Development and validation of a biomarker-based model to predict persistent hypoxemic respiratory failure among mechanically ventilated adults

Neha Alhad Sathe

A thesis

Submitted in partial fulfillment of the

Requirements for the degree of

Master of Science

University of Washington

2021

Committee:

Alison E. Fohner (Chair)

Mark M. Wurfel

Program Authorized to Offer Degree:

Epidemiology

University of Washington

Abstract

Development and validation of a biomarker-based model to predict persistent hypoxemic respiratory failure among mechanically ventilated adults

Neha Alhad Sathe

Chair of the Supervisory Committee:

Dr. Alison E. Fohner, Assistant Professor

Department of Epidemiology

Background. Acute hypoxemic respiratory failure (HRF) requiring invasive mechanical ventilation (IMV) is a common and morbid condition, but treatment is largely supportive and most pharmacologic trials have been negative. The early clinical trajectory is heterogeneous, with most patients resolving within days while others require prolonged IMV, which impairs both clinical resource allocation and the detection of therapeutic effects in trials. Our objectives were to (1) compare mortality between those with persistent HRF at day 3 to those with resolving HRF, and (2) develop and validate a model to predict persistent HRF.

Methods. I performed a secondary analysis of patients with acute HRF in the intensive care unit (ICU), enrolled in 2 independent prospective cohorts (discovery cohort n = 632, validation cohort n = 906). I used generalized linear models to estimate the relative risk (RR) of persistent HRF compared to resolving HRF in each cohort, adjusting for age, sex and the acute physiology and chronic health evaluation score (APACHE-II). For development of the prediction model, I split the entire discovery cohort into training (n=474) and testing (n=158) sets, and used the second cohort for external validation. I applied LASSO to 33 candidate clinical predictors (spanning demographics, comorbidities, vitals, standard laboratory values, and illness severity scores) for parsimonious model selection. Then I examined whether a combination of log2 transformed research biomarkers previously predictive of ICU outcomes (interleukin-6 [IL-6], interleukin-8 [IL-8], angiopoietin-2 [Ang-2], soluble tumor necrosis factor receptor 1 [sTNFR-1],

measured in a subset) could improve model performance. I calculated area under the curve (AUC) for model performance.

Results. Approximately half of patients in the discovery and validation cohorts had persistent HRF at day 3 (298/632 in discovery and 514/906 in validation). At cohort enrollment, patients with persistent HRF had higher proportion of shock, pneumonia, as well as higher ICU illness severity scores. In both cohorts, persistent HRF was associated with an over 2-fold higher risk of death compared to resolving HRF (aRR 2.33, 95% CI 1.42, 3.82 in discovery; aRR 2.05, 95% CI 1.51, 2.78) after adjusting for age, sex, and a measure of baseline arterial oxygenation (P/F). In models adjusting for an overall illness severity score (APACHE-II) instead of P/F, this effect was attenuated in the discovery cohort but remained significant in the validation cohort (aRR 1.49, 95% CI 0.92, 2.41 in discovery; aRR 1.81, 95% CI 1.32, 2.44 in validation). A LASSO derived model to predict persistent HRF had moderate performance in the training (AUC 0.79, 95% CI 0.75, 0.83) and test (AUC 0.76, 95% CI 0.68, 0.83) cohorts, but had limited performance in external validation (AUC 0.69, 85% CI 0.65, 0.72). A model adding IL-6 and Ang-2 significantly improved performance across all 3 cohorts (training AUC 0.82, 95% CI 0.77, 0.86; test AUC 0.78, 95% CI 0.69, 0.87; validation AUC 0.72, 95% CI 0.65, 0.79).

Conclusions. Persistent HRF is associated with a high risk for mortality compared to resolving HRF even when adjusting for initial illness severity. A parsimonious model of P/F, APACHE-II, IL-6, and Ang-2 has moderate performance for predicting persistent HRF. This can be implemented to direct care to high-risk patients early, and to improve prognostic enrichment in trials for acute HRF.

BACKGROUND

Acute hypoxemic respiratory failure (HRF) requiring mechanical ventilation is associated with high morbidity, mortality, resource utilization and cost(1–4). Mortality ranges from 22-52% depending on underlying conditions and severity of hypoxemia(2–4). Nearly all patients mechanically ventilated for 1 week develop weakness and functional limitations that impair daily activities(5). Total hospital charges for adults with respiratory failure were estimated at $1 billion in 2002(6), and costs have increased dramatically since then(7). The ongoing coronavirus epidemic has only accentuated the disease burden of acute HRF(8, 9), yet treatment remains supportive.

