***- SUPPLEMENTAL DATA FILE -***

**SPIRES score: Stratification for identification of Prognostic categories In acute RESpiratory distress syndrome**

**Jesús Villar**, MD, PhD; **Jesús M. González-Martín**, PhD; **Alfonso Ambrós**, MD, PhD;

**Fernando Mosteiro**, MD, PhD; **Domingo Martínez**, MD; **Lorena Fernández**, MD;

**Juan A. Soler**, MD, PhD; **Laura Parra**, MD; **Rosario Solano**, MD; **Marina Soro**, MD, PhD;

**Rafael del Campo**, MD; **Raúl I. González-Luengo**, MD; **Belén Civantos**, MD; **Raquel Montiel**, MD; **Lidia Pita-García**, MD; **Anxela Vidal**, MD; **José M. Añón**, MD, PhD; **Carlos Ferrando**, MD, PhD; **Francisco J. Díaz-Dominguez**, MD; **Juan M. Mora-Ordoñez**, MD; **M. Mar Fernández**, MD, PhD; **Cristina Fernández**, MSc; **Rosa L. Fernández**, MSc; **Pedro Rodríguez-Suárez**, MD;

**Ewout W. Steyerberg,** PhD; **Robert M. Kacmarek**, RTT, PhD;

for the “Spanish Initiative for Epidemiology, Stratification and Therapies of ARDS (SIESTA) Network

**Corresponding author**:

Jesús Villar, MD, PhD

Multidisciplinary Organ Dysfunction Evaluation Research Network

Hospital Universitario Dr. Negrín

Barranco de la Ballena s/n - 4th floor, South wing

35019 Las Palmas de Gran Canaria. Spain.

E-mail: jesus.villar54@gmail.com

**SUPPLEMENTARY METHODS**

This study has been conducted according to the principles of the Declaration of Helsinki approved by the World Medical Association (1), the European Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the application of Biology and Medicine, and within the requirements established by Spanish legislation for biomedical research. Our studies were approved by the Ethics Committees for Clinical Research at the Hospital Universitario Dr. Negrín (Las Palmas de Gran Canaria, Spain, approval No. 2008-0915-EPI), Hospital Virgen de La Luz (Cuenca, Spain, approval No. 2014/PI 1114), Hospital Clínico Universitario de Valladolid (Valladolid, Spain, approval No. PI17-594) and the local institutional review boards of all participating hospitals. The studies were granted a waiver of the need for informed consent. Patients’ data were anonymized and recorded in a secure, computer-based case report form specifically designed for the study. None of the findings reported in the present study have been published elsewhere.

**Justification of the study**

Although the clinical definition and classification of the acute respiratory distress syndrome (ARDS) has been refined since its first description (2), current definition of ARDS has a limited ability in identifying categories of ARDS patients with similar severity and/or outcome (3,4). At present, patients are classified as having mild, moderate, or severe ARDS based on the ratio of arterial oxygen partial pressure to fractional inspired oxygen (PaO2/FiO2) at ARDS onset. Attempting to predict outcomes based solely in Pa2/FiO2 on this basis is inherently flawed because it does not consider the sensitivity of PaO2 to ventilator settings and the effects of routine care during the first 24 hours after its diagnosis (4,5). There is a need for a new conceptual approach to ARDS classification. There is not a scoring system to assess on how an ARDS patient is presenting or progressing. Scoring systems provide an average prediction value for ranking patients according to how well the score anticipate the true outcome in the intensive care unit (ICU or in the hospital. We hypothesized that a clinical ARDS score could stratify and identify distinct prognostic classes of ARDS (termed the SPIRES score), and could be helpful for selecting ARDS patients for therapeutic clinical trials.

**Patient population**

This study was conducted in three steps. For the first two steps (model development and internal/internal-external validation), we performed a secondary analysis of an unrestricted set of pooled data from 1,000 adult patients included prospectively in three multicenter, observational cohorts enrolling consecutive patients with moderate-to-severe ARDS managed with lung-protective mechanical ventilation (MV) in a network of ICUs in Spain under an Initiative for Epidemiology, Stratification and Therapies of ARDS (SIESTA) Program (6-9). In the ALIEN cohort (ClinicalTrials.gov NCT00736892), 22 participating ICUs included 300 patients from September 2008 to May 2010, from which 255 patients were used to estimate the 1-year incidence of moderate/severe ARDS in 13 geographical areas of Spain (6). In the STANDARDS cohort (NCT02288949), 24 participating ICUs included 300 patients from September 2013 to July 2015, and were used to quantify the risk of death in ARDS (7) and for testing whether driving pressure was superior to the variables that define it in predicting outcome in ARDS patients (8). The STANDARDS-2 cohort was designed as a continuation of the STANDARDS cohort with the purpose of having a large database of ARDS patients combining our collective efforts. In the STANDARDS-2 cohort (NCT02836444), 21 participating ICUs included 400 patients from August 2015 to April 2017, and were used to determine whether an enrichment strategy could be useful for selecting patients into future clinical trials (9).

Patients admitted to participating ICUs were screened daily during the study periods. All consecutive patients meeting the American-European Consensus Conference (AECC) criteria for ARDS (10) on positive end-expiratory pressure (PEEP) ≥5 cmH2O (in the ALIEN cohort) and the Berlin criteria for moderate or severe ARDS (3) (in the STANDARDS and STANDARDS-2 cohorts) were included. By leaving the assessment of PaO2/FiO2 essentially unchanged, the AECC definition and the Berlin criteria are essentially identical (4). The requirement of a minimum PEEP level of 5 cmH2O has no impact on the definition since it is hard to conceive that a patient with ARDS would be managed with PEEP<5 cmH2O. Thus, our screening applies only to patients with moderate-to-severe ARDS, which include: (i) having an initiating clinical condition (pneumonia, aspiration, inhalation injury, sepsis, trauma, acute pancreatitis, etc.), (ii) within one week of the known clinical insult or new or worsening respiratory symptoms, (iii) bilateral pulmonary infiltrates on chest imaging (chest radiograph or computed tomography scan), (iv) absence of left atrial hypertension or no clinical signs of left heart failure, (v) hypoxemia (as defined by a PaO2/FiO2 ≤200 mmHg on PEEP≥5 cmH2O, regardless of FiO2). We did not enroll patients with mild ARDS (PaO2FiO2>200 mmHg). However, we are confident that no patients with mild ARDS were excluded during our observational periods if they moved to a more severe category, although we do not have data on the precise number of those patients.

Onset of ARDS was defined as the day in which the patient first met moderate/severe ARDS criteria. For qualifying values of PaO2/FiO2, physicians considered blood gases while patients were clinically stable and did not consider a transient fall in PaO2 resulting from acute events unrelated to the disease process. We excluded patients younger than 18 years old, patients with severe chronic pulmonary disease, acute cardiac failure, brain death, patients with a do-not-resuscitate orders, or postoperative patients receiving MV for <24 hours. Also, because diagnostic confusion could occur with other diseases that cause hypoxemia and bilateral pulmonary infiltrates, physicians excluded lymphangitic carcinoma, acute eosinophilic pneumonia, hypersensitivity pneumonitis, and idiopathic pulmonary fibrosis.

