injury, cardiovascular disease, respiratory disease, or sepsis), vital signs (heart rate, mean arterial pressure, respiratory rate, and temperature), serum creatinine levels, receipt of vasopressors or inotropes, time from mechanical ventilation to randomization, mode of mechanical ventilation (volume-based mode, pressure-based mode, pressure regulated volume control, or missing or other ventilation mode indicator), positive end-expiratory pressure, arterial partial pressure of carbon dioxide ($Paco_2$), missing indicator for $Paco_2$, and predicted risk of 28-day mortality (see eMethods in Supplement 1).

The predicted risk of 28-day mortality was derived in each trial separately using the nonrespiratory Sequential Organ Failure Assessment (SOFA) score[20] in the PILOT trial and the Acute Physiology and Chronic Health Evaluation (APACHE) II score[21] in the ICU-ROX trial (see eMethods in Supplement 1). The predicted risk of 28-day mortality was included as a predictor variable in the models because variation in patients' baseline risk of the primary outcome across a trial population may contribute to heterogeneity of treatment effect.[22] Diagnoses were harmonized between the 2 trials using the categories listed in eMethods in Supplement 1.[23] Missing predictor data in the PILOT and ICU-ROX trials were imputed using bagged trees derived from the PILOT trial (see the eMethods in Supplement 1). Nonmissing baseline predictor variables did not undergo any preprocessing.

Model Derivation

Six machine learning algorithms were evaluated in the PILOT dataset using 5-fold cross-validation based on the mean out-of-sample adjusted qini statistic. Only the best performing algorithm, Rboost, was selected for model derivation on the full PILOT dataset and eventual external validation in the ICU-ROX validation cohort. Rboost is an XGBoost implementation of the R-learner. XGBoost is a tree-based nonparametric model that builds a collection of decision trees with advanced regularization to reduce overfitting.[24] The R-Learner framework uses the Robinson transformation to predict the treatment effect estimation as a function of baseline covariates (see the eMethods in Supplement 1).[25,26] Hyperparameters were tuned in the PILOT derivation cohort using 5-fold cross-validation. To improve stability, 5 Rboost models were constructed in the PILOT derivation cohort using different seed initialization in sequential runs (see eMethods of Supplement 1).[27]

Model Validation

The Rboost models trained in the PILOT derivation cohort were applied to the ICU-ROX validation cohort. To obtain a value for the individualized treatment effect (the predicted difference in 28-day mortality from the use of a lower vs higher Spo_2 target) for each patient in the ICU-ROX validation cohort, a mean was calculated from the 5 individualized treatment effect values of the iterations of the 5 models.

Patient Characteristics Associated With Predicted Individualized Treatment Effects

All patients in the ICU-ROX validation cohort were ranked by their predicted individualized treatment effect and categorized

into 3 equally-sized groups using cutoffs determined in the ICU-ROX validation cohort (see eMethods in Supplement 1).[15] Patients in the lower third of predicted individualized treatment effect were those for whom the model predicted a benefit from a lower Spo_2 target. Patients in the middle third of predicted individualized treatment effect were those for whom the model predicted the least difference in outcomes between a lower vs higher Spo_2 target. Patients in the upper third of predicted individualized treatment effect were those for whom the model predicted a benefit from a higher Spo_2 target.

Patient characteristics and outcomes were compared between the lower, middle, and upper thirds using Kruskal-Wallis rank-sum and χ^2 tests based on variable distributions. A variable importance plot was used to evaluate the relative weights of the baseline patient characteristics used by the Rboost model. To illustrate the relative influence of each baseline patient characteristic in predicting the individualized treatment effect, the value of each baseline patient characteristic was sequentially replaced 1 at a time by the median value from the PILOT derivation cohort while all other baseline patient characteristics were held constant. For continuous variables, partial dependence plots were used to explore the effect of each baseline patient characteristic on the average marginal effect of a lower Spo_2 target on 28-day mortality. See the eMethods in Supplement 1 for further methods regarding the development of the variable importance plot, spaghetti plot, individual patient example plots, and partial dependence plots.

Testing for Effect Modification

To assess whether the predicted effect of treatment on outcomes for individual patients generated by the model modified the actual effect of lower vs higher Spo_2 targets on outcomes in the ICU-ROX trial, in a logistic regression model with the dependent variable of 28-day mortality, a likelihood ratio test evaluated the significance of the interaction term between the thirds of predicted individualized treatment effect (lower third, model predicted benefit from a lower Spo_2 target; middle third, model did not predict benefit from a lower or higher Spo_2 target; and upper third, model predicted benefit from a higher Spo_2 target) and the randomized trial group assignment. Statistical significance was indicated by a 2-sided α level of .05.

Outcome Reporting

For patients in the lower, middle, and upper third of predicted individualized treatment effect in the ICU-ROX trial, the absolute risk difference in 28-day mortality within the third was calculated as the difference in the incidence of 28-day mortality for patients randomized to the lower Spo_2 group vs patients randomized to the higher Spo_2 group.

