Predicting ICU Mortality in Acute Respiratory Distress Syndrome Patients Using Machine Learning: The Predicting Outcome and STratifiCation of severity in ARDS (POSTCARDS) Study*

OBJECTIVES: To assess the value of machine learning approaches in the development of a multivariable model for early prediction of ICU death in patients with acute respiratory distress syndrome (ARDS).

DESIGN: A development, testing, and external validation study using clinical data from four prospective, multicenter, observational cohorts.

SETTING: A network of multidisciplinary ICUs.

PATIENTS: A total of 1,303 patients with moderate-to-severe ARDS managed with lung-protective ventilation.

INTERVENTIONS: None.

MEASUREMENTS AND MAIN RESULTS: We developed and tested prediction models in 1,000 ARDS patients. We performed logistic regression analysis following variable selection by a genetic algorithm, random forest and extreme gradient boosting machine learning techniques. Potential predictors included demographics, comorbidities, ventilatory and oxygenation descriptors, and extrapulmonary organ failures. Risk modeling identified some major prognostic factors for ICU mortality, including age, cancer, immunosuppression, Pao_2/Fio_2, inspiratory plateau pressure, and number of extrapulmonary organ failures. Together, these characteristics contained most of the prognostic information in the first 24 hours to predict ICU mortality. Performance with machine learning methods was similar to logistic regression (area under the receiver operating characteristic curve [AUC], 0.87; 95% CI, 0.82–0.91). External validation in an independent cohort of 303 ARDS patients confirmed that the performance of the model was similar to a logistic regression model (AUC, 0.91; 95% CI, 0.87–0.94).

CONCLUSIONS: Both machine learning and traditional methods lead to promising models to predict ICU death in moderate/severe ARDS patients. More research is needed to identify markers for severity beyond clinical determinants, such as demographics, comorbidities, lung mechanics, oxygenation, and extrapulmonary organ failure to guide patient management.

KEY WORDS: acute respiratory distress syndrome; clinical trials; ICU mortality; lung-protective ventilation; machine learning; observational studies; stratification.

Jesús Villar, MD, PhD, FCCM[1,2,3]
Jesús M. González-Martín, PhD[1,2]
Jerónimo Hernández-González, PhD[4]
Miguel A. Armengol, PhD[5]
Cristina Fernández, MSc[2]
Carmen Martín-Rodríguez, MD[6]
Fernando Mosteiro, MD, PhD[7]
Domingo Martínez, MD[8]
Jesús Sánchez-Ballesteros, MD[9]
Carlos Ferrando, MD, PhD[10]
Ana M. Domínguez-Berrot, MD[11]
José M. Añón, MD, PhD[12]
Laura Parra, MD[13]
Raquel Montiel, MD[14]
Rosario Solano, MD[15]
Denis Robaglia, MD[16]
Pedro Rodríguez-Suárez, MD, PhD[1,17]
Estrella Gómez-Bentolila, RN, MSc[2]
Rosa L. Fernández, MSc[1,2]
Tamas Szakmany, MD, PhD, FCCM[18,19]
Ewout W. Steyerberg, PhD[20]
Arthur S. Slutsky, MD[3,21]
for the Predicting Outcome and STratifiCation of severity in ARDS (POSTCARDS) Network

***See also p. 1814.**

DOI: 10.1097/CCM.0000000000006030

The acute respiratory distress syndrome (ARDS) is a severe form of acute hypoxemic respiratory failure associated with high morbidity and mortality (1). The ability to accurately predict mortality of ARDS patients' remains challenging despite an array of existing prediction models which combine multiple variables thought to influence prognosis (2–4). A reliable

KEY POINTS

Purpose: To develop and validate an early prediction model for ICU death in patients with moderate-to-severe acute respiratory distress syndrome (ARDS) using machine learning.

Findings: Both machine learning and traditional methods were effective in predicting patients likely to die in the ICU.

Meanings: Major prognostic factors for ICU mortality in patients with moderate/severe ARDS patients include plateau pressure, oxygenation, age, number of extrapulmonary organ failures, cancer, and immunosuppression. We identified a seven-variable prediction model that might improve prediction beyond models currently available to help guide therapeutic choices.

prediction tool for assessment of outcome in the ICU may prove beneficial for decision-making in these patients.

Despite advances in the management of ARDS patients over the past 2 decades (5–9) reported mortality rates (at ICU and hospital discharge) in observational studies are between 35% and 45%. Predicting ARDS outcome could inform clinicians' decision-making by targeting specific therapeutic interventions to enhance organ recovery, reduce iatrogenic harm, and increase survival. ARDS outcome is influenced by a broad spectrum of clinical factors dependent and independent of pulmonary function (1). Identifying key clinical variables, which are associated with mortality in ARDS, might suggest therapeutic alternatives to lower the high fatality rate of ARDS.

Since its initial clinical description, the management of ARDS has understandably evolved (2, 10). Few studies have investigated the prediction of ICU mortality in ARDS patients in the era of lung-protective mechanical ventilation (MV) using the current ARDS definition (11). Recently developed prediction models interpretable at the bedside combine clinically relevant variables derived from a population that represents the type of patients seen in clinical practice. Machine learning (ML) approaches hold promise to capture the complex interactions among these variables (12, 13) and, in recent years, have been used for predicting mortality in critically ill patients, including those with ARDS (14–16).

Using a large number of patients with moderate-to-severe ARDS admitted to a network of ICUs from several geographical areas of Spain, we aimed to assess the value of ML techniques to predict ICU mortality using variables collected within the first 24 hours of diagnosis of moderate-to-severe ARDS. We compared the prognostic abilities of ML and logistic regression models with each other, and with our previously described Stratification for identification of Prognostic categories In acute RESpiratory distress syndrome (SPIRES) score (4).

METHODS

This study was approved by the Ethics Committee for Clinical Research at Hospital Universitario Dr. Negrín (Las Palmas de Gran Canaria, Spain). The need for informed consent was waived based on Spanish legislation for biomedical research (CEI/CEIm No. 2021-321-1) under the Royal Decree 1090/2015 December 2015 and Royal Decree 957/2020 November 2020, due to the retrospective nature of this secondary analysis, the anonymization/dissociation of data, and no potential harm or benefit to patients (**Supplemental File**, http://links.lww.com/CCM/H413). The study followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis guidelines for prediction models (17).

