Published in final edited form as:

Crit Care Med. 2011 May ; 39(5): 1023–1028.

Short-Term Mortality Prediction for Acute Lung Injury Patients: External Validation of the ARDSNet Prediction Model

Abdulla Damluji, MBChB, MPH¹, Elizabeth Colantuoni, PhD^{2,3}, Pedro A. Mendez-Tellez, MD², Jonathan E. Sevransky, MD, MHS⁴, Eddy Fan, MD⁴, Carl Shanholtz, MD⁵, Margaret Wojnar, M.D.⁶, Peter J. Pronovost, MD, PhD², and Dale M. Needham, MD, PhD⁴

¹ Department of Internal Medicine, Pennsylvania State University, Hershey, Pennsylvania ² Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University, Baltimore, Maryland ³ Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland ⁴ Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, Maryland ⁵ Division of Pulmonary and Critical Care Medicine, University of Maryland, Baltimore, Maryland ⁶ Division of Pulmonary and Critical Care Medicine, Pennsylvania State University, Hershey, Pennsylvania

Abstract

Objective—An independent cohort of acute lung injury (ALI) patients was used to evaluate the external validity of a simple prediction model for short-term mortality previously developed using data from ARDS Network (ARDSNet) trials.

Design, Setting, and Patients—Data for external validation were obtained from a prospective cohort study of ALI patients from 13 ICUs at four teaching hospitals in Baltimore, Maryland.

Measurements and Main Results—Of the 508 non-trauma, ALI patients eligible for this analysis, 234 (46%) died in-hospital. Discrimination of the ARDSNet prediction model for inhospital mortality, evaluated by the area under the receiver operator characteristics curves (AUC), was 0.67 for our external validation dataset versus 0.70 and 0.68 using APACHE II and the ARDSNet validation dataset, respectively. In evaluating calibration of the model, predicted versus observed in-hospital mortality for the external validation dataset was similar for both low risk (ARDSNet model score = 0) and high risk (score = 3 or 4+) patient strata. However, for intermediate risk (score = 1 or 2) patients, observed in-hospital mortality was substantially higher than predicted mortality (25.3% vs. 16.5% and 40.6% vs. 31.0% for score = 1 and 2, respectively). Sensitivity analyses limiting our external validation data set to only those patients meeting the ARDSNet trial eligibility criteria and to those who received mechanical ventilation in compliance with the ARDSNet ventilation protocol, did not substantially change the model's discrimination or improve its calibration.

Conclusions—Evaluation of the ARDSNet prediction model using an external ALI cohort demonstrated similar discrimination of the model as was observed with the ARDSNet validation dataset. However, there were substantial differences in observed versus predicted mortality among intermediate risk ALI patients. The ARDSNet model provided reasonable, but imprecise, estimates of predicted mortality when applied to our external validation cohort of ALI patients.

Corresponding author (No reprints will be ordered): Dale M. Needham, MD, PhD, Pulmonary and Critical Care Medicine, Johns Hopkins University, 1830 E Monument St., Fifth Floor, Baltimore, MD 21205; dale.needham@jhmi.edu, Phone: 410-955-3467; Fax: 410-955-0036.

The other authors have not disclosed any potential conflicts of interest.

Keywords

respiratory distress syndrome; adult; statistical model; mortality determinants; prognosis; health status indicators; intensive care units

Although mortality from acute lung injury (ALI) has declined over recent decades (1, 2), it remains a significant challenge in the intensive care unit (ICU). ALI mortality varies by age, comorbidity, and ICU severity of illness and organ dysfunction (3). Prior studies identified a number of early predictive factors for mortality in ALI and the acute respiratory distress syndrome (ARDS) (4–6). A recent study used data from the National Institutes of Heart, Lung and Blood ARDS Network (ARDSNet) low tidal volume trial (ARMA study)(7), to develop a simple, ALI-specific, multivariable predictive model of mortality (8). This ARDSNet model was validated using an independent dataset from the subsequent ARDSNet trial of higher versus lower positive end-expiratory pressure (ALVEOLI study) (9).