Assessing Model Performance

The qini curve, adjusted qini value, and C-for-benefit were used to quantify the model's ability to discriminate between patients likely to benefit from a lower vs a higher Spo_2 target.[28,29] The qini curve and adjusted qini value represent the positive gain or "uplift" in 28-day survival that resulted from a patient being randomized to the lower Spo_2 group based on the ordering of patients by their predicted likelihood to benefit from a lower Spo_2 target. The C-for-benefit assesses model performance by giving the probability of concordance between predicted and observed benefit. An adjusted qini value greater than 0 and C-for-benefit value greater than 0.5 provides evidence of discrimination that is better than chance. See the eMethods in Supplement 1 for further details. To quantify model calibration, the absolute risk difference in 28-day mortality within the lower, middle, and upper third was compared with the average predicted individualized treatment effect within the corresponding third.

Sensitivity Analyses

In 3 separate post hoc sensitivity analyses, model performance was assessed using (1) an alternative definition of sepsis in the ICU-ROX trial, (2) exclusion of patients' baseline risk of 28-day mortality as a predictor variable, and (3) reversing the order of the PILOT and ICU-ROX trials for model derivation and validation.

All analyses were performed using R version 3.6.3 (R Foundation for Statistical Computing).

Results

Patients

Among 1682 patients in the PILOT derivation cohort, 808 patients were in the lower Spo_2 group and 874 patients were in the higher Spo_2 group (eFigure 1 in Supplement 1). Among 965 patients in the ICU-ROX validation cohort, 484 patients were in the lower Spo_2 group and 481 patients were in the higher Spo_2 group. Patients in these temporally and geographically distinct cohorts differed in their baseline characteristics, receipt of vasopressor or inotropic support, mode of mechanical ventilation, predicted 28-day mortality risk, and incidence of the primary outcome of 28-day mortality (eTable 1 in Supplement 1).

Predicted Individualized Treatment Effects

Although the original trials demonstrated no statistically significant average treatment effect of lower vs higher Spo_2 groups on mortality at any time point, the individualized treatment effects for the 965 patients in the ICU-ROX validation cohort, which were predicted by the model derived in the PILOT derivation cohort, ranged from a 27.2% absolute reduction in 28-day mortality with use of a lower Spo_2 target to a 34.4% absolute reduction in mortality with use of a higher Spo_2 target (**Figure 1**).

Patient Characteristics Associated With Predicted Individualized Treatment Effects

Mean arterial pressure, heart rate, age, and $Paco_2$ were the most important characteristics in the model for predicting the individualized treatment effects (eFigure 2 in Supplement 1). The relative influence of the 10 most-important baseline characteristics on patients' predicted individualized

Figure 1. Predicted Individualized Treatment Effects of Lower vs Higher Oxygen-Saturation Targets on Mortality

A. The y-axis displays the effect of treatment with a lower vs a higher peripheral oxygen-saturation (SpO_2) target on 28-day mortality in the Intensive Care Unit Randomized Trial Comparing Two Approaches to Oxygen Therapy (ICU-ROX) trial. For 100 of the 965 patients in the ICU-ROX trial, a colored line displays how the treatment effect predicted for that patient changed from the average treatment effect (point estimate and 95% CI at the far left) to the final predicted individualized treatment effect (point estimate at the far right) with the sequential addition of the patient's values for the each of the 10 most-important baseline patient characteristics. B, Displays the distribution of predicted individualized treatment effects for all 965 patients in the ICU-ROX trial.

treatment effects is illustrated in Figure 1A. Similarly, eFigure 3 in Supplement 1 shows how the values for each baseline characteristic together determined the predicted individualized treatment effect for one example patient in each third of predicted individualized treatment effect. Partial dependence plots for each continuous predictor variable are in eFigure 4 in Supplement 1.

In the ICU-ROX validation cohort, patients in the lower third (predicted to benefit from a lower Spo_2 target) were older; were more likely to be male; and had a higher prevalence of hypoxic brain injury, nonhypoxic brain injury, and cardiovascular disease. Patients in the upper third (predicted to benefit from a higher Spo_2 target) were younger; had a higher prevalence of sepsis and respiratory disease; and had higher values of baseline heart rate, mean arterial pressure, respiratory rate, and temperature (**Table 2**). The separation in oxygen exposure between patients randomized to the lower Spo_2 group and patients randomized to the higher Spo_2 group was similar for patients in the lower, middle, and upper third of predicted individualized treatment effect (eTable 2 in Supplement 1).