Patient Population

This study is an extension of the Spanish Initiative for Epidemiology, Stratification and Therapies of Acute Respiratory Distress Syndrome program (3, 18–20) (Supplemental File, http://links.lww.com/CCM/H413). We performed a comprehensive analysis, termed the Predicting Outcome and STratifiCation of severity in ARDS (POSTCARDS) study, of an unrestricted set of data derived from 1,303 adult (≥ 18 yr old) patients with moderate-to-severe ARDS (11) treated with lung-protective MV in a network of ICUs from several geographical areas of Spain (Supplemental File, http://links.lww.com/CCM/H413).

The study was conducted in four steps. In the first two steps (model development and testing), we analyzed data derived from 1,000 patients included in three independent, prospective, multicenter, observational

cohorts enrolling consecutive patients meeting current criteria for moderate-to-severe ARDS (11). In the third step, we tested the performance of the model in an independent cohort of 303 patients with moderate/severe ARDS included in the multicenter observational Prevalence and outcome of acute hypoxemic respiratory failure (PANDORA) study (20) with sufficient number of events (ICU deaths) required for reliable external validation (21, 22) (Supplemental File, http://links.lww.com/CCM/H413). Finally, we compared the predictive performance of the different models with the previously reported SPIRES score (4).

Variables, Primary Outcome, and Predefined Rules

To build the models, we used variables, including demographics, comorbidities, cause of ARDS, Acute Physiology and Chronic Health Evaluation II score (23) during the first 24 hours of ARDS diagnosis, data from ventilator settings and lung mechanics (tidal volume [Vt], respiratory rate [RR], positive end-expiratory pressure [PEEP], plateau pressure [Pplat]), and gas exchange ([Pao_2, $Paco_2$, Fio_2, Pao_2/Fio_2, pH]) at the time of diagnosis of moderate/severe ARDS and 24 hours later. For the purpose of this study, values of Pao_2/Fio_2 and Pplat at 24 hours were assessed under standardized ventilatory settings (PEEP = 10 cm H_2O and Fio_2 = 0.5) (24). When patients required PEEP greater than 10 or Fio_2 greater than 0.5 and could not tolerate a decrease in PEEP or Fio_2, a set of rules for setting PEEP and Fio_2 were applied only during the standardized assessment, as described and validated previously by our group (3, 24). At other times, PEEP and Fio_2 levels were set at the discretion of managing clinicians. We recorded the Sequential Organ Failure Assessment (SOFA) score (25) and occurrence of extrapulmonary organ failures (OFs) included in the SOFA scale at diagnosis of moderate-to-severe ARDS and 24 hours later. We recorded date and status (alive or dead) of patients at ICU and hospital discharge. Primary outcome was all-cause ICU mortality.

Although we collected data from 165 variables in each patient during ICU stay, feature selection is of vital importance in building a prediction model that is easily actionable and interpretable in clinical daily practice. Based on previous work by our group (26, 27), we focused our analysis on variables collected within the first 24 hours of diagnosis of moderate/severe ARDS to estimate the early probability of ICU death, independent of the underlying disease or cause of death (**Fig. S1**, http://links.lww.com/CCM/H413). Our aim for variable selection was to incorporate clinically relevant variables while avoiding noise/redundant variables. We analyzed the following variables as potential predictors of ICU outcome: age, gender, comorbidities, SOFA score, number of extrapulmonary OF, Pao_2, Pao_2/Fio_2, $Paco_2$, pH, Fio_2, Vt, RR, PEEP, Pplat, driving pressure (calculated as Pplat minus PEEP), and minute ventilation, at the time of moderate/severe ARDS diagnosis and 24 hours later (**Tables S1** and **S2**, http://links.lww.com/CCM/H413). We prespecified the analysis before final statistical analyses were conducted (Supplemental File, http://links.lww.com/CCM/H413).

Statistical Analysis

We performed descriptive statistical analyses to analyze patients until ICU discharge. We performed univariable analysis to predict ICU outcome. We identified potential variables that could be included in the prediction model based on our predefined rules, their contribution to the area under the receiver operating characteristic curve (AUC), and their *p* values. Since the inclusion of all available variables in ML practice can lead to complex models that are difficult to interpret, we screened the collected variables employing a genetic algorithm variable selection method (28) to achieve parsimony and to identify a subset of relevant variables (subset selection) for an accurate prediction model, while excluding noise/redundant variables. We decided to use genetic algorithm for variable selection since it demonstrated better performance than other advanced variable selection methods in large ICU datasets and due to our previous successful experience (29). We applied the genetic algorithm to optimize the subset of selected variables by minimizing the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) (30). We calculated the variance inflation factor, a measure of multicollinearity in regression logistic analysis. Multicollinearity exists when there is a correlation between multiple independent variables in a multiple regression model. A two-sided *p* value of less than 0.005 was considered for identification of important prognostic variables, to keep the false discovery rate below 5% (31).

We built the POSTCARDS prediction model by considering the minimum number of variables obtained by a genetic algorithm that provided a similar performance as all variables prediction model. We used a five-fold cross-validation for splitting randomly the 1,000-patient dataset as 800 patients for training and 200 patients for validating and repeating this process 100 times (Supplemental File, http://links.lww.com/CCM/H413). We evaluated the final minimum number of variables model using two ML approaches: random forest (RF) and extreme gradient boosting (XGBoost) (32, 33) (Supplemental File, http://links.lww.com/CCM/H413). Calculations were performed using the R Core Team 2022 software (R Version 4.2.2) (https://www.r-project.org) (R Foundation for Statistical Computing, Vienna, Austria), package "randomForest" and "xgboost" (34, 35).

We also examined whether the ML models provided an improvement in the prediction of ICU mortality when compared with the SPIRES score that we previously reported (4). We calculated measures to assess the validation of the prediction models, related to calibration and discrimination, by studying the external validity of the models developed in 1,000 patients and tested in 303 patients (36, 37) (detailed in Supplemental File, http://links.lww.com/CCM/H413).

Figure 1 summarizes the study design.