Although patient care was not strictly protocolized, clinicians were asked to follow current guidelines for the general critical care management, which included the following: (i) in case of sepsis, physicians were urged to ensure early identification of causative microorganism, intravenous administration of antibiotics as soon as sepsis was suspected or recognized, and to optimize antibiotic selection and timely administration on the basis of antibiogram; (ii) fluid resuscitation and vasopressor use were individualized with the goal of maintaining a systolic blood pressure ≥90 mmHg or a mean arterial pressure ≥65 mmHg; (iii) to maintain hemoglobin between 7 to 10 g/dL. For ventilatory management, clinicians were encouraged to follow current recommendations for lung-protective MV with a tidal volume (VT) of 4-8 ml/kg predicted body weight (PBW), a plateau pressure (Pplat) <30 cmH2O, a ventilatory rate (RR) to maintain a PaCO2 between 35-50 mmHg (permissive hypercapnia was allowed to target VT), and PEEP and FiO2 combinations according to the PEEP-FiO2 table of the ARDS protocol (11), ensuring that among the PEEP and FiO2 combinations, clinicians should use the PEEP levels that allowed the reduction of FiO2 to the lowest level for maintaining a PaO2 within a target range of 60 to 100 mmHg or a peripheral capillary oxygen saturation (SpO2) within a target range of 90 to 98%. The choice of drugs for sedation and analgesia, early neuromuscular blockade, prone positioning, recruitment maneuvers, hemodynamic management modalities, and the decision to perform a tracheotomy were left to the discretion of the attending physician. PBW was calculated using the following equations: 50 + 0.91 x [height (cm) – 152) for men and 45.5 + 0.91 x [height (cm) – 152] for women (11). None of the patients were included in any clinical trial or received nitric oxide, activated protein C, or high frequency ventilation. Although prone positioning and recruitment maneuvers were used in some patients, we do not have data on timing of prone positioning, or whether prone ventilation and recruitment maneuvers were applied as a rescue therapy, as a routine practice, or following any specific protocol. Weaning off MV was not strictly protocolized, but could be started when the attending physician considered it clinically appropriate. Patients were assessed daily for readiness for a spontaneous breathing trial (SBT) based on the ARDSnet protocol (11). In general, pre-requisites for the SBT included a PaO2/FiO2>200 with PEEP<10 cmH2O, no vasopressors, continuous sedation minimized, and ability to cough during tracheal aspirations. Spontaneous ventilation was tested with a T-piece or with pressure support at 8 cmH2O. The duration of the SBT was at least 30 min and no longer than 120 minutes. If the patient passed the trial, a decision for extubation was taken, unless there was a specific reason not to extubate. Weaning and the decision to extubate were left to the discretion of the responsible clinician. Decisions about noninvasive ventilation, reintubation or extubation, were dictated by common clinical criteria.

**Data collection and follow-up**

Data were collected in each participating ICU using standardized case report forms, and transmitted to the coordinating center (Hospital Universitario Dr. Negrin) when the patient was discharged from hospital. Before exporting the data into a computerized database, a trained data collector from the coordinating center checked the completeness and the quality of information. Logical checks were performed for missing data and to find inconsistencies, especially regarding clinical diagnosis, date, and severity scores. If necessary, the data collector contacted the investigator by phone to validate the data or reformat the data for entry into the database. For the purpose of this study, we analyzed information from 121 clinical variables including demographics, comorbidities, main cause of ARDS, Acute Physiology And Chronic Health Evaluation II (APACHE II) score (12) during the first 24 hours of ARDS diagnosis, hemodynamics (blood pressure and heart rate) and data from ventilator settings and lung mechanics (VT, RR, PEEP, peak inspiratory pressure, Pplat), gas exchange (PaO2, PaCO2, FiO2, PaO2/FiO2, pH), at ARDS onset and at 24 hours of ARDS diagnosis. We recorded the occurrence of extrapulmonary organ system failures (OF) included in the Sequential Organ Failure Assessment (SOFA) scale (13) at ARDS onset and after 24 hours of ARDS diagnosis. Those extrapulmonary organ systems are: cardiovascular system, liver, kidney, coagulation system, and central nervous system. Since the term “organ dysfunction” is unclear and because organ dysfunction may emerge for reasons other than sepsis, extrapulmonary organ failure was defined as an acute change in organ-specific SOFA score ≥2 (14,15). The baseline SOFA score was assumed to be zero in patients not known to have preexisting organ dysfunction. Sepsis was defined according to 2001 International Consensus Conference criteria (16). We also collected complications, administration of vasopressors, and temporal changes in organ dysfunction as assessed by the SOFA scale at ARDS onset and at 24 hours. We recorded the actual duration of MV, the length of stay in the ICU and the length of stay in the hospital. We recorded date and status (alive or dead) of the patient at ICU and hospital discharge, and causes of death.

**Statistical analysis plan**

We first described the full dataset of 1,000 patients. In each patient, we collected data from 121 clinical variables. Based on previous work by our group (7,8), we focused our analysis on clinically relevant variables collected within the first 24 hours of ARDS diagnosis to estimate the probability of ICU death, independent of the underlying disease or cause of death (**Figure S1**). We listed in sequential order the values of each variable for all 1,000 patients at ARDS diagnosis (Day 0) and after 24 hours of general management and treatment (Day 1). Although the distribution of values of all variables identified patients with a wide range of ICU mortality risk, we narrowed the search to the following variables, as potential early predictors of outcome: age, gender, and APACHE II score, SOFA score, number of extrapulmonaray OF, PaO2, PaO2/FiO2, PaCO2, pH, FiO2, VT, RR, PEEP, Pplat, driving pressure, minute ventilation (as an indirect measurement of dead space, and defined by VT x RR, in liters/min) at ARDS onset and 24 hours later (**Table S1**). For the purpose of this study, the values of PaO2/FiO2 and Pplat at 24 hours were measured under a standardized ventilatory setting. At 24 hours after meeting moderate/severe ARDS criteria, PaO2/FiO2 and Pplat were assessed in all patients under standardized ventilatory settings (PEEP=10 cmH2O and FiO2=0.5) (5). When patients required PEEP>10 or FiO2>0.5 and could not tolerate a decrease in PEEP or FiO2, a set of rules for setting PEEP and FiO2 were applied *only* during the standardized assessment, as described and validated previously by our group (5,9). At other times, PEEP and FiO2 levels were up to the discretion of managing clinicians. We calculated driving pressure (Pplat minus PEEP) in all patients. However, since whether driving pressure relates causally to outcome remains to be established in randomized controlled trial, we valued Pplat over driving pressure for prognosis in our scoring model based on our previous work (8).