Outcomes

In a logistic regression analysis of all patients in the ICU-ROX validation cohort, the patient's quantile of predicted individualized treatment effect (lower, middle, or higher third) significantly modified the observed effect of randomized Spo_2 group assignment on the incidence of the primary outcome of 28-day mortality (likelihood ratio test for effect modification $P = .02$; **Table 3**). Among patients in the lower third of predicted individualized treatment effect (for whom a lower Spo_2 target was predicted to be beneficial), 28-day mortality was 6.1% lower (95% CI, −4.3% to 16.5%) for those randomized to the lower Spo_2 group than those randomized to the higher Spo_2 group. Among patients in the middle third of predicted individualized treatment effect (for whom outcomes were predicted to be similar between lower and higher Spo_2 targets), mortality was similar between those randomized to the lower Spo_2 group (28.7%) vs the higher Spo_2 group (28.5%). Among patients in the upper third of predicted individualized treatment effect (for whom a higher Spo_2 target was predicted to be beneficial), mortality was 13.0% higher (95% CI, 3.5% to 22.6%) in those randomized to the lower Spo_2 group than those randomized to the higher Spo_2 group (**Figure 2**). Overall, mortality would have been 6.4% absolute percentage points lower (95% CI, 1.9% to 10.9%) if all patients in the ICU-ROX validation cohort had been treated with the Spo_2 target predicted to best for them by the model rather than the randomly assigned Spo_2 group (Table 3).

Model Performance

The uplift in the qini curve demonstrated the gain in 28-day survival that resulted from patients being randomized to the lower Spo_2 group relative to the ordering of patients by their predicted likelihood to benefit from a lower Spo_2 target (eFigure 5 in Supplement 1).[30] The adjusted qini value was 2.27 and C-for-benefit was 0.55 (bootstrapped 95% CI, 0.50 to 0.60), consistent with the model's ability to discriminate treatment effects better than random chance. The model was well calibrated (eFigure 6 in Supplement 1).

Sensitivity Analyses

In the post hoc sensitivity analyses, an alternative definition of sepsis in the ICU-ROX trial yielded similar results (see eTable 3 in Supplement 1). The performance of the model decreased without inclusion of patients' baseline risk of 28-day mortality as a predictor variable (see eTable 4 in Supplement 1). In deriving the model in the ICU-ROX trial and validating in the PILOT trial, Rboost was still the best performing model and heart rate, age, and creatinine values remained among the 5 most important variables. However, model performance decreased, likely due to the smaller sample size in the ICU-ROX derivation cohort (see eMethods in Supplement 1).

Discussion

In 2 temporally and geographically distinct randomized trials, patients' individual characteristics modified the effect of lower vs higher Spo_2 targets on 28-day mortality. For example, the use of a lower Spo_2 target may decrease mortality for patients with acute brain injury, whereas use of a higher Spo_2 target may decrease mortality for patients with sepsis and abnormally elevated vital signs. These findings suggest that the use of Spo_2 targets that are individualized using machine learning analyses of randomized trials may improve outcomes for critically ill adults receiving mechanical ventilation.

These findings increase the understanding of the effects of oxygen-saturation targets on clinical outcomes in several ways. First, the data suggest that heterogeneity in the effect of treatment with lower vs higher Spo_2 targets on mortality exists and is consistent across 2 temporally and geographically diverse randomized trials. Such heterogeneity of treatment effect may partly explain the findings, on average, of no difference between groups in multiple previous randomized trials of oxygen-saturation targets. Second, the patient characteristics and direction of effect modification identified by the model are consistent with the findings of several previous studies, including an individual-level patient data meta-analysis suggesting that patients with hypoxic brain injury may benefit from a lower Spo_2 target[31] and a secondary analysis of a trial suggesting that patients with sepsis may benefit from a higher Spo_2 target.[10] The findings that patients with higher blood pressure, heart rate, respiratory rate, and temperature experienced decreased mortality with a higher Spo_2 target supports prior research that baseline vital signs could inform differential treatment effect.[32] The selection of continuous patient characteristics as high variable importance compared with binary patient characteristics is additionally consistent across tree-based models like Rboost because continuous variables provide more levels on which a split can occur. Third, variation in the baseline risk of the primary outcome across a trial population may result in heterogeneity of treatment effects,[22] and the performance of models decreased in a sensitivity

Table 2. Baseline Characteristics of Patients in the ICU-ROX Validation Cohort by Predicted Individualized Treatment Effect