Major barriers to identifying effective, targeted therapeutics in acute HRF is clinical and biologic heterogeneity across patients as well as mortality not attributable to respiratory failure (10, 11). In order to improve detection of therapeutic effect in acute HRF trials, enrollment should focus on reliably-defined sub-populations with a high likelihood of a disease related event or response to therapy(11). Defining subsets of patients with these key features, often called subphenotypes, within heterogeneous syndromes has led to the identification of important new therapeutics in conditions such as asthma, chronic obstructive pulmonary disease, and certain cancers (12–14). Previous efforts to subphenotype acute HRF have been focused within the acute respiratory distress syndrome (ARDS) (15, 16). However, ARDS is similarly heterogeneous in prognosis(15, 17) and biology(16, 18–22), has limited reliability in research(23–25), is poorly recognized by clinicians(26), and represents a small subset of all acute HRF patients(26). Identifying a subphenotype of acute HRF with high likelihood of a disease related event population may improve prognostic enrichment in future trials and expand who we target for investigational therapeutics.

In other heterogeneous conditions, defining subphenotypes by the clinical trajectory has helped identify patients at high risk for poor outcomes and can help determine the early biologic events that lead to subsequent clinical course (27, 28). Prior epidemiologic studies and secondary trial analyses suggest a large proportion of HRF patients improve within days of admission to an intensive care unit (ICU) while others go on to develop prolonged need for mechanical ventilation (15, 17, 29–31). Furthermore, patients with prolonged respiratory failure exhibit distinct biologic responses that can be identified in the first 4 days of illness (32–37). Overall, leveraging these distinct trajectories of acute HRF, characterizing the

differences in mortality, and developing approaches that use the high volume of clinical and biologic data to discriminate between them could help target both clinical care and trial enrollment to patients most likely to benefit.

The overall goal of this study was to define and characterize a novel subphenotype of acute HRF based on persistence of hypoxemic respiratory failure on ICU day 3, termed persistent HRF. The first specific aim was to determine the relative risk of mortality associated with persistent HRF, compared to resolving HRF. The second aim was to develop and validate a model to predict persistent HRF using clinical and biologic data obtained within one day of ICU admission.

METHODS

Discovery Cohort

The cohort included adults ≥ 18 years of age admitted to the ICU at Harborview Medical Center (Seattle, WA) between 2006 and 2010 (18, 19). Patients were enrolled within 24 hours of ICU admission if meeting criteria for the systemic inflammatory response syndrome (40). Exclusion criteria included admission for trauma; admission for intracranial hemorrhage; and severe immunosuppression; and active cancer diagnosis. The study was granted a waiver of consent and approved by the University of Washington Institutional Review Board (IRB).

External Validation Cohort

The external validation cohort included adults from the Validating Acute Lung Injury Biomarkers for Diagnosis (VALID) cohort study at Vanderbilt University Medical Center (Nashville, TN) (41). Patients were enrolled on the day following ICU admission from 2006-2020. Exclusion criteria included severe chronic lung disease on home supplemental oxygen; cardiac arrest prior to admission; and anticipated discharge from ICU on the day of enrollment. This study was approved by the IRB at Vanderbilt University Medical Center.

Study Definitions

I restricted analyses of the cohorts to adults in medical or surgical ICUs with acute HRF, defined here as new invasive mechanical ventilation and a ratio of arterial partial pressure of oxygen to fraction of inspired oxygen ratio (P/F) ≤ 300 on the day of study enrollment. The P/F ratio is a commonly used

measure of hypoxemia and HRF severity (42). If P/F was unavailable, I used a ratio of % oxygen saturation (measured by pulse oximeter) to the fraction of inspired oxygen ≤ 315 (43). In addition, I excluded patients with chronic respiratory disease prior to enrollment to further ensure that HRF was acute.

Patients with acute HRF were further classified as persistent or resolving. Persistent HRF was defined by need for invasive mechanical ventilation and P/F ≤ 300 or S/F ≤ 315 on day 3 following ICU admission. Patients with brief improvement before day 3 (e.g. patients extubated but reintubated, or P/F > 300 on day 2) were still classified as persistent HRF as long as they met criteria on day 3. Patients who were liberated from mechanical ventilation, who had P/F >300 or S/F > 315, or who were missing both P/F and S/F on day 3 were classified as resolving HRF. Of note, all measures of hypoxemia in this study were obtained as part of clinical care. When multiple values were available, the lowest (i.e. physiologically worst) values for each ICU day were used. Patients who died before day 3 were classified as persistent HRF if they were on mechanical ventilation with P/F ≤ 300 on the day of death, and resolving if otherwise.

Statistical Analysis

We singly imputed missing predictor data with clinical normal values, which is consistent with real-world clinical practice and other prediction models(40).

We first compared demographics, pre-existing chronic conditions, acute conditions on cohort enrollment, and outcomes by persistent and resolving HRF.

In order to estimate the relative risk of mortality associated with persistent HRF, compared to resolving HRF I used generalized linear models with Poisson distribution and robust standard errors (44). I *a priori* chose to adjust for age, sex, and either P/F or acute physiology and chronic health evaluation (APACHE-II) score. I did not include both P/F and APACHE-II in models to avoid over-adjustment, as APACHE-II includes a measure of oxygenation. In this analysis, patients who died before qualifying for persistent HRF (i.e. day 3) were excluded.