Simple mortality prediction models can be useful for prognostication of patients cohorts. (10). Before adopting a new prediction model, it is important to evaluate its “generalizability” in two ways (11). First, one should evaluate the “reproducibility” of the model’s mortality predictions among patients not included in the development model, but drawn from a similar population, as occurred via evaluation of the ARDSNet model using the ALVEOLI study data (8). Second, to evaluate “transportability”, the accuracy of the model’s predictions should be examined in a new patient population with the same disease, but with different enrollment methods and patient characteristics (11).

In this study, our primary objective was to evaluate the “transportability” of the ARDSNet mortality model in comparing the mortality prediction results from the ARDSNet’s development cohort (ARMA) and internal validation cohort (ALVEOLI) versus an “external validation” cohort of ALI patients.

Methods

Study Population

Our external validation patient cohort used for this analysis is derived from a clinical study of ALI patients enrolled from 2004 to 2007 from 13 ICUs at four academic teaching hospitals in Baltimore, Maryland (12). Patients were eligible for inclusion in the clinical study if they were mechanically ventilated adults (age>18 years) with ALI, as defined by the American-European Consensus Conference (13). Relevant exclusion criteria for this cohort study included: (1) >96 hours between ALI diagnosis and enrollment or pre-existing ALI of >24 hours duration when transferred to a study site ICU, (2) >5 days of mechanical ventilation during the current hospitalization before ALI and study enrollment, (3) pre-existing illness with a life expectancy of <6 months or any limitation in the escalation of ICU care at the time of enrollment (with exception of a “No Cardio-Pulmonary Resuscitation (CPR)” order), (4) prior lung resection, and (5) no fixed address or inability to speak or understand English (12). Similar to the approach taken in development of the ARDSNet prediction model, patients with trauma as their primary ALI risk factor were excluded in this analysis, due to their markedly lower mortality rate (8).

Patient Demographic and Severity of Illness Exposure Variables

To be consistent with the development process undertaken for the ARDSNet model, the following categories of baseline exposure variables, which were prospectively collected from the medical record during the clinical study, are reported in this analysis: demographics, severity of illness, primary ALI risk factor, and laboratory and physiology

data at the time of ALI diagnosis. In this study, illness severity was measured by the Charlson comorbidity index (14), APACHE II score at ICU admission (15), and Sequential Organ Failure Assessment (SOFA) score at study enrollment (16). As in development of the ARDSNet prediction model, cumulative fluid balance was defined as total fluid intake minus output in the first day after ALI diagnosis.

Outcomes Variables

Consistent with the ARDSNet model, the primary outcome of this analysis was in-hospital mortality, with 28-day mortality evaluated as a secondary outcome. Discrimination of the ARDSNet model was also compared to the APACHE II scoring system calculated during the first 24 hours after ICU admission.

Sensitivity Analyses

Two a priori sensitivity analyses were performed. First, the primary analyses were replicated in a subgroup of our external validation cohort which was matched for the eligibility criteria of the ARMA trial originally used to develop the ARDSNet prediction model (7). In this sensitivity analysis, patients were excluded from our external validation cohort if they met any of the following criteria: pregnant, tricyclic antidepressant overdose, hemoglobin SS/SC, severe chronic respiratory disease, morbid obesity, bone marrow transplant, lung transplant, and severe chronic liver disease. Second, this subgroup of patients from the first sensitivity analysis was further restricted to include only those patients whose mechanical ventilation was compliant with the ARDSNet protocol for low tidal volume ventilation during at least 50% of ventilator days.

Statistical Analysis

Continuous variables were reported as medians with interquartile range (IQR) and categorical variables as proportions, with comparison made using Wilcoxon rank sum, Student's t, and chi-square tests, as appropriate. A mortality prediction score was calculated for each patient using the original prediction model (Box 1) (8). As in the development and validation of the ARDSNet model, its predictive ability was evaluated for both discrimination and calibration. Discrimination is the ability of the prediction model to distinguish between those persons with and those without the outcome (17). Calibration refers to the agreement between predicted and observed (i.e. actual) outcomes across the spectrum of risk strata (17). To evaluate discrimination of the ARDSNet model using our external validation cohort, receiver operating characteristics (ROC) curves of the model's point score with cohort's observed mortality were plotted and assessed by calculating the area under the ROC curve (AUC). For comparison purposes, using data from our external validation model, the AUC was also calculated for the APACHE II score and statistically compared to the AUC of the ARDSNet model using the methods of Delong et al (18). To evaluate calibration of the ARDSNet model, the observed mortality rates from our external validation cohort data were calculated and compared to the predicted mortality from the model for each strata of model's scores (0, 1, 2, 3, and 4+). Two sided p-values were considered statistically significant when <0.05 . All analyses were performed using Stata 11.0 software (Stata Corporation, College Station, Tx, USA). The Institutional Review Board (IRB) at Johns Hopkins University and all participating sites approved the study.