***Predefined rules and pre-specified statistical analysis for building the score***

For building the final SPIRES scoring model, we considered the minimum number of variables that provided a similar performance as the full 28-predictor model. We defined and specified in advance rules and expectations before the final statistical analysis was conducted, realizing that overly detailed analyses could produce overoptimistic results due to a combination of reduced statistical power to detect real differences, or due to an increase in the variance around the mean estimate, and/or an increased statistical likelihood of a false finding when many subgroups are examined. First, for selecting thresholds of risk for ICU death among the minimum selected variables, we separated patients into subgroups by identifying the cut-offs where there appeared to be a stepwise increase in ICU mortality. Second, we required at least an absolute 15% ICU mortality difference between subgroups of severity and/or outcome for each risk variable at a 0.005 significance level (17). Third, we tried to stratify the ranges of each variables into a low, middle, and high-risk category, akin to how clinicians routinely categorize patients into risk groups. Forth, we expected at least 100 patients and/or 50 deaths in each subgroup for selecting the risk variables. The decision for demanding a minimum number of 50 deaths was based on a review of 159 randomized clinical trials that tested a variety of interventions in patients with ARDS (18). Fifth, for rating the thresholds of the prognostic variables, we assigned 1 point for the category with the lower risk of ICU death, 2 points for the category within the proximity of the overall ICU mortality of the sample population (middle risk), and 3 points for the category associated with an absolute risk of ICU death of at least 15% higher than the middle risk subgroup. Each patient was given a score based on the total points for each risk threshold and their overall ICU mortality was determined. Sixth, we attempted to build a scoring model in which the individual scores could aggregate patients into four or less classes or prognostic phenotypes with clear separation of survival from the first week of ARDS diagnosis.

***Internal and Internal-external validation of the score***

Hundreds of models for clinical prediction of outcome are published every year. A popular approach is to randomly split the data into two parts: one to develop the model (training cohort) and another to measure its performance (validation cohort). However, this split-sample method is often inefficient (19,20). For internal validation, bootstrapping is attractive, a computerized resampling technique which replicates the process of sample generation by drawing samples with replacement from the original dataset (21). We searched in the data for model specification since the model was not pre-specified. Once the predictors or risk variables were identified by performing univariate logistic regression analysis, we performed a multivariable logistic regression analysis. In our bootstrap analysis, we repeated the stepwise selection of variables with forward and backward techniques in 2,000 samplings. We used bootstrapping for internal validation because it gives reasonably valid estimates of the expected optimism in predictive performance provided that any selection of predictors is taken into account (22). With the probabilities obtained on the logistic model, we evaluated the areas under the receiver operating characteristic (ROC) curves (AUC), and estimated the 95% confidence intervals (CI). Test performance was defined as the performance of models from the bootstrap samples when applied to the original sample. The difference between the bootstrap performance and the test performance (“optimism”) was averaged over 200 replications for obtaining a stable estimate. The optimism was also approximated by internal validation of the full 28-predictor model. Then, we calculated the optimism-corrected performance, defined as the difference between the apparent performance in the sample and the optimism.

To adjust for optimism in the final SPIRES scoring model, we emphasize that we validated the performance of the final score as the apparent performance of the minimum-variables-predictor model minus the optimism calculated for the 28-predictor model. The final minimum-variables score should provide a similar performance as the full model.

Since prediction models often perform poorly when assessed in external validation studies, we performed internal-external validation in independent parts of the data by leaving each of the three parent cohorts out once (20-22). Therefore, we assessed the apparent performance of a model developed in the other cohorts (“internal-external validation”) (20) (**Figure S2**). The strength for assessing external validation increases when studies include patients from different hospitals, as it occurred in our patient population.

Based on the results of the scoring model, we aggregated patients with the same or similar scores into major prognostic classes with clear differences in ICU mortality. Since each patient was given a score based on the risk variables, we compared the predictive value of the SPIRES score with four other models: baseline PaO2/FiO2 at the time of diagnosis of moderate-to-severe ARDS -as mandated by current ARDS definition-, APACHE II score on the day of ARDS diagnosis, SOFA score at the time of ARDS diagnosis, and PaO2/FiO2 at 24 hours of ARDS diagnosis.

***External validation of the score***

As the third step, and for solving the complexity of validating our scoring model by internal and internal-external validation with 1,000 patients enrolled in three independent cohorts, we tested the performance of the model in fully new patients. We analyzed a cohort of 301 consecutive patients with moderate-to-severe ARDS included in the multicenter, observational “Prevalence AND Outcome of acute hypoxemic Respiratory fAilure (PANDORA)” study (NCT03145974, unpublished data). With this approach, we studied the temporal aspect of external validity since this new cohort contains recently treated ARDS patients. Patients were admitted in a network of 22 ICUs from May 2017 to March 2018. This external validation cohort has a sufficient number of events required for external validation (23). As recommended by recent guidelines (24), we avoided the retraining on the external dataset (**Figure S2**).

***Statistical Analysis Plan***

We calculated the mean, standard deviation (SD), median and 25-75% percentiles (P25%-75%) of quantitative variables. We used the Shapiro-Wilk to test normal distribution of data. We calculated the frequency and percentage of qualitative variables. Data are reported as percentages or mean ± SD unless otherwise specified. We analyzed differences between distributions of categorical variables with the Fisher’s exact test. We did not choose any preselected cut-off points for defining subgroups of risk. Instead, the range of values for each variable were based on the distribution of each variable in the full dataset. We performed a univariate analysis of each variable as a predictor of outcome. We determined the overall significance for each independent association between the variable and the ICU outcome. Once we had determined the ICU mortality associated with each subgroup, we identified the variables that could be included in the score based on our predefined rules, the AUC, and the pre-specified p-values. We performed logistic regression analyses to test risk variables in relation to ICU mortality. With the probabilities obtained in the logistic model, we calculated the AUC, and estimated the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Test performance was defined as the performance of models from the bootstrap samples when applied to the original sample. The difference between the bootstrap performance and the test performance (“optimism”) was averaged over 200 replications (25,26). For the final SPIRES score, we considered the minimum number of variables that provided similar performance as the full 28-predictor model. We emphasize that to adjust for optimism, we internally validated the performance of the score as the apparent performance of the minimum-variables-predictor model minus the optimism calculated for the 28-predictor model. We performed internal-external validation in independent parts of the data by leaving each of the three parent cohorts out once (20,21). Each patient was given a score based on risk variables. We aggregated patients with similar scores into major prognostic categories and classes.

We examined whether the survival time for each class could be predicted beyond the first week of ICU treatment. We analyzed the probability of ICU survival for the major prognostic classes using the Kaplan-Meier method with the log-rank test. We used the R Core Team 2019 software (R version 3.6.1) (https://www.r-project.org) for statistical computing (R Foundation for Statistical Computing, Vienna, Austria). The R Code to perform the analyses is available in a separate Supporting Material File. For all comparisons, a two-sided significance level of p-value <0.005 was considered a real effect size (17).

**SUPPLEMENTARY RESULTS**

Although the values of the selected 28 variables identified patients with a wide range of ICU mortality risk, only patient’s age, and APACHE II score, SOFA score, FiO2, PaO2/FiO2, Pplat, driving pressure, and number of extra-pulmonary OF at 24 h had an AUC≥0.65 and clear potential for being early predictors of ICU outcome (**table S1**). For the purpose of this study, APACHE II, SOFA score, FiO2 and driving pressure were intentionally not included as potential predictors in our final model for the following reasons. Although APACHE II score was available, it is cumbersome and not routinely collected at the bedside or during trial enrollment decisions, it requires numerous data elements, and relies on laboratory data that are not uniformly collected. We valued the number of extrapulmonary organ failures at 24 h over the SOFA scores at 24 h, since the 95% AUC confidence intervals were better (0.690-0.753 vs 0.682-0.747, respectively). FiO2 is a component of PaO2/FiO2 ratio and PaO2/FiO2 had better accuracy for ICU death. Pplat is a component of driving pressure and Pplat at 24 h had better accuracy for ICU death (8).