To develop a model to predict persistent HRF I randomly split the discovery cohort into training (75%) and test (25%) datasets. The VALID cohort provided external validation. I considered all clinical variables obtained by the day following ICU admission as candidate predictors for model development

(Table 1). I used the physiologically worst values available if multiple were collected. I evaluated predictors for quality (e.g. missingness, frequency of outliers, range) and for associations with persistent HRF. Variables with significant skew were log2 transformed to fit their distribution.

These candidate predictors were entered into a least absolute shrinkage and selection operator (LASSO) regression, a machine learning method that penalizes less predictive variables to prevent over-fitting and improve performance in external cohorts(38). I used 10-fold crossvalidation to calculate the mean standard predictive error, and then selected the parsimonious LASSO model within 1 standard deviation of that error. In order to assess model discrimination I calculated the area under the curve. To assess model calibration I calculated Hosmer-Lemeshow $\chi 2$ statistic.

In addition, I examined research biomarkers of lung injury as candidate predictors, which were measured in a subset of patients in each cohort. Since these are not typically available in clinical practice but have strong relationships with ICU outcomes, I determined whether one or a combination of these biomarkers could help improve the LASSO derived prediction model.

Analyses were conducted in STATA (version 17.0).

Biomarker Measurements

Plasma samples were collected at enrollment for (38, 41). For this study I focused on markers of inflammation (interleukin-6 [IL-6], interleukin-8 [IL-8], soluble tumor necrosis factor receptor-1 [sTNFR-1] and endothelial dysfunction/activation (angiopoietin-2 [Ang-2]), which were measured in both cohorts and have been associated with poor outcomes in HRF(45). All biomarkers were measured using electrochemiluminiscent immunoassays (MesoScale Discovery, Rockville, MD) in the discovery cohort. In the validation cohort, IL-6 and IL-8 were measured on the same platform, while sTNFR-1 and Ang-2 were measured by ELISA (R&D Systems, Minneapolis, MN).

RESULTS

Cohort description

Of 632 patients with acute HRF in the discovery cohort, 322 had resolving HRF at day 3, 298 had persistent HRF; of 12 patients died by day 3, 11 were classified as persistent HRF. The validation cohort

(n=906) had a slightly higher proportion of patients with persistent HRF (n=514); of 48 patients who died by day 3, all were classified as persistent HRF.

Comparing features of patients with persistent and resolving HRF in each cohort, I observed that age, sex and pre-existing chronic conditions were similar across groups (Table 2). Acute diagnoses of shock, sepsis, pneumonia, and ARDS were more common among those with persistent HRF in both cohorts. Patients with persistent HRF also had lower P/F and higher APACHE-II scores on enrollment. Finally, patients with persistent HRF had fewer ventilator-free days (VFD) and higher mortality compared to resolving HRF (20% vs 8% in discovery, 28% vs 13% in validation).

Mortality of persistent HRF

After adjusting for age, sex, and baseline P/F in the discovery cohort, persistent HRF was associated with 2.33 (95% CI 1.42, 3.82) times higher risk of mortality compared to resolving HRF (Table 3). Adjusting for APACHE-II attenuated the relative risk, although the estimated relative risk was still 1.49 (95% CI 0.92, 2.41). In the validation cohort, both models adjusting for P/F and APACHE-II demonstrated a significantly higher risk of mortality associated with persistent HRF compared to resolving HRF (Table 3).

LASSO model to predict persistent HRF

In order to develop the prediction model, I split the discovery cohort randomly into training (n=474) and internal test (n=158) sets, and used VALID for external validation (n=906). Select features of the 3 datasets are listed in Table 4. Notably, the proportion of persistent HRF was highest in the external validation cohort, as was overall illness severity (as evidenced by higher mortality, shock, pneumonia). However, many other cohort features, such age, sex, and pre-existing chronic conditions were similar across the 3 datasets.

Of 33 candidate clinical predictors, two predictors (APACHE-II and P/F) were selected in the most parsimonious model that was within one standard deviation of the lowest mean square error. This model had moderate performance with an AUC 0.79 (95% CI 0.75, 0.83) in the training set, and adequate reliability on the basis of the Hosmer Lemeshow test (Table 5). The model was a significant improvement

compared to either APACHE-II ($p < 0.001$) or P/F ($p < 0.001$) alone in the training set based on Delong test. Performance was similar in the internal test set with an AUC 0.76 (0.68, 0.83) although calibration was not optimal based on Hosmer Lemeshow test. The model performance was weaker in the external validation cohort, with an AUC 0.69 (95% CI 0.65, 0.72) and poor model fit based on a high Hosmer-Lemeshow statistic.