Results

Study Population

Of the 520 patients enrolled in our external validation cohort, 12 patients (2%) were excluded, for purposes of this analysis, due to trauma as the primary risk ALI factor, leaving

a sample size of 508 eligible patients. Demographics, primary ALI risk factor, illness severity, and laboratory and physiologic data for this 508 patient external validation cohort are presented in Table 1. Within this cohort, patient mortality at 28 days and hospital discharge was 37% and 46%, respectively. Patients who died were older with a greater severity of illness and physiologic and laboratory abnormalities (Table 1).

In-Hospital Mortality

The AUC for in-hospital mortality using the ARDSNet and the APACHE II prediction models with our external validation patient cohort data (Figure) was 0.67 and 0.70, respectively ($p=0.207$ for comparison), which was similar to the AUC of 0.68 in the ARDSNet internal validation cohort (ALVEOLI study) (8). The predicted versus the observed mortality for the ARDSNet development cohort (ARMA study), internal validation cohort (ALVEOLI study) and our external validation cohort was similar in both low risk (score = 0) and high risk (score = 3 or 4+) patient strata (Table 2). However, for patients in the intermediate risk strata (score = 1 or 2), observed in-hospital mortality in our external validation cohort was substantially higher than predicted (25.3% vs. 16.5% and 40.6% vs. 31.0% for scores = 1 and 2, respectively).

Sensitivity Analysis—The *a priori* sensitivity analysis, which excluded patients from our external validation cohort who did not meet the eligibility criteria of the development cohort (ARMA study), left a sample size of 402 patients with in-hospital mortality of 42%. In this sensitivity analysis, the AUC for inhospital mortality with our external validation model and APACHE II was 0.64 and 0.69, respectively ($p=0.089$ for comparison) versus 0.68 in the ARDSNet internal validation cohort (ALVEOLI study). Compared with calibration in the primary results, this sensitivity analysis demonstrated similar differences in predicted versus observed mortality in intermediate risk patients (score = 1 or 2) and growing differences in high risk patients (score = 3 or 4+) with predicted mortality substantially higher than observed mortality in our external validation cohort (Table 2). When the sensitivity analysis was further restricted to patients compliant with the ARDSNet ventilation protocol, there was no improvement in discrimination or calibration of the model.

28-Day Mortality

For 28-day mortality, the AUC of the ARDSNet model and APACHE II, using data from our external validation cohort, was 0.64 and 0.70, respectively (Figure: $p=0.039$ for comparison) versus 0.71 in the ARDSNet internal validation cohort (ALVEOLI study) (8). The predicted versus observed 28-day mortality for the development (ARMA trial), internal validation (ALVEOLI trial), and our external validation cohorts was lower in both the low risk (score = 0) and intermediate risk (score = 1 or 2) patient groups (Table 3). However, in patients with the highest risk (score = 4+), the predicted mortality was substantially higher than our external validation cohort (60.0% vs. 50.3%, respectively).

Sensitivity Analysis—The mortality at 28 days, after applying the ARDSNet eligibility criteria (ARMA study) to our external validation cohort, was 33.6%. In this sensitivity analysis, the AUC for 28-day mortality model and APACHE II was 0.62 and 0.71, respectively ($p=0.006$ for comparison). With respect to calibration, similar results were obtained in the sensitivity analysis compared to the primary analysis, with the substantial difference between observed and predicted mortality being even greater in the highest risk (score = 4+) patient strata (Table 3). In the second sensitivity analysis, which further restricted this patient subgroup to only patients compliant with the ARDSNet ventilation protocol, there was no substantial improvement in the model discrimination and calibration (Table 3).