Characteristic	No. (%) of participants			*P* value[a]
	Lower third: predicted to benefit from lower Spo_2 target (n = 322)	Middle third: predicted to have similar outcomes with either target (n = 322)	Upper third: predicted to benefit from higher Spo_2 target (n = 321)	
Demographics				
Age, median (IQR), y	67 (55 to 74)	58 (45 to 68)	57 (49 to 65)	<.001
Sex				
Female	89 (27.6)	128 (39.8)	140 (43.6)	<.001
Male	233 (72.4)	194 (60.2)	181 (56.4)	<.001
Source of admission to ICU				
Hospital ward	44 (13.7)	58 (18.0)	87 (27.1)	<.001
Emergency department	155 (48.1)	120 (37.3)	124 (38.6)	.01
Operating room or postanesthesia care unit	101 (31.4)	109 (33.9)	85 (26.5)	.12
Transfer from another ICU or another hospital	22 (6.8)	35 (10.9)	25 (7.8)	.16
Time from initiation of mechanical ventilation to enrollment, median (IQR), h	3.6 (1.9 to 5.9)	3.2 (1.6 to 5.7)	2.6 (1.2 to 4.3)	<.001
ICU admitting diagnosis[b]				
Cardiovascular disease	94 (29.2)	63 (19.6)	48 (15.0)	<.001
Acute nonhypoxic brain injury	90 (28.0)	86 (26.7)	41 (12.8)	<.001
Acute hypoxic brain injury	84 (26.1)	46 (14.3)	36 (11.2)	<.001
Respiratory disease	25 (7.8)	46 (14.3)	67 (20.9)	<.001
Sepsis	20 (6.2)	33 (10.2)	43 (13.4)	<.001
Baseline physiology[c]				
Heart rate, median (IQR), beats/min	80 (66 to 100)	92 (80 to 106)	102 (88 to 122)	<.001
>80	160 (49.8)	228 (71.0)	277 (86.3)	<.001
Mean arterial pressure, median (IQR), mm Hg	73 (65 to 85)	82 (72 to 93)	82 (75 to 99)	<.001
<65	66 (20.6)	17 (5.3)	13 (4.1)	<.001
Respiratory rate, median (IQR), breaths/min	16 (14 to 20)	16 (14 to 18)	16 (14 to 20)	.02
>20	52 (16.2)	37 (11.5)	55 (17.2)	.09
Temperature, median (IQR), °C	36.0 (35.2 to 36.7)	36.3 (35.7 to 36.9)	36.5 (35.7 to 37.0)	<.001
<36	148 (46.7)	97 (30.4)	89 (28.4)	<.001
36-38	150 (47.3)	201 (63.0)	202 (64.5)	<.001
>38	17 (5.4)	18 (5.6)	21 (6.7)	.75
Baseline creatinine, median (IQR), mg/dL	1.14 (0.87 to 1.61)	1.04 (0.75 to 1.48)	1.07 (0.75 to 1.65)	.01
$Paco_2$, median (IQR), mm Hg[d]	44.0 (39.8 to 49.8)	43.00 (38.3 to 49.5)	41.4 (35.0 to 48.4)	<.001
Missing $Paco_2$ indicator	19 (5.9)	19 (5.9)	17 (5.3)	.93
Vasopressor or inotropic support	181 (56.2)	182 (56.5)	162 (50.5)	.22
Mode of mechanical ventilation				
Volume-based modes[e]	220 (68.3)	223 (69.3)	221 (68.8)	.97
Pressure-based modes[e]	54 (16.8)	52 (16.1)	84 (26.2)	.01
Pressure-regulated volume control	46 (14.3)	47 (14.6)	12 (3.7)	<.001
Missing or other ventilation mode indicator	2 (0.6)	0	4 (1.2)	.13
PEEP of mechanical ventilation, median (IQR), cm H_2O	6.5 (5.0 to 10.0)	6.6 (5.0 to 10.0)	5.0 (5.0 to 10.0)	.93
Predicted risk of 28-d mortality, median (IQR), %[f]	30.8 (25.7 to 43.1)	28.3 (25.7 to 30.8)	28.2 (25.7 to 30.8)	<.001
Randomized to the lower Spo_2 group	165 (51.2)	164 (50.9)	155 (48.3)	.71

Abbreviations: ICU-ROX, Intensive Care Unit Randomized Trial Comparing Two Approaches to Oxygen Therapy; $Paco_2$, arterial partial pressure of carbon dioxide; PEEP, positive end-expiratory pressure; PILOT, Pragmatic Investigation of Optimal Oxygen Targets; Spo_2, peripheral oxygenation-saturation by pulse oximetry.

SI conversion factor: To convert creatinine from mg/dL to µmol/L, multiply by 88.4.

[a] Testing for the difference *P* value were χ^2 for categorical variables and Kruskal-Wallis rank-sum test for continuous variables.

[b] The conditions that define the composite ICU diagnoses from PILOT and ICU-ROX variables are listed in the eMethods section of Supplement 1 and were created under guidance provided by the Australian and New Zealand Intensive Care Society Adult Patient Database Data Dictionary.[23]

[c] Baseline physiological measures are the last value recorded prior to randomization. Heart rate was missing in 2 patients (0.21%), respiratory rate in 4 (0.41%), mean arterial pressure in 5 (0.52%), temperature in 16 (1.7%), baseline creatinine in 26 (2.7%), and $Paco_2$ in 55 (5.7%).

[d] Normal range, 35 to 45 mm Hg.

[e] The modes of mechanical ventilation that define the composite of volume-based and pressure-based modes are eMethods in Supplement 1.

[f] Predicted risk of 28-day mortality was derived from the Acute Physiology and Chronic Health Evaluation (APACHE) II score in the ICU-ROX trial. Grouped increments of the score were correlated with the estimated 28-day mortality to create a predicted risk of 28-day mortality (eMethods, Supplement 1).