RESULTS

From the 1,000 patients used for model development, 375 patients died (37.5%) in ICU (**Table 1**). We analyzed 34 clinically relevant variables collected within

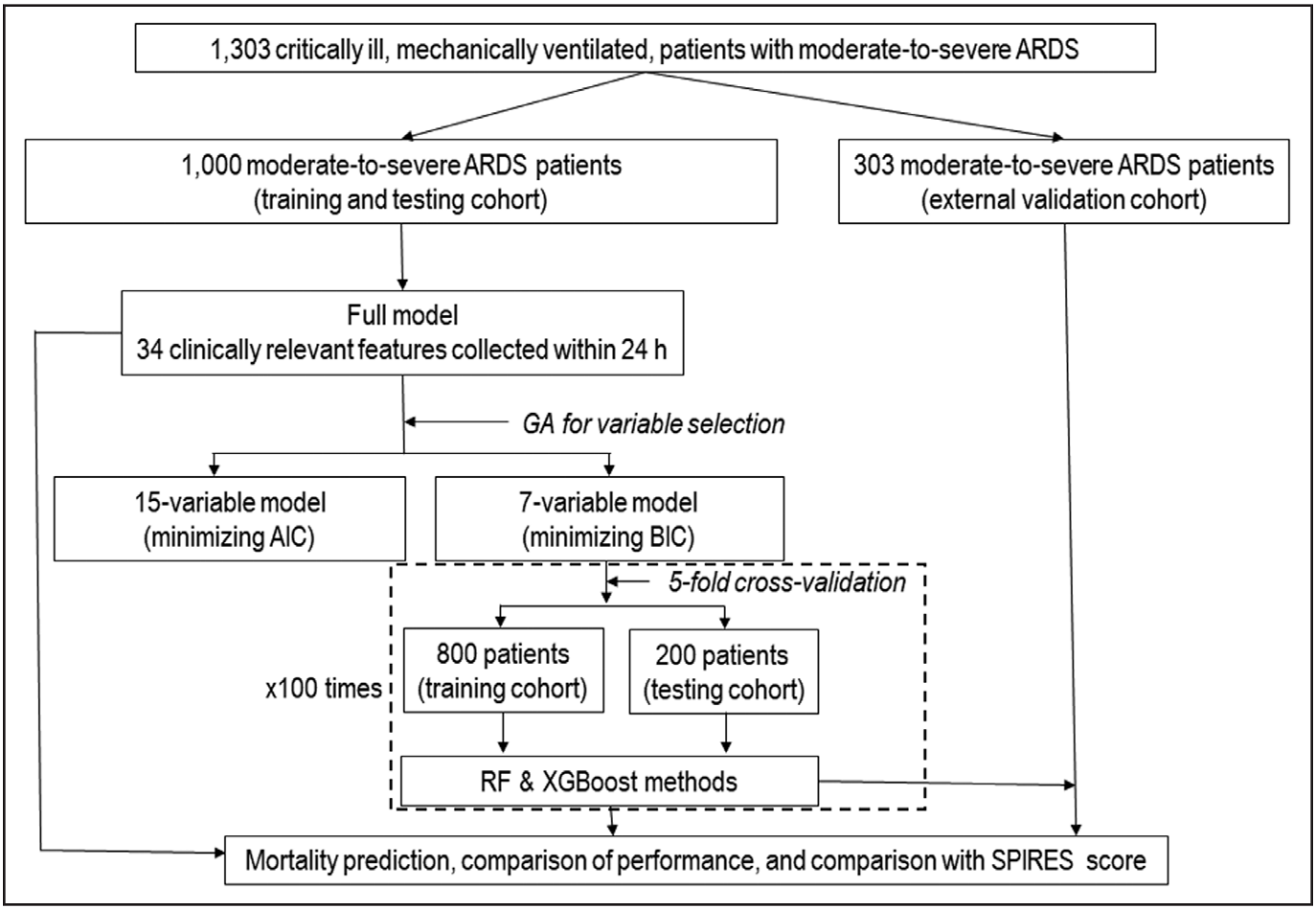

Figure 1. Diagram representing the study design. The flowchart illustrates the scheme for the database with 1,303 patients with moderate/severe acute respiratory distress syndrome (ARDS), selection of variables for final analysis, machine learning approaches, and comparisons among the prediction models. Once most relevant variables were selected by a genetic algorithm (GA) in the dataset of 1,000 patients, this dataset was divided into five folders to perform five-fold randomized cross-validation repeated 100 times using machine learning. AIC = Akaike information criterion, BIC = Bayesian information criterion, RF = random forest, SPIRES = a four-variable score as an acronym for "Stratification for Prognostic categories In the Acute RESpiratory distress syndrome" (see Villar et al [4]), XGBoost = extreme gradient boosting.

TABLE 1.
Baseline Characteristics and Outcome Data of 1,303 Patients With Moderate-to-Severe Acute Respiratory Distress Syndrome

Variables	Development Cohort ($n = 1,000$)	Testing Cohort ($n = 303$)	p
Age, yr, median (IQR)	57 (46–70)	60 (49–70)	0.157
Gender, *n* (%)			0.075
Male	680 (68.0)	223 (73.6)	
Female	320 (32.0)	80 (26.4)	
Etiology, *n* (%)			
Pneumonia	480 (48.0)	110 (36.3)	
Sepsis	286 (28.6)	78 (25.7)	
Aspiration	94 (9.4)	47 (15.5)	
Trauma	74 (7.4)	38 (12.5)	
Acute pancreatitis	32 (3.2)	13 (4.3)	
Multiple transfusions	10 (1.0)	3 (1.0)	
Others	24 (2.4)	14 (4.6)	
Degree of ARDS severity, *n* (%)			0.088
Severe	410 (41.0)	107 (35.3)	
Moderate	590 (59.0)	196 (64.7)	
Acute Physiology and Chronic Health Evaluation II score, mean ± SD	20.8±6.7	21.3±7.8	0.382
Sequential Organ Failure Assessment score, mean ± SD	9.25±3.5	9.9±3.6	0.005
FIO_2, mean ± SD	0.79±0.19	0.765±0.20	0.050
PaO_2, mm Hg, mean ± SD	85.9±26.3	86.3±24.9	0.817
PaO_2/FIO_2, mm Hg, mean ± SD	114.8±38.3	120.4±41.0	0.029
$PaCO_2$, mm Hg, mean ± SD	49.0±12.5	50.6±13.8	0.072
pH, mean ± SD	7.30±0.11	7.29±0.11	0.112
Tidal volume , mL/kg predicted body weight, mean ± SD	6.9±1.1	6.7±1.1	0.006
Respiratory rate, ventilator cycles/min, mean ± SD	21.4±4.9	22.3±4.6	0.005
Minute ventilation, L/min, mean ± SD	9.1±2.2	9.5±2.0	0.005
Positive end-expiratory pressure, cm H_2O, mean ± SD	12.1±3.3	11.0±3.0	< 0.001
Plateau pressure, cm H_2O, mean ± SD	26.5±4.8	25.2±4.9	< 0.001
Driving pressure, cm H_2O, mean ± SD	14.5±4.8	14.3±4.8	0.624
Number of extrapulmonary organ failure, mean ± SD	1.7±1.1	1.9±1.1	0.006
Length of ICU stay, d, median (IQR)	18 (11–31)	16 (8–26)	0.006
Days from ICU admission to ARDS onset/diagnosis, median (IQR)	1 (0–3)	1 (0–2)	0.588
Days from last day mechanical ventilation to ICU discharge, median (IQR)	2 (0–5)	2 (0–6)	0.752
All-cause ICU mortality, *n* (%)	375 (37.5)	112 (37.0)	0.920
All-cause hospital mortality, *n* (%)	415 (41.5)	124 (40.9)	0.860

ARDS = acute respiratory distress syndrome, IQR = interquartile range.