Thus, based on our own internal analysis (**Table S1**), we selected four variables (patient’s age, PaO2/FiO2 and Pplat recorded at 24 h under standardized ventilator settings, and number of extrapulmonary organ failure (OF) at 24 h of moderate/severe ARDS diagnosis) for characterizing our SPIRES score. For selecting the thresholds for those four variables (**Table S2**), we first assessed the overall ICU all-cause mortality by decades of patient’s age. Then, we examined the association of ICU mortality with the values of PaO2/FiO2 ratios and Pplat measured at 24 h of ARDS onset under standardized ventilator settings. PaO2/FiO2 values were grouped by increments of 50 mmHg; Pplat was analyzed from <20 to >34 cmH2O by increments of 1 cmH2O and then grouped depending on the magnitude or weights of the associated mortality rates. Then, we examined the overall ICU mortality associated with the number of extrapulmonary OF at 24 h of ARDS diagnosis. We chose the threshold value by creating an interval that separates patients into subgroups with clinically distinct ICU mortality rates with respect to the entire study population, independent of the underlying disease and the cause of death. Developing the SPIRES score based on individual values of those four variables entails three issues: (i) rating the selected thresholds in ordinal range categories, akin to how clinicians routinely categorize patients into risk groups by each variable; (ii) grouping patients with a similar outcome of interest (ICU death); (iii) determining whether the categories produced by all possible combinations of thresholds are meaningful. We assessed prediction of ARDS severity and ICU outcome based on the magnitude or weighs of the associated mortality rates in each interval of severity (as categorized with cut-points) for patient’s age (<50, 50-70, >70 years old), PaO2/FiO2 at 24 h of treatment (≤100, 101-200, >200 mmHg), Pplat at 24 h (<29, 29-29, >30 cmH2O), and number of extrapulmonary OF at 24 h of ARDS diagnosis (<2, 2, >2 organs). (**Table S2**).

Internal validation of the full 28-predictor model suggested minor statistical optimism in the score (**Table S3, Table S4**). The 4-variable model accomplished a similar performance as the larger model (AUC 0.860, 95%CI: 0.836-0.884). The optimism for the SPIRES score was 0.005, if considered as a pre-specified prediction model. The internally validated performance for the SPIRES score was calculated as the apparent performance of the 4-predictor model (0.860) minus the optimism calculated for the full predictor model (0.027), resulting in an AUC of 0.833 (95%CI: 0.810-0.855). Internal-external validation by leaving each of the three parent cohorts out once provided an average AUC of 0.860 (95%CI: 0.831-0.890) (**Table S5**).

Individual SPIRES scores achieved strong prediction for ICU death as reflected by an AUC of 0.860 (95%CI: 0.836-0.884), markedly superior than predictions achieved by PaO2/FiO2 at ARDS diagnosis (AUC 0.555, 95%CI: 0.518-0.592), by APACHE II (AUC 0.642, 95%CI: 0.607-0.677), by SOFA scores (AUC 0.681, 95%CI: 0.647-0.715), or by PaO2/FiO2 at 24 h (AUC 0.690, 95%CI: 0.656-0.724) (**Table S6**).

External validation cohort provided a very good confirmation of the performance of the SPIRES score (AUC 0.870, 95%CI 0.829-0.911) (**Table S9**).

Although the year of patient’s admission into the ICU was not accounted in the modelling of the SPIRES score, the internal-external validation confirmed that the score was accurate for the three time-periods of the derivation population. Of note, ICU mortality across the four cohorts (including the new external validation dataset) was similar (123/300, 41%; 114/300, 38%; 138/400, 35%; 111/301, 37%) and not significantly different (p=0.37).

**SUPPLEMENTARY DISCUSSION**

We aimed to develop a simple, integrative, and more systematic approach to ARDS stratification based on the values of commonly collected variables that are representative of the pulmonary and systemic status during the first 24 h after ARDS diagnosis that could be helpful for selecting candidate patients for enrollment into therapeutic RCTs. The SPIRES score confirmed that age, degree of hypoxemia, ventilating pressures, and systemic organ dysfunction are very important variables in the prognosis of ARDS patients: outcome is worse with increasing age (27), patients with severe lung damage have lower PaO2/FiO2 (28), there is a direct relationship between Pplat and mortality (29), and the greater the number of extrapulmonary OF the higher the mortality (30).

None of the three new ARDS classes paralleled current existing categories of ARDS. Clearly, our findings illustrate that ARDS cannot be viewed as a homogeneous disorder, and that the SPIRES score provides prognostic information that seems to be more important than the clinical diagnosis of moderate or severe ARDS for predicting ICU outcome.

Despite the pathobiological relevance of identifying biomarkers for targeting treatment and predicting outcome in future ARDS trials, we would like to emphasize the following issues as serious limitations for a two-class inflammatory model or ARDS phenotypes (hypoinflammatory and hyperinflammatory) reported by several authors (31-34). First, in those retrospective analysis of randomized controlled trials (31-33) and in a prospective observational study (34) used for describing the two-class inflammatory model, blood samples for biomarker measurements were collected within 48-72 h of meeting ARDS criteria. A recent prospective observational study in 104 ARDS patients from several etiologies admitted into a single-center ICU during a 7-year period, assessed the prognostic information of these two phenotypes/endotypes (34). Although the authors identified the two phenotypes in this small population, no significant differences in 30-day mortality between the two classes were found. Second, in all those studies, patients with mild ARDS were included, but no information of outcome by lung severity was ever reported. In our study, patients were enrolled at the time of meeting criteria for moderate/severe ARDS and assessed at 24 h under a standardized ventilatory approach. Third, since ARDS is a syndrome that shares the same molecular pathways of sepsis or severe pneumonia (more than 80% of patients in the studies used for developing the hyper/hypoinflammatory classes had sepsis or pneumonia), the hyper-inflammatory class could be just a reflection of the host response to the infectious cause of ARDS that resulted into this classification. Forth, since the pre-ARDS inflammatory state of patients classified as having hyper-inflammatory class was never reported or examined, it is plausible that most patients included in this class had higher baseline levels of inflammatory mediators due to several other comorbid conditions. Fifth, in the simvastatin trial (35), compliance with the use of 6-8 ml/kg PBW recommendation for tidal volume ranged from 20% to 39% of patients across all time points (36), despite that the trial protocol mandated the use of lung-protective ventilation and investigators from all participating 40 centers from UK and Ireland were encouraged to use low tidal volume. Since considerably less than half of the patients enrolled in that trial received lung-protective ventilation, it is plausible that the inflammatory response associated with the development of ventilator-induced lung injury (37) was part of the hyper-inflammatory signature class, making the results of those analyses very questionable. We are patiently waiting further data from prospective RCTs before adopting those phenotypes to alter clinical management of ARDS. Sixth, it is of particular importance that the vast majority of biomarker studies in critical care, even those based on physiologic mechanisms and reasoning, demonstrated no mortality benefit when studied in a robust manner (38).