Biomarkers to predict persistent HRF

Next I examined whether one or a subset of biomarkers could improve the performance of the LASSO model. After examining several combinations in the training and internal test set, I found that adding Ang-2 and IL-6 to the LASSO model raised AUC to 0.82 (95% CI 0.77, 0.86) in the training set, which was a significant improvement (Delong test $p = 0.041$). The performance internal test set was also slightly improved with an AUC 0.78 (95% CI 0.69, 0.87). I carried this model forward for external validation. In the validation cohort, the model performance was moderate, with an AUC 0.72 (95% CI 0.65, 0.79) and a good fit for the data based on Hosmer Lemeshow statistic. In total I assessed the model among 329 patients in the training set with the relevant biomarkers measured, 117 in the internal test set, and 223 patients in the external validation set.

I then determined a range of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) across a range of model thresholds (Table 6). Notably, at a model threshold of 0.8, the PPV was 93% in the training cohort, 92% in external validation, and 81% in internal test cohort. None of the tested thresholds gave an NPV that exceeded 80% in all three sets.

DISCUSSION

In summary, I showed that persistent HRF was common among critically ill adults with acute HRF, and associated with a higher risk for mortality compared to those with resolving HRF in 2 independent cohorts. In addition, I developed and validated a multivariable model with P/F, APACHE-II, IL-6, and Ang-2 to help discriminate between patients who will and will not go on to develop persistent HRF.

It is not unexpected that the overall mortality among those with persistent HRF is higher than those with resolving HRF, especially since the higher prevalence of such acute conditions as shock and pneumonia, the low P/F, and the higher APACHE-II indicated higher illness severity at baseline. However, persistent HRF was associated with a higher risk for mortality compared to resolving HRF even in analyses that adjust for the initial severity of respiratory failure. Although models adjusting for APACHE-II in the discovery cohort did attenuate the risk, the findings suggest other factors could be contributing that factors other than initial illness severity, and other patient-level (e.g. genomic), treatment-level (e.g. antibiotic timing), and/or illness-specific (e.g. type of pathogen) could also be contributing. Further investigating such differences between persistent and resolving HRF could offer new opportunities for clinical intervention. Even the relative magnitude of the mortality differences that I identified may guide efforts at prognostication in the early ICU course, as well as size of anticipated study populations in HRF trials with mortality endpoints. Overall, the findings extend prior literature in ARDS to the larger population of acute HRF. My findings were consistent with the prior ARDS studies, where over 13% of patients had resolution of ARDS within just 24 hours as well as substantially lower mortality than those with more prolonged disease (15, 17).

The ability to predict patients at highest risk for persistent HRF would enable early enrollment into clinical trials aiming to enrich for poor outcomes related to respiratory failure, and may inform how clinicians direct care and resources. The clinical variables selected by the LASSO model, P/F and APACHE-II, relate to severity of respiratory failure and overall physiologic illness severity and are used frequently to prognosticate in the ICU(46). While they improved model accuracy over either score alone, the model performance is still limited, especially in external validation. I hypothesize this may be related to the fact that the validation set enrolled a sicker population than the discovery cohort, which may challenge further discrimination. Notably, the difficulties of external validation mirrors experiences using clinical variables to predict ARDS, as with the lung injury prediction score (LIPS), which that had strong early data (47, 48) yet underperformed in real world settings (49). Furthermore, the limitations of using early clinical markers to predict persistent HRF also mirrors prior studies examining trajectories of ARDS, which suggest that the earliest measures of severity are not clear indicators for severity of respiratory failure at subsequent timepoints over the first week of ICU stays (30, 31). For example, Madotto in a

reanalysis of a large multicenter ARDS cohort, showed that 24% had resolution of ARDS by ICU day 2, and even though half of those had moderate-severe ARDS based on low P/F at baseline. Similarly Pham in this same cohort showed that 80% of patients with mild ARDS on ICU admission worsen over the first week. Nonetheless, even though the model of P/F and APACHE-II has limited performance in the validation cohort, it fills an important gap in the literature; indeed most ICU prediction models tend to focus on mortality(50), some on the duration of mechanical ventilation (51, 52), and I know of none that examine persistent HRF as an outcome other than the aforementioned ARDS studies.

Moreover, adding the research biomarkers IL-6, a marker of systemic inflammation, and Ang-2, a marker of endothelial dysfunction, to our LASSO model significantly improved model performance. These biomarkers have been associated with outcomes in multiple studies of ARDS, albeit to a lesser degree in other acute HRF populations (38, 45). Prior studies also suggest that similar research biomarkers of inflammation and lung injury are more important than clinical variables in classifying high-risk subgroups in the ICU(16, 38). In addition, I was able to identify threshold with PPV exceeding 90% in the training and validation sets. Although these thresholds do exclude many patients who do go on to develop persistent HRF, they would help enable the identification of a relatively homogenous population of high-risk patients. This may be valuable for administration of investigational therapies limited in quantity by cost or supply, or for therapies with an adverse side effect profile not intended for patients who may improve on their own. It may also help advance prognostic enrichment in clinical trials, which often carefully select populations to avoid diluting evidence of therapeutic effect. Altogether, my findings support further developing these biomarkers for point-of-care clinical use in patient populations.