Discussion

Using data from an independent cohort of 508 ALI patients, we evaluated the generalizability of a recent ARDSNet mortality prediction model. The ability of the ARDSNet model to distinguish between patients who do, and do not, die (i.e. model discrimination) using the external patient cohort was comparable to the original ARDSNet development model from ARMA study, and to the APACHE II scoring system. Moreover, in evaluating the predicted versus observed mortality across strata of the ARDSNet model scores (i.e. model calibration), we found that for both low-and high-risk patients (score = 0 and score = 3–4+, respectively), in-hospital mortality was similar with absolute differences in predicted versus observed mortality of <5.5%. However, for intermediate-risk patients (score = 1 and 2), the observed in-hospital mortality was substantially higher (absolute differences ~9–10%) than that predicted by the model. In two sensitivity analyses evaluating subsets of patients from our external validation cohort, the primary results did not substantially change. Thus, the ARDSNet model has reasonable discrimination but imprecise calibration for intermediate risk patients when applied to our external validation cohort.

We believe that development of the ARDSNet model and evaluation of its generalizability is potentially helpful because an accurate mortality prediction model that is faster and less complex than existing models (e.g. APACHE) may have value in stratifying cohorts of ICU patients according to their risk of mortality. Such accurate prognostic information is important for assessment of ICU performance and to help identify appropriate patient groups to exclude from enrollment into clinical trials (i.e. patients who may be considered either “too healthy” or “too sick” to benefit an intervention aimed at improving mortality) (10, 19, 20). However, the advantages of simple prediction models may become less relevant as automated calculations of more complex models are increasingly employed. Also, routine calibration of mortality prediction models, like that of the ARDSNet study, is needed to account for differences in demographics and risk factors in different underlying populations and changes that may occur over time.

The ARDSNet model was developed using patient data from the ARMA trial and then validated in a very similar patient population using data from the subsequent ARDSNet ALVEOLI trial (7). As suggested by developers of the ARDSNet model (8), we undertook external validation of its transportability using data from a more recent prospective cohort study of ALI patients from four teaching hospitals in Baltimore (12). External validation involves quantifying the accuracy of a prediction model in patients with the same disease, but from a population with different methods of patient selection, severity of illness, and ICU performance, staffing, resources and technology (10).

Some differences in patients’ demographic, laboratory, and physiology data were observed between the ARDSNet model development cohort (ARMA trial) and our external validation cohort. In our external validation versus the development cohort, the largest difference in demographic factors was a 2-fold greater proportion of African Americans patients, which was not associated with crude hospital mortality in either cohort. Among those who died, in our external validation cohort versus model development cohort (ARMA study), there was a much higher prevalence of both shock (i.e., vasopressor use: 71% vs. 49%) and impaired renal function (i.e. lower 24-hour urine output and higher median creatinine), perhaps due to a higher proportion of patients having either pulmonary and non-pulmonary sepsis (82% vs. 67%). These factors likely contribute to the higher in-hospital mortality (46%) in our external validation cohort versus the ARDSNet development model (33%) (8). Including additional variables in the prognostic model may have addressed these issues and improved its calibration for this particular external validation cohort. In other ALI cohorts (eg

transplant or oncology patients with ALI), other additional prognostic factors would need to be considered for optimal model calibration. Other issues also may have contributed to differences in mortality rates between the two cohorts. First, there are important differences in study design. The development cohort was part of interventional randomized trial requiring informed consent, while our external validation cohort was part of an observational study which had an IRB-approved waiver of consent for collection of existing data from the medical record. This difference eliminates previously-identified patient exclusions based on age- and race/ethnicity-related consent issues (21). The mortality rate in the external validation cohort is similar to other observational studies of ALI patients (5, 22–25). Second, the external validation cohort is derived from a study with more liberal eligibility criteria than the ARDSNet trials (12). When the ARDSNet eligibility criteria were applied to our external validation cohort, the in-hospital mortality rate decreased from 46% to 42%. Third, approximately 50% of patients in our external cohort received mechanical ventilation according to the ARDSNet protocol while the ARDSNet model was developed and validated only in patients randomized to receive that protocol (8, 26). When restricting our external validation cohort to only those patients who received mechanical ventilation compliant with the ARDSNet protocol, in-hospital mortality decreased to 33%. Fourth, although there are many measured variables to objectively compare the development cohort to our external validation cohort, additional unmeasured variables may differ between the cohorts, especially given the differences in study designs. For example, moribund patients were not enrolled in the randomized trials, but could be enrolled in the observational study, which served as the external validation cohort. Such differences may have accounted for the lower mortality in the development cohort versus our external validation cohort.