TABLE 2.

Univariate Logistic Regression of Clinically Relevant Variables in 1,000 Patients With Moderate-to-Severe Acute Respiratory Distress Syndrome

Variables	*n*	OR (95% CI)	*p*	Area Under the Receiving Operating Characteristic Curve (95% CI)
Age	1,000	1.03 (1.02–1.04)	1.1×10^{-13}	0.65 (0.61–0.68)
Gender	1,000	0.96 (0.73–1.3)	0.78	0.50 (0.47–0.53)
Cardiac disease	1,000	2.21 (1.37–3.61)	0.0013	0.53 (0.51–0.55)
Diabetes	1,000	1.23 (0.85–1.75)	0.26	0.51 (0.49–0.54)
Immunosuppressed	1,000	2.55 (1.68–3.92)	1.5×10^{-05}	0.54 (0.52–0.56)
Morbid obesity	1,000	0.80 (0.50–1.28)	0.36	0.51 (0.49–0.53)
Liver disease	1,000	2.47 (1.41–4.38)	0.0016	0.52 (0.51–0.54)
Neoplastic disease	1,000	3.34 (2.43–4.62)	1.8×10^{-13}	0.60 (0.57–0.63)
SOFA at T0	1,000	1.23 (1.18–1.29)	3.4×10^{-22}	0.69 (0.65–0.72)
SOFA at T24	1,000	1.27 (1.22–1.32)	8.8×10^{-29}	0.72 (0.69–0.75)
V_T at T0	1,000	1.03 (0.91–1.16)	0.68	0.50 (0.46–0.54)
V_T at T24	1,000	0.88 (0.76–1.01)	0.07	0.54 (0.50–0.57)
F_{IO_2} at T0	1,000	1.52 (0.78–3.0)	0.22	0.52 (0.49–0.56)
F_{IO_2} at T24	1,000	25.6 (12.3–54.11)	7.9×10^{-18}	0.66 (0.62–0.69)
Respiratory rate at T0	1,000	1.01 (0.99–1.04)	0.30	0.51 (0.47–0.54)
Respiratory rate at T24	1,000	1.05 (1.03–1.08)	9.2×10^{-05}	0.57 (0.54–0.61)
PEEP at T0	1,000	0.97 (0.94–1.01)	0.17	0.52 (0.49–0.56)
PEEP at T24	1,000	1.05 (1.0–1.10)	0.03	0.55 (0.52–0.59)
Plateau pressure at T0	1,000	1.06 (1.03–1.09)	5.4×10^{-05}	0.57 (0.53–0.61)
Plateau pressure at T24	1,000	1.22 (1.18–1.26)	2.6×10^{-28}	0.73 (0.70–0.77)
Driving pressure at T0	1,000	1.06 (1.03–1.09)	2.5×10^{-05}	0.58 (0.54–0.61)
Driving pressure at T24	1,000	1.18 (1.15–1.23)	1.7×10^{-23}	0.70 (0.67–0.73)
Minute ventilation at T0	1,000	1.01 (0.95–1.07)	0.83	0.49 (0.46–0.53)
Minute ventilation at T24	1,000	1.05 (0.99–1.11)	0.10	0.53 (0.49–0.56)
Pa_{O_2} at T0	1,000	0.99 (0.99–1.0)	0.005	0.56 (0.52–0.59)
Pa_{O_2} at T24	1,000	0.98 (0.98–0.99)	3.7×10^{-10}	0.64 (0.60–0.67)
Pa_{O_2}/F_{IO_2} at T0	1,000	1.0 (0.99–1.0)	0.006	0.55 (0.52–0.59)
Pa_{O_2}/F_{IO_2} at T24	1,000	0.99 (0.98–0.99)	2.9×10^{-20}	0.69 (0.66–0.73)
Pa_{CO_2} at T0	1,000	1.01 (1.0–1.02)	0.012	0.54 (0.51–0.58)
Pa_{CO_2} at T24	1,000	1.03 (1.02–1.05)	2.4×10^{-07}	0.59 (0.55–0.63)
pH at T0	1,000	0.10 (0.03–0.32)	0.00011	0.56 (0.53–0.60)
pH at T24	1,000	0 (0–0.01)	3.2×10^{-16}	0.64 (0.61–0.68)
Number of extrapulmonary OF at T0	1,000	2.1 (1.84–2.41)	2.0×10^{-26}	0.70 (0.67–0.73)
Number of extrapulmonary OF at T24	1,000	2.31 (2.03–2.66)	9.4×10^{-34}	0.73 (0.70–0.76)

OF = organ failure, OR = odds ratio, PEEP = positive end-expiratory pressure, SOFA = Sequential Organ Failure Assessment, T0 = at the time of moderate/severe acute respiratory distress syndrome diagnosis, T24 = at 24 hr after moderate/severe acute respiratory distress syndrome diagnosis, V_T = tidal volume.

the first 24 hours of moderate/severe ARDS diagnosis: 20 variables had a significant univariate prognostic relation with ICU death and 12 variables had an AUC greater than or equal to 0.6 (**Table 2**). These variables included some factors that could be influenced by clinical interventions such as number of extrapulmonary OF, SOFA score, oxygenation, and Pplat, all of them measured at 24 hours.