This work has several limitations. First, although I aimed to understand the relative risk of mortality in those with persistent HRF compared to resolving HRF after adjusting for initial illness severity, residual confounding may be present. Second, I could only assess the model performance of biomarker-based in the relatively small subset of patients with appropriate measurements. In the validation cohort in particular, only 223/906 (25%) of patients had biomarkers measured, and this selection could lead to bias. Third, many clinically available laboratory parameters of inflammation (such as C reactive protein, procalcitonin) were unavailable as candidate clinical predictors, and could offer similar value to the research biomarkers I examined. Nonetheless, this study clarifies details on the subgroup of patients with

persistent HRF at day 3, advancing our understanding of the heterogeneity among patients with acute HRF requiring mechanical ventilation.

Table 1: Candidate predictors

Demographics	Age, sex
Comorbidities	Cirrhosis, chronic kidney disease, end-stage renal disease, alcohol use disorder, heart failure, diabetes
Acute conditions on ICU admission	Pneumonia, pancreatitis, aspiration, multiple transfusion, sepsis
Vitals/hemodynamics	Vasopressors, heart rate, respiratory rate, mean arterial pressure, temperature
Laboratory	Ratio of oxygen partial pressure to fraction of inspired oxygen, pH, carbon dioxide partial pressure, total bilirubin, blood urea nitrogen, creatinine, albumin, white blood cell count, hematocrit, sodium, bicarbonate
Composite scores	Acute physiology and chronic health evaluation II score, sequential organ failure assessment score, Glasgow coma scale
Biomarkers	Interleukin-6, interleukin-8, angiopoetin-2, soluble tumor necrosis factor receptor-1

Table 2: Cohort description by persistent and resolving HRF

	Discovery		**Validation**	
	Resolving	**Persistent**	**Resolving**	**Persistent**
	N=322	**N=298**	**N=344**	**N=514**
Age, years	52 (43-62)	52 (40-60)	58 (46-66)	55 (44-66)
Female sex	124 (39%)	98 (33%)	153 (44%)	224 (44%)
Diabetes	91 (28%)	92 (31%)	117 (34%)	154 (30%)
Cirrhosis	36 (11%)	34 (11%)	48 (14%)	54 (11%)
Heart failure	28 (9%)	32 (11%)	39 (11%)	50 (10%)
Alcohol use disorder	91 (28%)	104 (35%)	62 (18%)	105 (20%)
ICU type				
Surgical	109 (34%)	94 (32%)	122 (35%)	162 (32%)
Medical	213 (66%)	204 (68%)	222 (65%)	352 (68%)
Shock	68 (21%)	132 (44%)	149 (43%)	285 (55%)
Sepsis	250 (78%)	256 (86%)	178 (52%)	364 (71%)
Pneumonia	69 (21%)	102 (34%)	123 (36%)	267 (52%)
ARDS	44 (14%)	95 (32%)	80 (23%)	241 (47%)
P/F	176 (130-230)	112 (77-165)	177 (118-218)	122 (80-193)
APACHE-II	19 (15-24)	25 (20-30)	28 (23-33)	31 (26-36)
VFD	25 (23-26)	16 (0-21)	26 (23-26)	16 (0-23)
Mortality	26 (8%)	60 (20%)	46 (13%)	144 (28%)

All continuous variables are expressed as median (interquartile range), and all categorical variables are expressed as n (%). Table excludes patients who died before day 3. ICU = intensive care unit. ARDS = acute respiratory distress syndrome. P/F = ratio of arterial oxygen partial pressure to fraction of inspired oxygen. APACHE = acute physiology and chronic health evaluation. VFD = ventilator free days.

Table 3: Mortality associated with persistent HRF

	Deaths	**Relative risk (95% Confidence Interval)**		
	N (%)	*Unadjusted*	*Model A*	*Model B*
Discovery Cohort				

Resolving	26 (8%)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Persistent	60 (20%)	2.49 (1.62, 3.84)**	2.33 (1.42, 3.82)*	1.49 (0.92, 2.41)
Validation Cohort				
Resolving	46 (13%)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Persistent	144 (28%)	2.10 (1.55, 2.84)**	2.05 (1.51, 2.78)**	1.81 (1.34, 2.44)**

Mortality is in-hospital mortality 28 days after ICU admission.
Model A: adjusted for age, sex, and P/F on enrollment.
Model B: adjusted for age, sex, APACHE-II
** $p < 0.001$ * $p < 0.01$

Table 4: Cohort description by training, internal test, and external validation cohorts