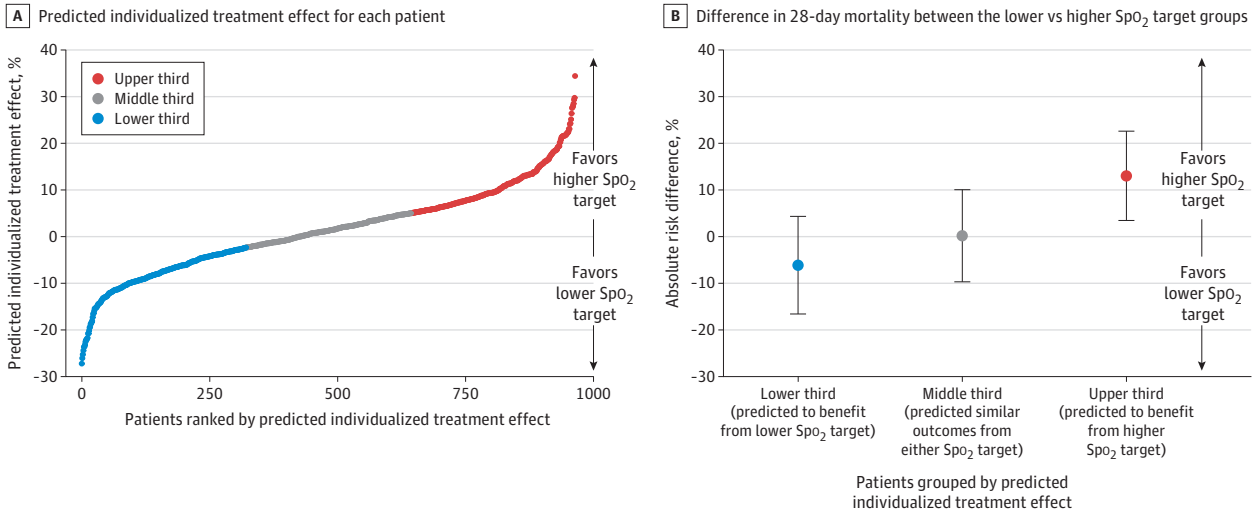
Table 3. Primary Outcome of 28-Day Mortality

Outcome	No. (%) of participants			*P* value
	Lower third: predicted to benefit from lower Spo_2 target (n = 322)	Middle third: predicted to have similar outcomes with either target (n = 322)	Upper third: predicted to benefit from higher Spo_2 target (n = 321)	
Overall	115 (35.7)	92 (28.6)	84 (26.2)	
Randomized to lower Spo_2 group	54 (32.7)	47 (28.7)	51 (32.9)	
Randomized to higher Spo_2 group	61 (38.9)	45 (28.5)	33 (19.9)	
Treatment effect (difference in incidence in 28-d mortality between lower vs higher Spo_2 groups), % (95% CI)	−6.1 (−16.5 to 4.3)	0.2 (−9.6 to 10.0)	13.0 (3.5 to 22.6)	.02[a]

Abbreviation: Spo_2, peripheral oxygenation-saturation by pulse oximetry.

[a] Likelihood ratio test *P* value for interaction term between the thirds of predicted individualized treatment effect and the randomized Spo_2 group assignment.

Figure 2. Absolute Risk Difference in 28-Day Mortality Between the Lower and Higher Spo_2 Target Groups by Predicted Individualized Treatment Effect



A, Displays the value of the predicted individualized treatment effect generated by the model derived from the Pragmatic Investigation of Optimal Oxygen Targets (PILOT) trial for each of the 965 patients in the Intensive Care Unit Randomized Trial Comparing Two Approaches to Oxygen Therapy (ICU-ROX) trial. Patients were ranked from the lowest value on the left to the highest value on the right.

B, For patients in the lower third, middle third, and upper third of predicted individualized treatment effect, the absolute risk difference in the incidence of 28-day mortality between patients randomized to the lower Spo_2 group and the higher Spo_2 group is shown (colored dots) with 95% CIs (whiskers). The predicted individualized treatment effect (lower, middle, or higher third) modified the effect of randomized Spo_2 group assignment on the incidence of 28-day mortality (*P* value for interaction = .02).

analysis that did not include patients' baseline risk of 28-day mortality as a predictor variable.

A strength of this study is that the model was derived in a large randomized trial in the United States and externally validated in a separate large randomized trial in Australia and New Zealand. This is important because individualized treatment effect models use supervised machine learning algorithms that can overfit the derivation cohort and produce excessively optimistic results. Prior studies examining predicted individualized treatment effects in a single randomized trial did not partition or partitioned the dataset using train-test or time-series splits.[16,17,33-35] The externally validation of models in a separate randomized trial, as done in this study, represents the next important step in accurately identifying individualized treatment effects.