In the multivariable logistic regression analysis, few characteristics associated with mortality in the univariate analysis remained statistically significant (Table S2, http://links.lww.com/CCM/H413). The performance of the model with all 34 variables (full model, including all variables, **Fig. S2**, http://links.lww.com/CCM/H413) had a cross-validated AUC of 0.88 (95% CI, 0.86–0.90) (Table S2, http://links.lww.com/CCM/H413), but most variables were correlated (**Table S3**, http://links.lww.com/CCM/H413). When applying the genetic algorithm for variable selection using optimization of AIC, the resulting model included 15 variables (15-variable model, **Table S4** and Fig. S2, http://links.lww.com/CCM/H413). The performance of 15-variable model had an AUC of 0.88 (95% CI, 0.86–0.90) (**Table S5**, http://links.lww.com/CCM/H413). When the genetic algorithm for subset selection optimized BIC, the resulting model reduced the number of predictors from 34 to 7 variables (seven-variable model, **Table 3**; and Fig. S2, http://links.lww.com/CCM/H413), with an AUC of 0.87 (95% CI, 0.85–0.90) (**Table S6**, http://links.lww.com/CCM/H413). The seven variables showing strong relations with ICU mortality were: age, Pplat at 24 hours, Pao_2/Fio_2 at 24 hours, number of extrapulmonary OF at 24 hours, history of neoplastic disease, immunosuppression, and Pplat at baseline (Table S6, http://links.lww.com/CCM/H413). Of note, the first four variables of that model characterized the SPIRES score (4) with an AUC of 0.86 (95% CI, 0.84–0.89) (**Table S7**, http://links.lww.com/CCM/H413). When comparing 15-variable and seven-variable models with the SPIRES scoring model, there were no significant differences in their performance (15-variable model vs SPIRES; $p = 0.060$ and seven-variable model vs SPIRES; $p = 0.30$) (**Table S8** and **Fig. S3**, http://links.lww.com/CCM/H413). The four variables of the SPIRES score had the higher importance in the seven-variable model (**Fig. S4**, http://links.lww.com/CCM/H413).

Validated performance using XGBoost and RF was similar to performance by logistic regression (AUC, 0.86; 95% CI, 0.81–90 vs AUC, 0.87; 95% CI, 0.82–0.91 vs AUC, 0.87; 95% CI, 0.82–0.91, respectively) (**Table S9**, http://links.lww.com/CCM/H413). The external

TABLE 3.

Performance of a Logistic Regression Model of Predicting ICU Mortality (Seven-Variable Model) Within 24 Hours of Diagnosis of Moderate-to-Severe Acute Respiratory Distress Syndrome Using Logistic Regression Analysis and Minimizing the Bayesian Information Criterion

Variable	b	SE	OR (95% CI)	p
Intercept	−8.25	0.88	0 (0–0.001)	< 0.001
Age	0.05	0.01	1.048 (1.036–1.060)	< 0.001
Immunosuppressed	1.09	0.29	2.986 (1.685–5.341)	< 0.001
Neoplastic disease	1.13	0.21	3.093 (2.059–4.682)	< 0.001
Plateau pressure at the time of diagnosis/onset of moderate/severe acute respiratory distress syndrome	−0.08	0.02	0.920 (0.879–0.962)	< 0.001
Plateau pressure at T24	0.25	0.03	1.278 (1.212–1.351)	< 0.001
Pao_2/Fio_2 at T24	−0.01	0	0.991 (0.988–0.994)	< 0.001
Number of extrapulmonary OF at T24	0.82	0.08	2.276 (1.938–2.694)	< 0.001
Akaike information criterion			869.47	
Bayesian information criterion			908.73	
Area under the receiving operating characteristic curve			0.87 (95% CI: 0.85–0.89)	

OR = odds ratio, T24 = at 24 hr of diagnosis of moderate/severe acute respiratory distress syndrome.
This model reduced the number of relevant variables from 34 to 7. Data are expressed as mean values of logistic coefficients.

validation cohort (n = 303) had baseline characteristics and ICU mortality (112 deaths, 37%) similar to 1,000 patients for model development (Table 1). The POSTCARDS models as optimized by the two ML techniques provided similar performance to the one derived by logistic regression (AUC, 0.91; 95% CI, 0.87–0.94) (**Table S10**, http://links.lww.com/CCM/H413). Calibration results suggest good reliability of predictions, with logistic regression best overall (**Fig. S5**, http://links.lww.com/CCM/H413).

DISCUSSION

We found that prediction models of ICU mortality developed by ML methods provided similar discriminative ability and calibration to regression approaches and to the parsimonious SPIRES score, when applied to datasets of moderate-to-severe ARDS patients. Seven characteristics (age, history of cancer, history of immunosuppression, baseline Pplat, and Pplat at 24 hr, Pao_2/Fio_2 at 24 hr, and number of extrapulmonary OF at 24 hr) contained most of the prognostic information for ICU death within the first 24 hours after diagnosis of moderate/severe ARDS. The validity of the POSTCARDS model was confirmed in a contemporary external validation cohort.

It has been known for decades that ICU outcome is worse with higher age (38); comorbidities have a notable impact on ARDS survival (39); patients with severe lung damage have lower Pao_2/Fio_2 (19); there is a direct relationship between Pplat and mortality (40); and the greater the number of extrapulmonary OFs the higher the mortality (41). We have previously shown that restricting ARDS severity to the hypoxemia level could lead to discrepancies in outcome prediction (3, 4). Since baseline Pao_2/Fio_2 is impacted by clinician-set ventilatory strategies (3, 4, 42), an important feature of our work is that the oxygenation parameters included in our models were obtained using standardized ventilatory settings after a 24-hour stabilization period (3, 9, 19, 24).

Most clinical trials in ARDS tested the effectiveness of therapies in populations with highly variable baseline characteristics and lack of assessment of the progression of modifiable clinical features (43, 44). It is unsurprising, that a recent systematic review has found significant unexplained heterogeneity in the 28-day mortality of control groups (44). Our findings emphasize the importance of standardized assessment of ventilation following the initial diagnosis of ARDS. Two out of the four modifiable clinical variables responsible for the predictive power of our POSTCARDS model were recorded during these standardized MV settings. In our study, which focused on prognostic, risk features did not identify which patients are likely to respond to any specific treatment. More research is needed to identify predictive variables that are modifiable (10).

A novel finding of our study is that two different ML techniques were not able to outperform logistic regression or the previously developed SPIRES scoring system. While prediction modeling is one of the most common ML applications, we should be realistic on the role of ML in this clinical domain (45). Previous efforts to improve generalizable predictive models have been hampered by lack of standardized datasets obtained from heterogeneous populations (1, 4, 19, 44, 46). Regardless of which tool we chose to predict ARDS outcome, it must be assessed under precise standardized conditions (46) to make clinical decisions and actions replicable (47). We have addressed this problem by ensuring that the ML algorithms started from a comparable disease severity state across all patient cohorts. Interestingly, we found that the SPIRES score (4), a simple scoring system with high explanatory predictive power, can predict ICU death in moderate/severe ARDS with a parsimonious four-variable model. Although critical care physicians must deal with about 200 variables when caring for an ARDS patient (48), human working memory is limited to 4 ± 1 constructs (49) with a degradation in clinical decision-making once the limit of four constructs is exceeded (48, 49).