	Training N=474	Internal Test N=158	External Validation N=906
Age	53 (42-61)	52 (41-61)	56 (45-66)
Female sex	173 (36%)	52 (33%)	396 (44%)
Diabetes	141 (30%)	45 (28%)	283 (31%)
Cirrhosis	57 (12%)	13 (8%)	116 (13%)
Heart failure	47 (10%)	14 (9%)	94 (10%)
Alcohol use disorder	146 (31%)	52 (33%)	177 (20%)
ICU type			
Surgical	151 (32%)	57 (36%)	289 (32%)
Medical	323 (68%)	101 (64%)	617 (68%)
Shock	162 (34%)	49 (31%)	475 (52%)
Sepsis	392 (83%)	125 (79%)	579 (64%)
Pneumonia	137 (29%)	36 (23%)	418 (46%)
ARDS	106 (22%)	36 (23%)	342 (38%)
P/F	140 (92-204)	152 (101-218)	143 (89-204)
APACHE-II	22 (17-28)	20 (17-26)	30 (25-35)
IL-6, pg/mL	155 (63-393)	137 (63-478)	77 (22-398)
IL-8, pg/mL	15 (7-33)	15 (8-43)	25 (10-99)
sTNFR-1, pg/mL	9321 (5559-17746)	10532 (6634-15943)	3729 (2063-6951)
Ang-2, pg/mL	16649 (8408-38191)	15752 (8824-36974)	7364 (3564-14793)
Persistent HRF	242 (51%)	67 (42%)	562 (62%)
VFD	21 (5-25)	22 (6-25)	21 (0-25)
Mortality	74 (16%)	24 (15%)	238 (26%)

All continuous variables are expressed as median (interquartile range), and all categorical variables are expressed as n (%). ICU = intensive care unit. ARDS = acute respiratory distress syndrome. P/F = ratio of arterial oxygen partial pressure to fraction of inspired oxygen. APACHE = acute physiology and chronic health evaluation. IL-6 = Interleukin 6, measured in 329, 117, and 515 patients across training, internal test, and external validation cohorts respectively. IL-8, measured in 398, 117, and 516 patients. sTNFR-1, measured in 329, 117, and 477 patients. Angiopoetin-2, measured in 354, 122, and 226 patients. VFD = ventilator free days.

Table 5: Model Performance in training, internal test, and external validation cohorts

	Training		Internal Test		External Validation	
	N = 474		N = 158		N = 906	
	AUC	HL	AUC	HL	AUC	HL
Model 1 (P/F + APACHE-II)	0.79 (0.75, 0.83)	15.50 (p = 0.050)	0.76 (0.68, 0.83)	22.19 (p = 0.005)	0.69 (0.65, 0.72)	93.99 (p < 0.001)
Model 2 (Model 1 + Ang2 + IL6)	0.82 (0.77, 0.86)	10.54 (p=0.229)	0.78 (0.69, 0.87)	18.47 (p=0.018)	0.72 (0.65, 0.79)	14.56 (p=0.068)
APACHE-II	0.73 (0.69, 0.78)	4.86 (p = 0.773)	0.72 (0.64, 0.80)	9.06 (p=0.337)	0.63 (0.60, 0.67)	101.94 (p <0.001)
P/F	0.73 (0.68, 0.77)	6.01 (p=0.646)	0.70 (0.61, 0.78)	12.50 (p = 0.130)	0.66 (0.62, 0.70)	70.68 (p < 0.001)

AUC area under the curve. HL Hosmer-Lemeshow test statistic.
Model 1 was not carried forward for external validation.

Table 6: Negative and positive predictive values for final model

Cohort	**Model Score threshold**	**Patients above/below threshold**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Training	0.80	57/272	33	98	93	61
	0.50	159/170	74	75	74	75
	0.30	215/114	84	52	62	77
Internal Test	0.80	16/101	27	96	81	65
	0.50	52/65	73	75	67	80
	0.30	80/37	90	46	54	86
External validation	0.80	74/149	42	90	92	37
	0.50	160/63	79	48	80	46
	0.30	205/18	95	16	75	56

PPV = Positive predictive value. NPV = Negative predictive value.
Model score threshold given by: Log(odds) = 0.02 + 1.08 * (APACHE-II) + 0.99 (P/F) + 1.18 * (Ang-2) + 1.21 * (IL-6). Ang-2 and IL-6 were log(2) transformed.

REFERENCES

1. Herridge MS, Cheung AM, Tansey CM, et al.: One-Year Outcomes in Survivors of the Acute Respiratory Distress Syndrome. *N Engl J Med* 2003; 348:683–693

2. Vincent J-L, Akça S, de Mendonça A, et al.: The Epidemiology of Acute Respiratory Failure in Critically Ill Patients. *Chest* 2002; 121:1602–1609

3. Esteban A, Anzueto A, Frutos F, et al.: Characteristics and Outcomes in Adult Patients Receiving Mechanical Ventilation: A 28-Day International Study. *JAMA* 2002; 287:345–355

4. Rubenfeld GD, Caldwell E, Peabody E, et al.: Incidence and Outcomes of Acute Lung Injury. *N Engl J Med* 2005; 353:1685–1693