There are important strengths of this study. Our external validation cohort is derived from a well-defined, prospective, multi-site, observational study of ALI patients with good variability in patient and clinical factors. Moreover, its more recent timing, larger number of mortality outcomes, and less strict eligibility criteria than the ARDSNet cohorts are important factors in evaluating generalizability of the ARDSNet prediction model.

There are also several potential limitations to this study. Our external validation cohort recruited patients from ICUs which participate in ARDSNet trials; however, our external cohort was more recent, and did enroll patients with broader eligibility criteria and differing characteristics, from the earlier ARDSNet cohorts used in developing and internal validating the prediction model. In addition, our external validation cohort was recruited from four teaching hospitals in one city and hence, may not be representative of patients at non-teaching hospitals or patients located in other geographic locations. These issues may have biased the external validation results toward better discrimination and calibration. Consequently, further evaluation of the ARDSNet prediction model in other ALI cohorts would be helpful.

In conclusion, we examined the generalizability of the ARDSNet mortality prediction model using data from a multi-site prospective cohort study. The model's ability to discriminate inhospital survival is similar to APACHE II, and to the ARDSNet development and internal validation models. However, the observed mortality among intermediate risk patients was higher than that predicted by the model. Thus, the ARDSNet model provides reasonable, but imprecise, estimates of predicted mortality when applied to our external validation cohort of ALI patients.

Acknowledgments

Dr. Colantuoni, Dr. Sevransky, Dr. Shanholtz, Dr. Pronovost, and Dr. Needham received funding from the NIH.

Funding: This research was supported by the National Institutes of Health (Acute Lung Injury SCCOR Grant # P050 HL 73994). Dr. Fan is supported by a Canadian Institutes of Health Research Fellowship Award. The funding bodies had no role in the study design, manuscript writing or decision to submit the manuscript for publication.

References

1. Zambon M, Vincent JL. Mortality rates for patients with acute lung injury/ARDS have decreased over time. *Chest*. 2008; 133(5):1120–1127. [PubMed: 18263687]
2. Erickson SE, Martin GS, Davis JL, et al. Recent trends in acute lung injury mortality: 1996–2005. *Crit Care Med*. 2009; 37(5):1574–1579. [PubMed: 19325464]
3. Bernard GR. Acute respiratory distress syndrome: a historical perspective. *Am J Respir Crit Care Med*. 2005; 172(7):798–806. [PubMed: 16020801]
4. Monchi M, Bellenfant F, Cariou A, et al. Early predictive factors of survival in the acute respiratory distress syndrome. A multivariate analysis. *Am J Respir Crit Care Med*. 1998; 158(4):1076–1081. [PubMed: 9769263]
5. Seeley E, McAuley DF, Eisner M, et al. Predictors of mortality in acute lung injury during the era of lung protective ventilation. *Thorax*. 2008; 63(11):994–998. [PubMed: 18566110]
6. Luecke T, Muench E, Roth H, et al. Predictors of mortality in ARDS patients referred to a tertiary care centre: a pilot study. *Eur J Anaesthesiol*. 2006; 23(5):403–410. [PubMed: 16469204]
7. The 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(18):1301–1308. [PubMed: 10793162]
8. Cooke CR, Shah CV, Gallop R, et al. A simple clinical predictive index for objective estimates of mortality in acute lung injury. *Crit Care Med*. 2009; 37(6):1913–1920. [PubMed: 19384214]
9. Brower RG, Lanken PN, MacIntyre N, et al. Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome. *N Engl J Med*. 2004; 351(4):327–336. [PubMed: 15269312]
10. Lemeshow S, Klar J, Teres D. Outcome prediction for individual intensive care patients: useful, misused, or abused? *Intensive Care Med*. 1995; 21(9):770–776. [PubMed: 8847434]
11. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999; 130(6):515–524. [PubMed: 10075620]
12. Needham DM, Dennison CR, Dowdy DW, et al. Study protocol: The Improving Care of Acute Lung Injury Patients (ICAP) study. *Crit Care*. 2006; 10(1):R9. [PubMed: 16420652]
13. 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(3 Pt 1):818–824. [PubMed: 7509706]
14. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987; 40(5):373–383. [PubMed: 3558716]
15. Knaus WA, Draper EA, Wagner DP, et al. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985; 13(10):818–829. [PubMed: 3928249]
16. Ferreira FL, Bota DP, Bross A, et al. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA*. 2001; 286(14):1754–1758. [PubMed: 11594901]
17. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010; 21(1):128–138. [PubMed: 20010215]
18. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988; 44(3):837–845. [PubMed: 3203132]
19. Fan E, Needham DM. Deciding who to admit to a critical care unit. *BMJ*. 2007; 335(7630):1103–1104. [PubMed: 17975253]
20. Angus DC, Mira JP, Vincent JL. Improving clinical trials in the critically ill. *Crit Care Med*. 2010; 38(2):527–532. [PubMed: 19851096]