**TABLE S1. Univariate logistic regression of clinically relevant variables in 1,000 patients with moderate-to-severe acute respiratory distress syndrome (ARDS).**

|  |  |
| --- | --- |
| Variables | Univariate logistic regression |
| N | Beta | SE | OR | 95% CI | p-value | AUC (95%CI) |
| Age  | 1000 | 0.03 | 0.005 | 1.03 | 1.02-1.04 | 1.1e-13 | **0.65 (0.61-0.68)** |
| Gender  | 1000 | -0.04 | 0.14 | 0.96 | 0.73-1.26 | 0.78 | 0.50 (0.47-0.53) |
| APACHE II - D0  | 1000 | 0.08 | 0.01 | 1.08 | 1.06-1.1 | 2.7e-13 | 0.64 (0.61-0.68) |
| APACHE II - D1  | 967 | 0.11 | 0.01 | 1.11 | 1.09-1.14 | 3.7e-22 | 0.70 (0.67-0.74) |
| SOFA at T0  | 1000 | 0.2 | 0.02 | 1.22 | 1.17-1.28 | 4.2e-21 | 0.68 (0.65-0.71) |
| SOFA at T24  | 1000 | 0.23 | 0.02 | 1.26 | 1.21-1.31 | 2.2e-27 | 0.715 (0.68-0.75) |
| VT at T0  | 1000 | 0.04 | 0.06 | 1.04 | 0.92-1.18 | 0.51 | 0.50 (0.46-0.54) |
| VT at T24  | 1000 | -0.11 | 0.07 | 0.9 | 0.78-1.04 | 0.14 | 0.53 (0.49-0.57) |
| FiO2 at T0  | 1000 | 0.42 | 0.34 | 1.52 | 0.78-3 | 0.22 | 0.52 (0.49-0.56) |
| FiO2 at T24  | 1000 | 3.22 | 0.38 | 25.03 | 12.04-52.75 | 1.2e-17 | 0.65 (0.62-0.69) |
| Respiratory rate at T0  | 1000 | 0.01 | 0.01 | 1.01 | 0.99-1.04 | 0.35 | 0.51 (0.47-0.54) |
| Respiratory rate at T24  | 1000 | 0.05 | 0.01 | 1.05 | 1.02-1.08 | 0.0001 | 0.57 (0.53-0.61) |
| PEEP at T0  | 1000 | -0.01 | 0.02 | 0.99 | 0.96-1.03 | 0.72 | 0.50 (0.46-0.54) |
| PEEP at T24  | 1000 | 0.05 | 0.02 | 1.05 | 1-1.09 | 0.03 | 0.55 (0.51-0.58) |
| Plateau pressure at T0  | 985 | 0.05 | 0.01 | 1.05 | 1.02-1.08 | 0.0002 | 0.57 (0.53-0.60) |
| Plateau pressure at T24 | 1000 | 0.2 | 0.02 | 1.22 | 1.18-1.26 | 2.6e-28 | **0.73 (0.70-0.77)** |
| Driving pressure at T0 | 985 | 0.05 | 0.013 | 1.05 | 1.03-1.08 | 0.00014 | 0.57 (0.53-0.60) |
| Driving pressure at T24 | 1000 | 0.17 | 0.017 | 1.18 | 1.14-1.22 | 2.2e-22 | 0.69 (0.66-0.72) |
| PaO2 at T0 | 1000 | -0.01 | 0.003 | 0.99 | 0.99-1 | 0.0056 | 0.56 (0.52-0.59) |
| PaO2 at T24 | 1000 | -0.02 | 0.003 | 0.98 | 0.98-0.99 | 8.1e-10 | 0.63 (0.60-0.67) |
| PaO2/FiO2 at T0  | 1000 | 0 | 0.002 | 1 | 0.99-1 | 0.0063 | 0.55 (0.52-0.59) |
| PaO2/FiO2 at T24  | 1000 | -0.01 | 0.001 | 0.99 | 0.98-0.99 | 7.2e-20 | **0.69 (0.66-0.72)** |
| PaCO2 at T0  | 1000 | 0.01 | 0.01 | 1.01 | 1-1.02 | 0.011 | 0.54 (0.51-0.58) |
| PaCO2 at T24  | 1000 | 0.03 | 0.01 | 1.03 | 1.02-1.05 | 1.7e-07 | 0.59 (0.55-0.63) |
| pH at T0  | 1000 | -2.31 | 0.6 | 0.1 | 0.03-0.32 | 0.0001 | 0.56 (0.53-0.60) |
| pH at T24  | 1000 | -6.22 | 0.76 | 0 | 0-0.01 | 3.2e-16 | 0.64 (0.61-0.68) |
| Number OF at T0 | 1000 | 0.73 | 0.07 | 2.08 | 1.82-2.39 | 6.9e-26 | 0.70 (0.66-0.73) |
| Number OF at T24  | 1000 | 0.82 | 0.07 | 2.27 | 1.99-2.61 | 7.7e-32 | **0.72 (0.69-0.75)** |
| Minute ventilation, T0  | 1000 | 0.01 | 0.03 | 1.01 | 0.95-1.07 | 0.83 | 0.49 (0.46-0.53) |
| Minute ventilation, T24  | 1000 | 0.05 | 0.03 | 1.05 | 0.99-1.11 | 0.12 | 0.53 (0.49-0.56) |

*APACHE II (acute physiology and chronic health evaluation II) score; AUC (area under the receiving operating characteristics curve); CI (confidence intervals); OF (extrapulmonary organ failure); OR (odds ratio); PEEP (positive end-expiratory pressure); SE (standard error); SOFA (sequential organ failure assessment) scale; VT (tidal volume); T0 (at ARDS diagnosis); T24 (at 24 h after ARDS diagnosis).*

**TABLE S2.** Distribution of 1,000 patients with moderate-to-severe acute respiratory distress syndrome (ARDS), based on thresholds for patient’s age, number of extrapulmonary organ failure (OF) at 24 h of ARDS onset, and PaO2/FiO2 and plateau pressure (Pplat) assessed under standardized ventilatory settings at 24 h after ARDS onset. Percentages have been rounded to zero decimals.