5. Herridge MS, Chu LM, Matte A, et al.: The RECOVER Program: Disability Risk Groups and 1-Year Outcome after 7 or More Days of Mechanical Ventilation. *Am J Respir Crit Care Med* 2016; 194:831–844

6. Carson SS, Cox CE, Holmes GM, et al.: The Changing Epidemiology of Mechanical Ventilation: A Population-Based Study. *J Intensive Care Med* 2006; 21:173–182

7. Halpern NA, Goldman DA, Tan KS, et al.: Trends in Critical Care Beds and Use Among Population Groups and Medicare and Medicaid Beneficiaries in the United States: 2000-2010. *Crit Care Med* 2016; 44:1490–1499

8. Richardson S, Hirsch JS, Narasimhan M, et al.: Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* 2020; 323:2052–2059

9. Auld SC, Caridi-Scheible M, Blum JM, et al.: ICU and Ventilator Mortality Among Critically Ill Adults With Coronavirus Disease 2019 [Internet]. *Crit Care Med* 2020; [cited 2020 May 31] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7255393/

10. Santacruz CA, Pereira AJ, Celis E, et al.: Which Multicenter Randomized Controlled Trials in Critical Care Medicine Have Shown Reduced Mortality? A Systematic Review. *Critical Care Medicine* 2019; 47:1680–1691

11. Prescott HC, Calfee CS, Thompson BT, et al.: Toward Smarter Lumping and Smarter Splitting: Rethinking Strategies for Sepsis and Acute Respiratory Distress Syndrome Clinical Trial Design. *Am J Respir Crit Care Med* 2016; 194:147–155

12. Wu AC, Kiley JP, Noel PJ, et al.: Current Status and Future Opportunities in Lung Precision Medicine Research with a Focus on Biomarkers. An American Thoracic Society/National Heart, Lung, and Blood Institute Research Statement. *Am J Respir Crit Care Med* 2018; 198:e116–e136

13. Moore WC, Fitzpatrick AM, Li X, et al.: Clinical Heterogeneity in the Severe Asthma Research Program. *Ann Am Thorac Soc* 2013; 10:S118–S124

14. Han MK, Agusti A, Calverley PM, et al.: Chronic obstructive pulmonary disease phenotypes: the future of COPD. *Am J Respir Crit Care Med* 2010; 182:598–604

15. Schenck EJ, Oromendia C, Torres LK, et al.: Rapidly Improving ARDS in Therapeutic Randomized Controlled Trials. *Chest* 2019; 155:474–482

16. Calfee CS, Delucchi K, Parsons PE, et al.: Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med* 2014; 2:611–620

17. Sanchez E, Price DR, Chung K-P, et al.: Persistent severe acute respiratory distress syndrome for the prognostic enrichment of trials. *PLOS ONE* 2020; 15:e0227346

18. Calfee CS, Delucchi KL, Sinha P, et al.: Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomised controlled trial. *Lancet Respir Med* 2018; 6:691–698

19. Sinha P, Delucchi KL, Thompson BT, et al.: Latent class analysis of ARDS subphenotypes: a secondary analysis of the statins for acutely injured lungs from sepsis (SAILS) study. *Intensive Care Med* 2018; 44:1859–1869

20. 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

21. Thille AW, Esteban A, Fernández-Segoviano P, et al.: Chronology of histological lesions in acute respiratory distress syndrome with diffuse alveolar damage: a prospective cohort study of clinical autopsies. *The Lancet Respiratory Medicine* 2013; 1:395–401

22. Bos LDJ, Scicluna BP, Ong DSY, et al.: Understanding Heterogeneity in Biologic Phenotypes of Acute Respiratory Distress Syndrome by Leukocyte Expression Profiles. *Am J Respir Crit Care Med* 2019; 200:42–50

23. Sjoding MW, Hofer TP, Co I, et al.: Interobserver Reliability of the Berlin ARDS Definition and Strategies to Improve the Reliability of ARDS Diagnosis. *Chest* 2018; 153:361–367

24. Sjoding MW, Hofer TP, Co I, et al.: Differences between Patients in Whom Physicians Agree and Disagree about the Diagnosis of Acute Respiratory Distress Syndrome. *Annals ATS* 2018; 16:258–264

25. Rubenfeld GD, Caldwell E, Granton J, et al.: Interobserver Variability in Applying a Radiographic Definition for ARDS. *Chest* 1999; 116:1347–1353

26. Bellani G, Laffey JG, Pham T, et al.: Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. *JAMA* 2016; 315:788–800

27. Bhatraju PK, Mukherjee P, Robinson-Cohen C, et al.: Acute kidney injury subphenotypes based on creatinine trajectory identifies patients at increased risk of death. *Critical Care* 2016; 20:372

28. Ross JC, Castaldi PJ, Cho MH, et al.: Longitudinal Modeling of Lung Function Trajectories in Smokers with and without Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2018; 198:1033–1042