We recognize that the major focus of clinicians is not the prediction of outcomes in individual patients using population level data. As suggested by experts in the field of critical illness, we believe that the current ARDS-based framework of illness should be reconsidered (10). Clinicians are interested in actionable and modifiable variables for improving expected outcomes (10). Based on our findings, the following are reasonable targets within the first 24 hours of ARDS management: improving the oxygenation to achieve Pao_2/Fio_2 greater than 150 mm Hg, reducing the lung strain by using Pplat less than 29 cm H_2O, and aiming to reduce the number of extrapulmonary OFs. Targeting these variables have high external validity as it has been shown that improving oxygenation and limiting

lung-stress with early interventions such as prone positioning can decrease mortality (50). The presence of extrapulmonary OF has also been highlighted by others as a significant risk variable (51). However, the best strategies to achieve improvement of nonpulmonary OF in ARDS has not been elucidated and should be subject to further research.

Our study has a several strengths. First, our prediction model was developed from a large patient population that reflects current clinical practice for patients with moderate/severe ARDS. Second, ML methods that incorporate prior clinical knowledge satisfying face validity, have the benefit of leading to interpretations that are more relevant, and less likely to generate unreasonable predictions (52). Third, our models provided good fit to the various datasets obtained from the multicenter studies, further increasing generalizability.

We acknowledge some limitations. First, the POSTCARDS prediction model was developed using data from patients managed with lung-protective ventilation, and hence may not be valid for patients ventilated with large Vt or high Pplat. Second, some overfitting may have occurred because the initial number of candidate variables was 34, but our variable selection method is based on a survival of the fittest approach to modeling data. Using genetic algorithm-based modeling has been shown to perform better than other advanced variable selection methods in large ICU datasets (53). Third, we considered only a limited number of ML methods; other approaches may lead to better prediction models. Finally, our model needs further validation beyond our current cross-validation and external validation approach.

In summary, ML models may not provide advantages over regression models or simple scores for predicting ICU death in patients with moderate-to-severe ARDS. Clinical determinants of ICU death in ARDS are multifactorial. Our study clarifies that biology, as represented by a limited number of key characteristics, is key to prediction rather than the specific learning method (classic regression or ML). Demographics, comorbidities, and potentially modifiable variables such as lung mechanics, oxygenation, and extrapulmonary OF may predict outcome. Even if prediction models are highly accurate, they are unlikely to improve clinical outcomes unless they are linked to effective interventions, and recommendations or actions are integrated into ARDS management. Future research should address precision medicine in ARDS, invoking the concept of treatable traits (10), specific physiologic derangements that portend a response to a particular intervention.

ACKNOWLEDGMENTS

We thank Mr. Yasser and Lily B of LP Bahrain for their support in loving memory of Lily Bendahan. We also recognize Dr. R. M. Kacmarek (now deceased) for his contribution to the initial study concept and design.

1 *CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain.*

2 *Research Unit, Hospital Universitario Dr. Negrín, Las Palmas de Gran Canaria, Spain.*

3 *Li Ka Shing Knowledge Institute at St. Michael's Hospital, Toronto, ON, Canada.*

4 *Departament de Matemàtiques i Informàtica, Universitat de Barcelona (UB), Barcelona, Spain.*

5 *Big Data Department, PMC-FPS, Regional Ministry of Health and Consumer Affairs, Sevilla, Spain.*

6 *Intensive Care Unit, Hospital General Universitario de Ciudad Real, Ciudad Real, Spain.*

7 *Intensive Care Unit, Hospital Universitario de A Coruña, La Coruña, Spain.*

8 *Intensive Care Unit, Hospital Universitario Virgen de Arrixaca, Murcia, Spain.*

9 *Intensive Care Unit, Hospital Universitario Río Hortega, Valladolid, Spain.*

10 *Surgical Intensive Care Unit, Department of Anesthesia, Hospital Clinic, IDIBAPS, Barcelona, Spain.*

11 *Intensive Care Unit, Complejo Asistencial Universitario de León, León, Spain.*

12 *Intensive Care Unit, Hospital Universitario La Paz, IdiPaz, Madrid, Spain.*

13 *Intensive Care Unit, Hospital Clínico Universitario de Valladolid, Valladolid, Spain.*

14 *Intensive Care Unit, Hospital Universitario NS de Candelaria, Santa Cruz de Tenerife, Spain.*

15 *Intensive Care Unit, Hospital Virgen de La Luz, Cuenca, Spain.*

16 *Intensive Care Unit, Hospital Universitario Fundación Jiménez Díaz, Madrid, Spain.*

17 *Thoracic Surgery, Hospital Universitario Dr. Negrín, Las Palmas de Gran Canaria, Spain.*

18 *Department of Intensive Care Medicine & Anesthesia, Aneurin Bevan University Health Board, Newport, United Kingdom.*

19 *Cardiff University, Cardiff, United Kingdom.*

20 *Department Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands.*

21 Division of Critical Care Medicine, University of Toronto, Toronto, ON, Canada.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/ccmjournal).

Drs. Villar, González-Martín, Steyerberg, Hernández-González, Armengol, Szakmany, and Slutsky contributed to the initial study concept and design. Drs. Villar, González-Martín, Añón, Ferrando, and Rodríguez-Suárez obtained funding for the study. All authors contributed to the final study design, or participated in its coordination, or participated in drafting the first article. Drs. Ferrando, Añón, Martín-Rodriguez, Mosteiro, Martínez, Sánchez-Ballesteros, Dominguez-Berrot, Parra, Montiel, Solano, and Robaglia enrolled patients into the study and participated in the data collection and data analysis. Dr. Villar, Ms. R. L. Fernández, Ms. C. Fernández, Ms. Gómez-Bentolila, Dr. González-Martín, Dr. Steyerberg, Dr. Hernández-González, Dr. Armengol, Dr. Szakmany, and Dr. Slutsky are responsible for data analysis and/or interpretation of data. Dr. Villar, Ms. R. L. Fernández, Ms. C. Fernández, Ms. Gómez-Bentolila, and Dr. González-Martín had full access to all data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Dr. Villar, Dr. González-Martín, Dr. Hernández-González, Dr. Armengol, Ms. C. Fernández, Dr. Ferrando, Dr. Añón, Dr. Szakmany, Dr. Steyerberg, and Dr. Slutsky contributed to the final draft of the article. Drs. Villar, González-Martín, Szakmany, Steyerberg, and Slutsky contributed to the final revised article. All authors read and approved the final article.