|  |  |  |  |
| --- | --- | --- | --- |
| Patient’s age, years | N | No. ICU deaths (%) | ICU mortality after grouping |
| 18-29 | 58 | 12 (21) | 74/327 (23%) |
| 30-39 | 99 | 19 (19) |
| 40-49 | 170 | 43 (25) |
| 50-60 | 216 | 73 (34) | 161/417 (39%) |
| 60-70 | 201 | 88 (44) |
| 71-80 | 213 | 116 (55) | 140/256 (55%) |
| >80 | 43 | 24 (56) |
| Total | 1000 | 375 (38) |  |

|  |  |  |  |
| --- | --- | --- | --- |
| No. extrapulmonary OF | N | No. ICU deaths (%) | ICU mortality after grouping |
| 0 | 111 | 13 (12) | 111/498 (22%) |
| 1 | 387 | 98 (25) |
| 2 | 272 | 103 (38) | 103/272 (38%) |
| 3 | 149 | 95 (64) | 161/230 (70%) |
| 4 | 66 | 52 (79) |
| 5 | 15 | 14 (93) |

|  |  |  |  |
| --- | --- | --- | --- |
| PaO2/FiO2 at 24 h, mmHg | N | No. ICU deaths (%) | ICU mortality after grouping |
| ≤50 | 10 | 7 (70) | 142/218 (65%) |
| 51-100 | 208 | 135 (65) |
| 101-150 | 361 | 133 (37) | 203/611 (33%) |
| 151-200 | 250 | 70 (28) |
| 201-250 | 107 | 22 (21) | 30/171 (18%) |
| 251-300 | 45 | 7 (16) |
| >300 | 19 | 1 (5) |

|  |  |  |  |
| --- | --- | --- | --- |
| Pplat at 24 h, cmH2O | N | No. ICU deaths (%) | ICU mortality after grouping |
| <20 | 67 | 12 (18) | 143/635 (23%) |
| 20 | 43 | 12 (28) |
| 21 | 30 | 6 (20) |
| 22 | 43 | 10 (23) |
| 23 | 49 | 9 (18) |
| 24 | 68 | 17 (25) |
| 25 | 86 | 19 (22) |
| 26 | 70 | 12 (17) |
| 27 | 85 | 20 (24) |
| 28 | 94 | 26 (28) |
| 29 | 64 | 25 (39) | 76/161 (47%) |
| 30 | 97 | 51 (53) |
| 31 | 48 | 35 (73) | 156/204 (77%) |
| 32 | 65 | 47 (72) |
| 33 | 31 | 26 (84) |
| 34 | 19 | 14 (74) |
| >34 | 41 | 34 (83) |

**TABLE S3. Apparent performance of the scoring model in 1,000 patients with moderate-to-severe ARDS using logistic regression analysis and bootstrapping. Model 4 achieved the best apparent performance.** Data are expressed as mean values of logistic coefficients. ARDS: acute respiratory distress syndrome, *AUC: area under the receiving operating characteristic curve, CI: confidence intervals, NPV: negative predictive value, OF: extrapulmonary organ failures included in the sequential organ failure assessment (SOFA) scale, OR: odds ratio, P/F: PaO2/FiO2, Pplat: plateau pressure, PPV: positive predictive value, SE: standard error, T24: at 24 h after ARDS diagnosis.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Variable | b | SE | OR | 95% CI | *p* | OR Bootstrapping | 95% CI Bootstrapping |
| **1** Age | intercept | 1.23 | 0.13 | 0.29 | 0.22 - 0.38 | <0.001 | 0.29 | 0.22 - 0.38 |
| Age <50 |  |  | 1 (ref) |  |  | 1 (ref) |  |
| Age 50-70 | 0.77 | 0.16 | 2.16 | 1.57 – 3.0 | <0.001 | 2.17 | 1.59 - 3.05 |
| Age >70 | 1.46 | 0.19 | 4.30 | 3.0 - 6.21 | <0.001 | 4.30 | 2.97 - 6.20 |
| **2** Age + PaO2/FiO2 at T24 | intercept | 2.32 | 0.24 | 0.10 | 0.06 - 0.15 | <0.001 | 0.099 | 0.06 - 0.15 |
| Age <50 |  |  | 1 (ref) |  |  | 1 (ref) |  |
| Age 50-70 | 0.82 | 0.17 | 2.27 | 1.62 - 3.22 | <0.001 | 2.31 | 1.64 - 3.40 |
| Age >70 | 1.59 | 0.20 | 4.91 | 3.34 - 7.29 | <0.001 | 4.95 | 3.30 - 7.51 |
| P/F >200 |  |  | 1 (ref) |  |  | 1 (ref) |  |
| P/F 101-200 | 0.81 | 0.22 | 2.26 | 1.47 - 3.56 | <0.001 | 2.25 | 1.50 - 3.45 |
| P/F≤100 | 2.26 | 0.26 | 9.56 | 5.87 - 15.98 | <0.001 | 9.44 | 5.74 - 15.52 |
| **3**Age + PaO2/FiO2 at T24 + Pplat at T24 | intercept | 2.95 | 0.27 | 0.05 | 0.03 - 0.09 | <0.001 | 0.054 | 0.03 - 0.09 |
| Age <50 |  |  | 1 (ref) |  |  | 1 (ref) |  |
| Age 50-70 | 1.01 | 0.19 | 2.74 | 1.88 - 4.05 | <0.001 | 2.73 | 1.86 - 4.05 |
| Age >70 | 1.81 | 0.22 | 6.09 | 3.98 - 9.46 | <0.001 | 6.06 | 3.99 - 9.56 |
| P/F >200 |  |  | 1 (ref) |  |  | 1 (ref) |  |
| P/F 101-200 | 0.69 | 0.24 | 1.99 | 1.26 - 3.23 | 0.004 | 1.98 | 1.26 - 3.16 |
| P/F≤100 | 1.70 | 0.27 | 5.45 | 3.21 - 9.43 | <0.001 | 5.28 | 3.18 - 9.26 |
| Pplat <29 |  |  | 1 (ref) |  |  | 1 (ref) |  |
| Pplat 29-30 | 1.09 | 0.20 | 2.96 | 1.99 - 4.42 | <0.001 | 2.95 | 1.97 - 4.48 |
| Pplat >30 | 2.30 | 0.21 | 10.0 | 6.68 - 15.21 | <0.001 | 9.95 | 6.84 - 15.55 |
| **4**Age + PaO2/FiO2 at T24 + Pplat at T24 + OF at T24 | intercept | 3.77 | 0.32 | 0.02 | 0.01 - 0.04 | <0.001 | 0.024 | 0.01 - 0.04 |
| Age <50 |  |  | 1 (ref) |  |  | 1 (ref) |  |
| Age 50-70 | 1.07 | 0.21 | 2.92 | 1.94 - 4.45 | <0.001 | 2.91 | 1.97 - 4.57 |
| Age >70 | 2.0 | 0.24 | 7.38 | 4.64 - 11.92 | <0.001 | 7.19 | 4.44 - 12.31 |
| P/F >200 |  |  | 1 (ref) |  |  | 1 (ref) |  |
| P/F 101-200 | 0.64 | 0.26 | 1.90 | 1.16 - 3.19 | 0.012 | 1.90 | 1.19 - 3.38 |
| P/F≤100 | 1.46 | 0.29 | 4.30 | 2.43 - 7.76 | <0.001 | 4.21 | 2.45 – 8.0 |
| Pplat <29 |  |  | 1 (ref) |  |  | 1 (ref) |  |
| Pplat 29-30 | 1.20 | 0.22 | 3.33 | 2.16 - 5.15 | <0.001 | 3.27 | 2.19 - 5.14 |
| Pplat >30 | 2.40 | 0.22 | 11.0 | 7.13 - 17.28 | <0.001 | 10.78 | 7.00 - 17.35 |
| OF <2 |  |  | 1 (ref) |  |  | 1 (ref) |  |
| OF 2 | 0.80 | 0.20 | 2.22 | 1.51 - 3.27 | <0.001 | 2.18 | 1.51 - 3.26 |
| OF >2 | 2.19 | 0.22 | 8.92 | 5.87 - 13.73 | <0.001 | 8.67 | 5.74 - 13.53 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | AUC | 95% CI | cutoffs | Sensitivity | Specificity | PPV | NPV |
| 1 | 0.641 | 0.608 – 0.673 | 0.31 | 0.80 | 0.40 | 0.45 | 0.77 |
| 2 | 0.730 | 0.641 – 0.730 | 0.41 | 0.59 | 0.76 | 0.59 | 0.75 |
| 3 | 0.811 | 0.783 – 0.839 | 0.41 | 0.66 | 0.81 | 0.68 | 0.80 |
| 4 | 0.860 | 0.836 – 0.884 | 0.33 | 0.79 | 0.79 | 0.69 | 0.86 |