29. Béduneau G, Pham T, Schortgen F, et al.: Epidemiology of Weaning Outcome according to a New Definition. The WIND Study. *Am J Respir Crit Care Med* 2016; 195:772–783

30. Madotto F, Pham T, Bellani G, et al.: Resolved versus confirmed ARDS after 24 h: insights from the LUNG SAFE study. *Intensive Care Med* 2018; 44:564–577

31. Pham T, Serpa Neto A, Pelosi P, et al.: Outcomes of Patients Presenting with Mild Acute Respiratory Distress Syndrome: Insights from the LUNG SAFE Study. *Anesthesiology* 2019; 130:263–283

32. Burnham EL, Janssen WJ, Riches DWH, et al.: The fibroproliferative response in acute respiratory distress syndrome: mechanisms and clinical significance. *Eur Respir J* 2014; 43:276–285

33. Davey A, McAuley DF, O'Kane CM: Matrix metalloproteinases in acute lung injury: mediators of injury and drivers of repair. *European Respiratory Journal* 2011; 38:959–970

34. Morrell ED, Bhatraju PK, Mikacenic CR, et al.: Alveolar Macrophage Transcriptional Programs Are Associated with Outcomes in Acute Respiratory Distress Syndrome. *Am J Respir Crit Care Med* 2019; 200:732–741

35. Marshall RP, Bellingan G, Webb S, et al.: Fibroproliferation Occurs Early in the Acute Respiratory Distress Syndrome and Impacts on Outcome. *Am J Respir Crit Care Med* 2000; 162:1783–1788

36. Clark JG, Milberg JA, Steinberg KP, et al.: Type III Procollagen Peptide in the Adult Respiratory Distress Syndrome: Association of Increased Peptide Levels in Bronchoalveolar Lavage Fluid with Increased Risk for Death. *Annals of Internal Medicine* 1995; 122:17–23

37. Santos FB, Nagato LKS, Boechem NM, et al.: Time course of lung parenchyma remodeling in pulmonary and extrapulmonary acute lung injury. *J Appl Physiol* 2006; 100:98–106

38. Mikacenic C, Price BL, Harju-Baker S, et al.: A Two-Biomarker Model Predicts Mortality in the Critically Ill with Sepsis. *Am J Respir Crit Care Med* 2017; 196:1004–1011

39. Bhatraju PK, Robinson-Cohen C, Mikacenic C, et al.: Circulating levels of soluble Fas (sCD95) are associated with risk for development of a nonresolving acute kidney injury subphenotype [Internet]. *Crit Care* 2017; 21[cited 2021 Apr 27] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5559814/

40. Singer M, Deutschman CS, Seymour CW, et al.: The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016; 315:801–810

41. Ware LB, Koyama T, Zhao Z, et al.: Biomarkers of lung epithelial injury and inflammation distinguish severe sepsis patients with acute respiratory distress syndrome. *Crit Care* 2013; 17:R253

42. ARDS Definition Task Force, Ranieri VM, Rubenfeld GD, et al.: Acute respiratory distress syndrome: the Berlin Definition. *JAMA* 2012; 307:2526–2533

43. Rice TW, Wheeler AP, Bernard GR, et al.: Comparison of the Spo2/Fio2 Ratio and the Pao2/Fio2 Ratio in Patients With Acute Lung Injury or ARDS. *Chest* 2007; 132:410–417

44. Zou G: A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *Am J Epidemiol* 2004; 159:702–706

45. van der Zee P, Rietdijk W, Somhorst P, et al.: A systematic review of biomarkers multivariately associated with acute respiratory distress syndrome development and mortality [Internet]. *Crit Care* 2020; 24[cited 2020 Sep 19] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7245629/

46. Knaus WA, Draper EA, Wagner DP, et al.: APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13:818–829

47. Gajic O, Dabbagh O, Park PK, et al.: Early identification of patients at risk of acute lung injury: evaluation of lung injury prediction score in a multicenter cohort study. *Am J Respir Crit Care Med* 2011; 183:462–470

48. Soto GJ, Kor DJ, Park PK, et al.: Lung Injury Prediction Score in Hospitalized Patients at risk of Acute Respiratory Distress Syndrome. *Crit Care Med* 2016; 44:2182–2191

49. Kor DJ, Carter RE, Park PK, et al.: Effect of Aspirin on Development of ARDS in At-Risk Patients Presenting to the Emergency Department: The LIPS-A Randomized Clinical Trial. *JAMA* 2016; 315:2406–2414

50. Spragg RG, Bernard GR, Checkley W, et al.: Beyond Mortality. *Am J Respir Crit Care Med* 2010; 181:1121–1127

51. Sayed M, Riaño D, Villar J: Novel criteria to classify ARDS severity using a machine learning approach. *Critical Care* 2021; 25:150

52. Seneff MG, Zimmerman JE, Knaus WA, et al.: Predicting the Duration of Mechanical Ventilation: The Importance of Disease and Patient Characteristics. *CHEST* 1996; 110:469–479