Limitations

This study has many limitations. First, predictor variables and treatment groups were harmonized between the 2 randomized trials to perform model derivation and external validation. Differences between the trials in design (cluster crossover vs parallel group), baseline variables definitions, methods of data collection, extent of predictor variable missingness, and the Spo_2 ranges targeted could degrade the performance of the model. Improved harmonization of baseline variables and interventions across trials and collection of more detailed physiological and biological measures could improve the performance of individualized treatment-effect models. Second, the suggested benefit associated with lower Spo_2 targets in patients with acute nonhypoxic brain injury is at variance with findings of a prior subgroup analysis.[8] However, machine learning models may identify complex interactions among multiple predictor variables, which may add information beyond traditional 1-at-a-time subgroup analyses. Third, model performance decreased when deriving a model using the ICU-ROX trial and validating in the PILOT trial. This was likely due to the smaller sample size of the ICU-ROX trial compared with the PILOT trial. Fourth, further prospective validation is

required before results can be generalized to other patient populations and predicted individualized treatment effects are used to inform clinical care. The findings that patients with acute brain injury may experience lower mortality with a lower Spo_2 target and patients with sepsis may experience lower mortality with a higher Spo_2 target should not be interpreted as definitive treatment strategies.

Conclusion

Oxygenation targets that are individualized using machine learning analyses of randomized trials may reduce mortality for critically ill adults. A prospective trial evaluating the use of individualized oxygenation targets is needed.

ARTICLE INFORMATION

Accepted for Publication: February 28, 2024.

Published Online: March 19, 2024. doi:10.1001/jama.2024.2933

Author Affiliations: Division of Pulmonary and Critical Care, Department of Medicine, University of Chicago, Chicago, Illinois (Buell); Division of Pulmonary and Critical Care, Department of Medicine, University of Wisconsin-Madison, Madison (Spicer, Graham Linck, Churpek); Division of Allergy, Pulmonary, and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee (Casey, Seitz, Qian, Rice, Semler); Department of Emergency Medicine, Vanderbilt University Medical Center, Nashville, Tennessee (Self); Vanderbilt Institute for Clinical and Translational Research, Nashville, Tennessee (Self, Rice, Semler); Division of Clinical and Translational Research, Washington University School of Medicine, St Louis, Missouri (Sinha); Division of Critical Care, Department of Anesthesia, Washington University School of Medicine, St Louis, Missouri (Sinha); Intensive Care Unit, Wellington Hospital, Wellington, New Zealand (Young); Medical Research Institute of New Zealand, Wellington, New Zealand (Young); Australian and New Zealand Intensive Care Research Centre, Monash University, Melbourne, Victoria, Australia (Young); Department of Critical Care, University of Melbourne, Melbourne, Victoria, Australia (Young).

Author Contributions: Drs Buell, Young, Semler, and Churpek and Ms Spicer had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Dr Buell and Ms Spicer contributed equally to the work. Drs Young, Semler, and Churpek contributed equally to the work.
Concept and design: Buell, Casey, Seitz, Qian, Self, Rice, Sinha, Young, Semler, Churpek.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Buell, Spicer, Casey, Seitz, Graham Linck, Young, Semler, Churpek.
Critical review of the manuscript for important intellectual content: Buell, Casey, Seitz, Qian, Graham Linck, Self, Rice, Sinha, Young, Semler, Churpek.
Statistical analysis: Buell, Spicer, Seitz, Graham Linck, Semler, Churpek.
Obtained funding: Semler, Churpek.
Administrative, technical, or material support: Buell, Self.
Supervision: Casey, Self, Rice, Sinha, Churpek.

Conflict of Interest Disclosures: Ms Spicer reported being an employee of the University of Wisconsin Data Science Laboratory in the Department of Medicine, which is supported by grants from the Department of Defense and the National Institute of General Medical Sciences outside the submitted work. Dr Casey reported receiving travel support from Fisher & Paykel outside the submitted work. Dr Qian reported receiving speaker fees from Karl Storz Endoscopy outside the submitted work. Dr Graham Linck reported receiving a training grant from the National Institutes of Health during the conduct of the study. Dr Rice reported receiving personal fees for serving as director of medical affairs of Cumberland Pharmaceuticals Inc, as a member of the data and safety monitoring board of Sanofi, and as a consultant for Cytovale Inc outside the submitted work. Dr Sinha reported receiving personal fees from AstraZeneca and Prenosis Inc outside the submitted work. Dr Semler reported receiving personal fees for serving on the advisory board of Baxter International outside the submitted work. Dr Churpek reported receiving grants from the Department of Defense and the National Institute of General Medical Sciences outside the submitted work, having patent 11,410,777 to the University of Chicago for eCART (electronic cardiac arrest risk triage), an early warning score not used in this work, with royalties from the University of Chicago. No other disclosures were reported.

Funding/Support: This study was supported in part by grants T32HL007605 (Buell), K23HL153584 (Casey), T32HL087738 (Seitz and Qian), UL1RR024975 (Rice), K23HL143053 (Semler), and R01HL157262 (Churpek) from the National Heart, Lung, and Blood Institute; UL1TR002243 from the National Center for Advancing Translational Sciences (Self); and GM142992 from the National Institute of General Medical Sciences (Sinha). Research for this study was conducted during the tenure of a Clinical Research Practitioner Fellowship from the Health Research Council of New Zealand (Young). The Medical Research Institute of New Zealand receives Independent Research Organization funding from the Health Research Council of New Zealand.