**TABLE S4. Steps for internal validation of SPIRES score: apparent performance, bootstrap performance, test performance, optimism, and optimism-corrected performance.** Optimism was calculated as the difference between the test performance and bootstrap performance, and repeated 200 times for a stable estimate. AUC: area under the receiving operating characteristic curve; CI: confidence interval.

|  |  |
| --- | --- |
| Model | **Validation steps**,*mean AUC (95%CI)* |
| Apparent performance | Bootstrap performance (mean) | Test performance (mean) | Optimism (mean)(n=200 times) | Optimism-corrected performance |
| 28-variable prognostic model (full model) | 0.865(0.842-0.888) | 0.885 (0.883-0.887) | 0.858 (0.857–0.858) | 0.027 (0.025–0.029) | 0.838 (0.816–0.859) |
| 4-variable prognostic model (SPIRES score) | 0.860 (0.836-0.884) | 0.863 (0.861-0.864) | 0.857 (0.857-0.858) | 0.005 \*(0.004-0.007) | 0.833 \*\*(0.811-0.856) |

\*this estimate of optimism ignores the uncertainty in selection of the 4-variable model, and is hence an underestimate.

\*\*calculated with the optimism-correction of the 28-variable model.

**TABLE S5. Internal-external validation analysis of the SPIRES score** (see text for details), representing the performance in a cohort of a model developed in the other two cohorts.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | N | Events | AUC | 95% CI | Sensitivity | Specificity | PPV | NPV |
| All patients | 1000 | 375 | 0.860 | 0.836 – 0.884 | 0.79 | 0.79 | 0.69 | 0.86 |
| Test on ALIEN cohort | 300 | 123 | 0.864 | 0.836 – 0.893 | 0.79 | 0.80 | 0.69 | 0.87 |
| Test on STANDARDS | 300 | 114 | 0.868 | 0.839 – 0.896 | 0.83 | 0.78 | 0.69 | 0.89 |
| Test on STANDARDS 2 | 400 | 138 | 0.849 | 0.818 – 0.881 | 0.74 | 0.79 | 0.70 | 0.83 |

**TABLE S6. Performance of the SPIRES score and other models in 1,000 patients with moderate-to-severe ARDS using logistic regression analysis and the area under the receiver operating characteristic curves.** Data are expressed as mean values of logistic coefficients (see text for details). *AUC: area under the receiving operating characteristic curve, CI: confidence intervals, NPV: negative predictive value, SOFA: sequential organ failure assessment scale, OR: odds ratio, PPV: positive predictive value, ROC: receiving operating characteristics; T24: at 24 h after ARDS onset assessed under a standardized ventilator setting.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | N | AUC | 95% CI | Sensitivity | Specificity | PPV | NPV |
| PaO2/FiO2 at ARDS onset | 1000 | 0.555 | 0.518 – 0.592 | 0.57 | 0.53 | 0.42 | 0.67 |
| APACHE II at ARDS diagnosis | 1000 | 0.642 | 0.607 – 0.677 | 0.54 | 0.69 | 0.51 | 0.71 |
| SOFA score at ARDS diagnosis | 1000 | 0.681 | 0.647 – 0.715 | 0.58 | 0.69 | 0.53 | 0.74 |
| PaO2/FiO2 at T24 | 1000 | 0.690 | 0.656 – 0.724 | 0.67 | 0.62 | 0.52 | 0.76 |
| SPIRES Score | 1000 | 0.860 | 0.836 – 0.884 | 0.79 | 0.79 | 0.69 | 0.86 |

**TABLE S7. External validation cohort.** Characteristics of 301 patients with moderate-to-severe acute respiratory distress syndrome (ARDS) based on thresholds for patient’s age, number of extrapulmonary organ failures (included in SOFA scale) at 24 h of ARDS diagnosis, PaO2/FiO2 ratio and end-inspiratory plateau pressure measured at 24 h after ARDS diagnosis under standardized ventilator settings. Percentages have been rounded to zero decimals.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables** | **Thresholds** | **Number of patients** | **Number of deaths** |
| Patient’s age (*years)* | < 50 | 81 | 18 (22%) |
| 50 - 70 | 158 | 60 (38%) |
| > 70 | 62 | 33 (53%) |
|  |  |  |  |
| Number of extrapulmonary organ failures at 24 h of ARDS diagnosis | < 2 | 125 | 19 (15%) |
| 2 | 83 | 28 (34%) |
| > 2 | 93 | 64 (69%) |
|  |  |  |  |
| PaO2/FiO2 at 24 h on standardized ventilatory settings(*mmHg*) | > 200 | 79 | 14 (18%) |
| 101 - 200 | 167 | 56 (34%) |
| ≤ 100 | 55 | 41 (75%) |
|  |  |  |  |
| Plateau pressure at 24 h on standardized settings(*cmH2O*) | < 29 | 227 | 54 (24%) |
| 29 - 30 | 43 | 29 (67%) |
| > 30 | 31 | 28 (90%) |
|  |  |  |
| **Total** | **301** | **111 (37%)** |

**TABLE S8.** Overall all-cause mortality in the intensive care unit (ICU) in relation to the total SPIRES score in the external validation cohort of 301 patients with moderate-to-severe acute respiratory distress syndrome (ARDS). Mortality increased with each increment of the total score for the 9 scoring groups clustered in three classes of ARDS. CI: confidence intervals. Percentages have been rounded to zero decimals.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Total score** | **No. patients** | **ICU deaths** | **ICU mortality** (%) | **Classes** | **ICU mortality** **N (%, 95%CI)** |
| 4 | 7 | 0 | 0 | Class 1 | 24/190 (13, 8 to 18) |
| 5 | 52 | 1 | 2 |
| 6 | 66 | 10 | 15 |
| 7 | 65 | 13 | 20 |
| 8 | 49 | 34 | 69 | Class 2 | 34/49 (69, 54 to 81) |
| 9 | 27 | 22 | 81 | Class 3 | 53/62 (85, 74 to 93) |
| 10 | 25 | 21 | 84 |
| 11 | 7 | 7 | 100 |
| 12 | 3 | 3 | 100 |
|  | 301 | 111 | 37 |  |  |

**TABLE S9. External validation of the SPIRES score** (see text for details).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Cohort | N | Events | AUC | 95% CI | Sensitivity | Specificity | PPV | NPV |
| Development cohort | 1000 | 375 | 0.860 | 0.836 – 0.884 | 0.79 | 0.79 | 0.69 | 0.86 |
| External validation cohort | 301 | 111 | 0.870 | 0.830 – 0.911 | 0.78 | 0.83 | 0.73 | 0.87 |

**FIGURE S1. Timeline for prediction.** At the patient’s time of eligibility (i.e. when they develop moderate or severe ARDS), the patient’s risk of future outcome at ICU discharge (alive or death) was predicted using the first 24 hours of data.