This study was funded by the Instituto de Salud Carlos III, Madrid, Spain (PI/19/0141), The European Regional Development's Funds, Fundación Canaria Instituto de Investigación Sanitaria de Canarias, Spain (PIFIISC20-51, PIFIISC21-36), and Asociación Científica Pulmón y Ventilación Mecánica, Spain.

Drs. Martín-Rodríguez and Rodríguez-Suárez received support for article research from the National Institutes of Health. Dr. Rodríguez-Suárez disclosed work for government. Dr. Szakmany received funding from PAION UK and Thermo Fisher UK; he disclosed that they are a trustee of Intensive Care National Audit & Research Centre and an Associate Editor for Social Media for Critical Care Explorations*. Dr. Slutsky received funding from Signal-1. Dr. Villar, Dr. Añón, Dr. Ferrando, Ms. Fernández, and Dr. González-Martin received grant support from the Instituto de Salud Carlos III, Madrid, Spain (CB06/06/1088). Dr. Hernández-González is a Serra Húnter fellow. Dr. Slutsky was funded by the Canadian Institutes of Health Research (grants numbers 137772 and FDN143285). The remaining authors have disclosed that they do not have any potential conflicts of interest.*

For information regarding this article, E-mail: jesus.villar54@gmail.com

Members of the Predicting Outcome and STratifiCation of severity in ARDS (POSTCARDS) Project are listed in the Supplemental File (http://links.lww.com/CCM/H413).

All data needed to evaluate the conclusions in this article are presented and tabulated in the main text or the Supplemental File (http://links.lww.com/CCM/H413). Data are available from the corresponding author on reasonable request.

The study was approved by the Ethics Committees for Clinical Research of Hospital Universitario Dr. Negrín (Las Palmas de Gran Canaria, Spain), and the requirement for informed consent was waived (Reference CEI/CEIm 2021-321-1) under the Royal Decree 1090/2015 of December 2015, and the Royal Decree 957/2020 of November 2020 of the Spanish legislation for biomedical research based on the retrospective nature of the secondary analysis, the anonymization/dissociation of data, and with no harm and no benefit for the management of patients. This was a comprehensive analysis using unrestricted data from our previously published studies in patients with moderate-to-severe acute respiratory distress syndrome that were approved by the referral Ethics Committees of Hospital Universitario Dr. Negrín (Las Palmas de Gran Canaria, Spain), Hospital Virgen de La Luz (Cuenca, Spain), Hospital Clínico Universitario (Valladolid, Spain), Hospital Universitario La Paz (Madrid, Spain), and Hospital Clínico de Valencia (Valencia, Spain), and adopted by all participating centers, as required by the Spanish legislation. This study was in accordance with the fundamental principles established in the Declaration of Helsinki, the Convention of the European Council related to human rights and biomedicine, the Ethical Guidelines for Health-related Research Involving Humans by the Council for International Organization of medical Sciences of the World Health Organization, and within the requirements established by the Spanish legislation for biomedical research, the protection of personal data, and bioethics.

This study was registered on November 8, 2022, at ClinicalTrials.gov (NCT 05611177).

REFERENCES

1. Villar J, Slutsky AS: Golden anniversary of the acute respiratory distress syndrome: Still much work to do!. *Curr Opin Crit Care* 2017; 23:4–9
2. Ferring M, Vincent JL: Is outcome from ARDS related to the severity of respiratory failure? *Eur Respir J* 1997; 10:1297–1300
3. Villar J, Blanco J, del Campo R, et al; Spanish Initiative for Epidemiology, Stratification & Therapies for ARDS (SIESTA) Network: Assessment of PaO_2/FiO_2 for stratification of patients with moderate and severe acute respiratory distress syndrome. *BMJ Open* 2015; 5:e006812
4. Villar J, González-Martín JM, Ambrós A, et al; Spanish Initiative for Epidemiology, Stratification and Therapies of ARDS (SIESTA) Network: Stratification for identification of prognostic categories in the acute respiratory distress syndrome (SPIRES) score. *Crit Care Med* 2021; 49:e920–e930
5. Brower RG, Matthay MA, Morris A, et al; Acute Respiratory Distress Syndrome Network: Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med* 2000; 342:1301–1308
6. Guerin C, Reignier J, Richard JC, et al; PROSEVA Study Group: The PROSEVA Study Group: Prone positioning in severe acute respiratory distress syndrome. *N Engl J Med* 2013; 368:2159–2168
7. Moss M, Huang DT, Brower RG, et al; National Heart, Lung, and Blood Institute PETAL Clinical Trials Network: Early neuromuscular blockade in the acute respiratory distress syndrome. *N Engl J Med* 2019; 380:1997–2008
8. Combes A, Hajage D, Capellier G, et al; EOLIA Trial Group, REVA, and ECMONet: Extracorporeal membrane oxygenation

for severe acute respiratory distress syndrome. *N Engl J Med* 2018; 378:1965–1975