Role of the Funder/Sponsor: The sponsors 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.

Meeting Presentation: This paper was presented at the 43rd International Symposium on Intensive Care and Emergency Medicine; Brussels, Belgium, March 19, 2024.

Data Sharing Statement: See Supplement 2.

REFERENCES

1. Schjørring OL, Klitgaard TL, Perner A, et al; HOT-ICU Investigators. Lower or higher oxygenation targets for acute hypoxemic respiratory failure. *N Engl J Med*. 2021;384(14):1301-1311. doi:10.1056/NEJMoa2032510

2. Semler MW, Casey JD, Lloyd BD, et al; PILOT Investigators and the Pragmatic Critical Care Research Group. Oxygen-saturation targets for critically ill adults receiving mechanical ventilation. *N Engl J Med*. 2022;387(19):1759-1769. doi:10.1056/NEJMoa2208415

3. Mackle D, Bellomo R, Bailey M, et al; ICU-ROX Investigators and the Australian and New Zealand Intensive Care Society Clinical Trials Group; ICU-ROX Investigators the Australian and New Zealand Intensive Care Society Clinical Trials Group. Conservative oxygen therapy during mechanical ventilation in the ICU. *N Engl J Med*. 2020;382(11):989-998. doi:10.1056/NEJMoa1903297

4. Girardis M, Busani S, Damiani E, et al. Effect of conservative vs conventional oxygen therapy on mortality among patients in an intensive care unit: the oxygen-ICU randomized clinical trial. *JAMA*. 2016;316(15):1583-1589. doi:10.1001/jama.2016.11993

5. Schmidt H, Kjaergaard J, Hassager C, et al. Oxygen targets in comatose survivors of cardiac arrest. *N Engl J Med*. 2022;387(16):1467-1476. doi:10.1056/NEJMoa2208686

6. Gelissen H, de Grooth HJ, Smulders Y, et al. Effect of low-normal vs high-normal oxygenation targets on organ dysfunction in critically ill patients: a randomized clinical trial. *JAMA*. 2021;326(10):940-948. doi:10.1001/jama.2021.13011

7. van der Wal LI, Grim CCA, Del Prado MR, et al; ICONIC investigators. Conservative versus Liberal Oxygenation Targets in Intensive Care Unit Patients (ICONIC): a randomized clinical trial. *Am J Respir Crit Care Med*. 2023;208(7):770-779. doi:10.1164/rccm.202303-0560OC

8. Young PJ, Mackle D, Hodgson C, et al; ICU-ROX Investigators and the Australian and New Zealand Intensive Care Society Clinical Trials Group. Conservative or liberal oxygen therapy for mechanically ventilated adults with acute brain pathologies: a post-hoc subgroup analysis. *J Crit Care*. 2022;71:154079. doi:10.1016/j.jcrc.2022.154079

9. Young P, Mackle D, Bellomo R, et al; ICU-ROX Investigators and the Australian and New Zealand Intensive Care Society Clinical Trials Group. Conservative oxygen therapy for mechanically ventilated adults with suspected hypoxic ischaemic encephalopathy. *Intensive Care Med*. 2020;46(12):2411-2422. doi:10.1007/s00134-020-06196-y

10. Young P, Mackle D, Bellomo R, et al; ICU-ROX Investigators the Australian New Zealand Intensive Care Society Clinical Trials Group. Conservative oxygen therapy for mechanically ventilated adults with sepsis: a post hoc analysis of data from the Intensive Care Unit Randomized Trial Comparing Two Approaches to Oxygen Therapy (ICU-ROX). *Intensive Care Med*. 2020;46(1):17-26. doi:10.1007/s00134-019-05857-x

11. Demiselle J, Calzia E, Hartmann C, et al. Target arterial Po_2 according to the underlying pathology: a mini-review of the available data in mechanically ventilated patients. *Ann Intensive Care*. 2021;11(1):88. doi:10.1186/s13613-021-00872-y

12. Helmerhorst HJ, Roos-Blom MJ, van Westerloo DJ, de Jonge E. Association between arterial hyperoxia and outcome in subsets of critical illness: a systematic review, meta-analysis, and meta-regression of cohort studies. *Crit Care Med*. 2015;43(7):1508-1519. doi:10.1097/CCM.0000000000000998

13. Klitgaard TL, Schjørring OL, Lange T, et al. Lower versus higher oxygenation targets in critically ill patients with severe hypoxaemia: secondary Bayesian analysis to explore heterogeneous treatment effects in the Handling Oxygenation Targets in the Intensive Care Unit (HOT-ICU) trial. *Br J Anaesth*. 2022;128(1):55-64.