**FIGURE S2. Steps for the SPIRES prediction model development in patients with moderate-to-severe ARDS.** We performed logistic regression analysis, bootstrapping, and internal-external validation. For internal validation we used all 1,000 patients. We performed external validation by leaving each of the three parent cohorts out once (see text for details). External validation was tested in an independent cohort of 301 recently treated patients. ALIEN: Acute Lung Injury: Epidemiology and Natural history; STANDARDS: Stratification and outcome in ARDS; STANDARDS-2: Stratification and outcome in ARDS – 2.



**FIGURE S3. Observed probability vs. predicted probability across deciles of prediction of ICU mortality by the SPIRES score.**



**SUPPLEMENTARY REFERENCES**

1. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA* 2013; 310:2191-2194
2. Ashbaugh DG, Bigelow DB, Petty TL, Levine BE: Acute respiratory distress in adults. *Lancet* 1967; 2(7511):319-323
3. ARDS Definition Task Force, Ranieri VM, Rubenfeld GD, Thompson BT, et al: Acute respiratory distress syndrome: the Berlin Definition. *JAMA* 2012; 307:2526-2533
4. Villar J, Pérez-Méndez L, Kacmarek RM: The Berlin definition me tour needs: no. *Intensive Care Med* 2016; 42:648-650
5. Villar J, Pérez-Méndez L, López J, et al: An early PEEP/FIO2 trial identifies different degrees of lung injury in patients with acute respiratory distress syndrome. *Am J Respir Crit Care Med* 2007; 176:795-804
6. Villar J, Blanco J, Añón JM, et al: 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
7. Villar J, Ambrós A, Soler JA, et al: Age, PaO2/FiO2, and plateau pressure score: a proposal for a simple outcome score in patients with the acute respiratory distress syndrome. *Crit Care Med* 2016; 44:1361-1369.
8. Villar J, Martín-Rodríguez C, Domínguez-Berrot AM, et al: A quantile analysis of plateau and driving pressures: Effects on mortality in patients with acute respiratory distress syndrome receiving lung-protective ventilation. *Crit Care Med* 2017; 45:843-850.
9. Villar J, Ambrós A, Mosteiro F, et al: A prognostic enrichment strategy for selection of patients with acute respiratory distress syndrome in clinical trials. *Crit Care Med* 2019; 47:377-385
10. Bernard GR, Artigas A, Brigham KL, et al: The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. *Am J Respir Crit Care Med* 1994; 149:818-824
11. 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
12. Knaus WA, Draper EA, Wagner DP, Zimmerman JE: APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13:818-829
13. Vincent JL, de Mendonça A, Cantraine F, el 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
14. Eke G, Bloos F, Wilson DC, Meybohm P; SepNet Critical Care Trials Group: Identification of developing multiple organ failure in sepsis patients with low or moderate SOFA scores. *Crit Care* 2018; 22:147
15. Singer M, Deutschman CS, Seymour CW, et al: The Third International Consensus Definition for Sepsis and Septic Shock. *JAMA* 2016; 315:801-810
16. Levy MM, Fink MP, Marshall JC, et al: 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med* 2003; 31:1250-1256
17. Ioannidis JPA: The proposal to lower P value thresholds to 0.005. *JAMA* 2018; 319:1429-1430
18. Tonelli AR, Zein J, Adams J, Ioannidis JPA: Effects of interventions on survival in acute respiratory distress syndrome: an umbrella review of 159 published randomized trials and 29 meta-analyses. *Intensive Care Med* 2014; 40:769-787
19. Royston P, Parmar MKB, Sylvester R: Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer. *Statist Med* 2004; 23:907-926
20. Steyerberg EW, Harrel FE: Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 2016; 69:245-247.
21. Steyerberg EW, Harrel FE, Borsboom GJJM, Eijkemans MJCR, Vergouwe Y, Habbema JDF: Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2011; 54:774-781
22. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG: Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol* 2003; 56:441-447
23. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF: Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005; 58:475-483
24. Leisman DE, Harhay MO, Lederer DJ, Abramson MA, Adjei AA, Bakker J, 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
25. Harrel Jr. FE: Regression Modeling Strategies. With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. In: Springer Series in Statistics, Springer International Publishing Switzerland, 2015, pp 103-126
26. Steyerberg EW: Clinical Prediction Models. A practical approach to Development, Validation, and Updating. In: Springer Series Statistics for Biology and Health, Springer Nature Switzerland AG, 2019, pp 83-100
27. Gee MH, Gottlieb JE, Albertine KH, Kubis JM, Peters SP, Fish JE: Physiology of aging related to outcome in the adult respiratory distress syndrome. *J Appl Physiol (1985)* 1990; 69:822-829
28. Villar J, Pérez-Méndez L, Blanco J, et al: A universal definition of ARDS: the PaO2/FiO2 ratio under a standard ventilatory setting -a prospective, multicenter validation study. *Intensive Care Med* 2013; 39:583-592
29. 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
30. Villar J, Martínez D, Mosteiro F, et al: Is overall mortality the right composite endpoint in clinical trials of acute respiratory distress syndrome? *Crit Care Med* 2018; 46:892-899
31. Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA; NHLBI ARDS Network: Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med* 2014; 2: 611-620
32. Famous KR, Delucchi K, Ware LB, et al: Acute respiratory distress syndrome subphenotypes respond differently to randomized fluid management strategy. *Am J Respir Crit Care Med* 2017; 195:331-338
33. Calfee CS, Delucchi KL, Sinha P, et al: Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomized controlled trial. *Lancet Respir Med* 2018; 6:691-698
34. Kitsios GD, Yang L, Manatakis D, et al: Host-response subphenotypes offer prognostic enrichment in patients with or at risk for acute respiratory distress syndrome. *Crit Care Med* 2019; 47:1724-1734
35. McAuley DF, Laffey JG, O’Kane CM, et al: Simvastatin in the acute respiratory distress syndrome. *N Engl J Med* 2014; 371:1695-1703
36. Poole J, McDowell C, Lall R, et al: Individual patient data analysis of tidal volumes used in three large randomized control trials involving patients with acute respiratory distress syndrome. Brit J Anaesth 2017; 118:570-575
37. Villar J, Blanco J, Zhang H, Slutsky AS: Ventilator-induced lung injury and sepsis: two sides of the same coin? *Minerva Anestesiol* 2011; 77:647-653
38. Santacruz CA, Pereira AJ, Celis E, Vincent JL: Which multicenter randomized controlled trials in critical care medicine have shown reduced mortality? A systematic review. *Crit Care Med* 2019; 47:1680-1691