9. Villar J, Ferrando C, Martínez D, et al; dexamethasone in ARDS network: Dexamethasone treatment for the acute respiratory distress syndrome: A multicentre, randomised controlled trial. *Lancet Respir Med* 2020; 8:267–276
10. Maslove DM, Tang B, Shankar-Hari M, et al: Redefining critical illness. *Nature Med* 2022; 28:1141–1148
11. Ranieri VM, Rubenfeld GD, Thompson BT, et al; ARDS Definition Task Force: Acute respiratory distress syndrome: The Berlin definition. *JAMA* 2012; 307:2526–2533
12. Pirrachio R, Petersn ML, Carone M, et al: Mortality prediction in intensive care units with the Super ICU Learner Algorith (SICULA): A population-based study. *Lancet Respir Med* 2015; 3:42–52
13. Nemati S, Holder A, Razmi F, et al: An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med* 2018; 46:547–553
14. Ding XF, Li JB, Liang HY, et al: Predictive model for acute respiratory distress syndrome events: A secondary analysis of a cohort study. *J Transl Med* 2019; 17:326
15. Huang B, Liang D, Zon R, et al: Mortality prediction for patients with acute respiratory distress syndrome based on machine learning: A population-based study. *Ann Transl Med* 2021; 9:794
16. Sayed M, Riaño D, Villar J: Novel criteria to classify ARDS severity using machine learning approach. *Crit Care* 2021; 25:150
17. Collins GS, Reitsma JB, Altman DG, et al: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *J Clin Epidemiol* 2015; 68:112–121
18. Villar J, Blanco J, Añón JM, et al; ALIEN Network: The ALIEN study: Incidence and outcome of acute respiratory distress syndrome in the era of lung protective ventilation. *Intensive Care Med* 2011; 37:1932–1941
19. Villar J, Pérez-Méndez L, Blanco J, et al; Spanish Initiative for Epidemiology, Stratification, and Therapies for ARDS (SIESTA) Network: A universal definition of ARDS: The PaO_2/FiO_2 ratio under s standard ventilatory setting – a prospective, multicenter validation study. *Intensive Care Med* 2013; 39:583–592
20. Villar J, Mora-Ordoñez JM, Soler JA, et al: The PANDORA study: Prevalence and outcome of acute hypoxemic respiratory failure in the pre-COVID era. *Crit Care Explor* 2022; 4:e0684
21. Vergouwe Y, Steyerberg EW, Eijkemans MJC, et al: Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005; 58:475–483
22. Leisman DE, Harhay MO, Lederer DJ, et al: Development and reporting of prediction models: Guidance for authors from editors of respiratory, sleep, and critical care journals. *Crit Care Med* 2020; 48:623–633
23. Knaus WA, Draper EA, Wagner DP, et al: APACHE II: A severity of disease classification system. *Crit Care Med* 1985; 13:818–829
24. Villar J, Pérez-Méndez L, López J, et al; HELP Network: An early $PEEP/FiO_2$ trial identifies different degrees of lung injury in patients with acute respiratory distress syndrome. *Am J Respir Crit Care Med* 2007; 176:795–804
25. Vincent JL, de Mendonça A, Cantraine F, et al: Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: Results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. *Crit Care Med* 1998; 26:1793–1800
26. Villar J, Ambrós A, Mosteiro F, et al: Age, PaO_2/FiO_2 and plateau pressure score: A proposal for a simple outcome score in patients with acute respiratory distress syndrome. *Crit Care Med* 2016; 44:1361–1369
27. Villar J, Martín-Rodriguez C, Dominguez-Berrot AM, et al; Spanish Initiative for Epidemiology, Stratification and Therapies for ARDS (SIESTA) Investigators Network: A quantile analysis of plateau and driving pressure: Effects on mortality in patients with acute respiratory distress syndrome receiving lung-protective ventilation. *Crit Care Med* 2017; 45:843–850
28. Scrucca L: GA: A package for genetic algorithms in R. *J Stat Softw* 2013; 53:1–37
29. González-Martin JM, Sánchez-Medina AJ, Alonso JB: Optimization of the prediction of financial problems in Spanish private health companies using genetic algorithm. *Gac Sanit* 2019; 33:462–467
30. Vrieze SI: Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods* 2012; 17:228–243
31. Ioannidis JPA: The proposal to lower P value thresholds to 0.005. *JAMA* 2018; 319:1429–1430
32. Gutierrez G: Artificial intelligence in the intensive care unit. *Crit Care* 2020; 24:101
33. Rashid M, Ramakrishnan M, Pulikkel V, et al: Artificial intelligence in acute respiratory distress syndrome: A systematic review. *Artif Intell Med* 2022; 131:102361
34. Chen T, He T, Benesty M, et al: Xgboost: Extreme Gradient Boosting. R package version 0.4-2. 2015. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/
35. Liaw A, Wiener M: Classification and regression by randomForest. *R News* 2002; 2:18–22
36. Steyerberg EW, Vergouwe Y: Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *Eur Heart J* 2014; 35:1925–1931
37. Van Calster B, Nieboer D, Vergouwe Y, et al: A calibration hierarchy for risk models was defined: From utopia to empirical data. *J Clin Epidemiol* 2016; 74:167–176
38. Gee MH, Gottlieb JE, Albertine KH, et al: Physiology of aging related to outcome in the adult respiratory distress syndrome. *J Appl Physiol (1985)* 1990; 69:822–829
39. Soubani AO, Shehada E, Chen W, et al: The outcome of cancer patients with acute respiratory distress syndrome. *J Crit Care* 2014; 29:183.e7–183.e12
40. Shiu KK, Rosen MJ: Is there a safe plateau pressure threshold for patients with acute lung injury and acute respiratory distress syndrome? *Am J Respir Crit Care Med* 2006; 173:686; author reply 687
41. Villar J, Martínez D, Mosteiro F, et al; Stratification and Outcome of Acute Respiratory Distress Syndrome (STANDARDS)

Network: Is overall mortality the right composite endpoint in clinical trials of acute respiratory distress syndrome? *Crit Care Med* 2018; 46:892–899

42. Møller MH, Derde LPG, Sweeney RM: Focus on clinical trial interpretation. *Intensive Care Med* 2020; 46:790–792
43. Villar J, Ferrando C, Tusman G, et al: Unsuccessful and successful clinical trials in acute respiratory distress syndrome: Addressing physiology-based gaps. *Front Physiol* 2021; 12:774025
44. Juschten J, Tuinman PR, Guo T, et al: Between-trial heterogeneity in ARDS research. *Intensive Care Med* 2021; 47:422–434
45. Sennaar K: Machine Learning Medical Diagnostics—4 Current Applications. Emerj Artificial Intelligence Research. 2018. Available at: https://emerj.com/ai-sector-overviews/machine-learning-medical-diagnostics- 4-current- applications/. Accessed February 27, 2023
46. Kacmarek RM, Berra L: Prediction of ARDS outcome: What tool should I use? *Lancet Respir Med* 2018; 6:253–254
47. Morris AH, Stagg B, Lanspa M, et al: Enabling a learning healthcare system with automated computer protocols that produce replicable and personalized clinical actions. *J Am Med Inform Assoc* 2021; 28:1330–1344
48. Morris AH: Human cognitive limitations. Broad, consistent, clinical application of physiological principles will require decision support. *An Am Thorac Soc* 2018; 15:S53–S56
49. Cowan N: The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav Brain Sci* 2001; 24:87–114; discussion 114–185
50. Guerin C, Albert RK, Beitler J, et al: Prone position in ARDS patients: Why, when, how and for whom. *Intensive Care Med* 2020; 46:2385–2396
51. Kallet RH, Lipnick MS, Zhuo H, et al: Characteristics of nonpulmonary organ dysfunction at onset of ARDS based on the Berlin definition. *Respir Care* 2019; 64:493–501
52. Steyerberg EW, Vickers AJ, Cook NR, et al: Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology* 2010; 21:128–138
53. Zhang F, Luo C, Lan C, et al: Benchmarking feature selection methods with different prediction models on large-scale healthcare event data. *BenchCouncil Transactions on Benchmarks, Standards and Evaluation* 2021; 1:100004