14. Iwashyna TJ, Burke JF, Sussman JB, Prescott HC, Hayward RA, Angus DC. Implications of heterogeneity of treatment effect for reporting and analysis of randomized trials in critical care. *Am J Respir Crit Care Med*. 2015;192(9):1045-1051. doi:10.1164/rccm.201411-2125CP

15. Kent DM, Paulus JK, van Klaveren D, et al. The Predictive Approaches to Treatment Effect Heterogeneity (PATH) statement. *Ann Intern Med*. 2020;172(1):35-45. doi:10.7326/M18-3667

16. Seitz KP, Spicer AB, Casey JD, et al. Individualized treatment effects of bougie vs stylet for tracheal intubation in critical illness. *Am J Respir Crit Care Med*. 2023;207(12):1602-1611.

17. Goligher EC, Lawler PR, Jensen TP, et al; REMAP-CAP, ATTACC, and ACTIV-4a Investigators. Heterogeneous treatment effects of therapeutic-dose heparin in patients hospitalized for COVID-19. *JAMA*. 2023;329(13):1066-1077. doi:10.1001/jama.2023.3651

18. Hoogland J, IntHout J, Belias M, et al. A tutorial on individualized treatment effect prediction from randomized trials with a binary endpoint. *Stat Med*. 2021;40(26):5961-5981. doi:10.1002/sim.9154

19. Angus DC, Chang CH. Heterogeneity of treatment effect: estimating how the effects of interventions vary across individuals. *JAMA*. 2021;326(22):2312-2313. doi:10.1001/jama.2021.20552

20. 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;22(7):707-710. doi:10.1007/BF01709751

21. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13(10):818-829. doi:10.1097/00003246-198510000-00009

22. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ*. 2018;363:k4245. doi:10.1136/bmj.k4245

23. Australian and New Zealand Intensive Care Society. Evaluation ACfOaR. APD Data Dictionary. Updated April 2022. Accessed June 13th 2023. https://www.anzics.com.au/adult-patient-database-apd/

24. Hastie T, Tibshirani R, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Vol 2. Springer; 2009. doi:10.1007/978-0-387-84858-7

25. Robinson PM. Root-*N*-consistent semiparametric regression. *Econometrica*. 1988;56(4):931-954. doi:10.2307/1912705

26. Robins JM. *Optimal Structural Nested Models for Optimal Sequential Decisions*. Springer; 2004:189-326. doi:10.1007/978-1-4419-9076-1_11

27. Sinha P, Spicer A, Delucchi KL, McAuley DF, Calfee CS, Churpek MM. Comparison of machine learning clustering algorithms for detecting heterogeneity of treatment effect in acute respiratory distress syndrome: a secondary analysis of three randomised controlled trials. *EBioMedicine*. 2021;74:103697. doi:10.1016/j.ebiom.2021.103697

28. Radcliffe N. *Using Control Groups to Target on Predicted Lift: Building and Assessing Uplift Model*. Direct Marketing Analytics Journal; 2007:14-21.

29. van Klaveren D, Steyerberg EW, Serruys PW, Kent DM. The proposed "concordance-statistic for benefit" provided a useful metric when modeling heterogeneous treatment effects. *J Clin Epidemiol*. 2018;94:59-68. doi:10.1016/j.jclinepi.2017.10.021

30. Belbahri M, Murua A, Gandouet O, Partovi Nia V. Qini-based uplift regression. *Ann Appl Stat*. 2021;15(3):1247-1272. doi:10.1214/21-AOAS1465

31. Young PJ, Bailey M, Bellomo R, et al. Conservative or liberal oxygen therapy in adults after cardiac arrest: an individual-level patient data meta-analysis of randomised controlled trials. *Resuscitation*. 2020;157:15-22. doi:10.1016/j.resuscitation.2020.09.036

32. Bhavani SV, Semler M, Qian ET, et al. Development and validation of novel sepsis subphenotypes using trajectories of vital signs. *Intensive Care Med*. 2022;48(11):1582-1592. doi:10.1007/s00134-022-06890-z

33. Blette BS, Granholm A, Li F, et al. Causal Bayesian machine learning to assess treatment effect heterogeneity by dexamethasone dose for patients with COVID-19 and severe hypoxemia. *Sci Rep*. 2023;13(1):6570. doi:10.1038/s41598-023-33425-3

34. Sadique Z, Grieve R, Diaz-Ordaz K, Mouncey P, Lamontagne F, O'Neill S. A machine-learning approach for estimating subgroup- and individual-level treatment effects: an illustration using the 65 Trial. *Med Decis Making*. 2022;42(7):923-936. doi:10.1177/0272989X221100717

35. Zampieri FG, Damiani LP, Bagshaw SM, et al; BRICNet. Conditional treatment effect analysis of two infusion rates for fluid challenges in critically ill patients: a secondary analysis of Balanced Solution versus Saline in Intensive Care Study (BaSICS) Trial. *Ann Am Thorac Soc*. 2023;20(6):872-879. doi:10.1513/AnnalsATS.202211-946OC