21. Cooke CR, Erickson SE, Watkins TR, et al. Age-, sex-, and race-based differences among patients enrolled versus not enrolled in acute lung injury clinical trials. *Crit Care Med.* 2010; 38(6):1450–1457. [PubMed: 20386308]
22. Phua J, Badia JR, Adhikari NK, et al. Has mortality from acute respiratory distress syndrome decreased over time?: A systematic review. *Am J Respir Crit Care Med.* 2009; 179(3):220–227. [PubMed: 19011152]
23. Rubenfeld GD, Caldwell E, Peabody E, et al. Incidence and outcomes of acute lung injury. *N Engl J Med.* 2005; 353(16):1685–1693. [PubMed: 16236739]
24. Vincent JL, Sakr Y, Ranieri VM. Epidemiology and outcome of acute respiratory failure in intensive care unit patients. *Crit Care Med.* 2003; 31(4 Suppl):S296–299. [PubMed: 12682455]
25. Sakr Y, Vincent JL, Reinhart K, et al. High tidal volume and positive fluid balance are associated with worse outcome in acute lung injury. *Chest.* 2005; 128(5):3098–3108. [PubMed: 16304249]
26. Umoh NJ, Fan E, Mendez-Tellez PA, et al. Patient and intensive care unit organizational factors associated with low tidal volume ventilation in acute lung injury. *Crit Care Med.* 2008; 36(5): 1463–1468. [PubMed: 18434907]

Figure. Receiver operating characteristics (ROC) curves

Using data from the external validation model, this Figure compares the ROC curves for the ARDSNet prediction model versus the Acute Physiology Assessment and Chronic Health Evaluation II score, respectively, for in-hospital mortality (area under curve = 0.67 vs. 0.70, $p=0.207$) in Panel A, and for 28-day mortality (0.64 vs. 0.70, $p=0.039$) in Panel B.

Table 1**Baseline Characteristics of Eligible Patients**

Variable	Status at Hospital Discharge ^a		P-value ^b
	Alive (n=274)	Dead (n=234)	
Age, years, median (IQR)	50 (41 – 60)	55 (46 – 67)	< 0.001
Male, %	55	57	0.569
Race, %			0.876
White	59	59	
Black	39	38	
Other	2	3	
Hospital days before ALI, median (IQR)	2 (1, 5)	3 (1, 7)	0.009
Primary ALI risk factor, %			< 0.001
Pneumonia	49	39	
Sepsis	20	43	
Aspiration	16	9	
Multiple Transfusions	4	3	
Other	11	5	
Severity of illness, mean (SD)			
APACHE II score at ICU admission	24 (8)	30 (9)	< 0.001
Charlson comorbidity	2.4 (2.6)	3.1 (2.3)	0.004
SOFA score at enrollment	8 (4)	12 (4)	< 0.001
Net fluid volume in first day after ALI ^c (ml)	1,779 (2,450)	3,427 (3,590)	< 0.001
Vasopressor use in first day after ALI, %	38	71	< 0.001
Respiratory physiology,			
Minute ventilation (L/min), mean (SD) ^d	11 (3)	13 (4)	< 0.001
Plateau pressure (mm Hg), mean (SD) ^d	21 (8)	25 (8)	< 0.001
PEEP, mm Hg, median (IQR) ^d	8 (5 – 10)	10 (5 – 12)	< 0.001
PaO ₂ /FiO ₂ ratio, mean (SD) ^e	202 (96)	149 (80)	< 0.001
pH, mean (SD) ^e	7.35 (0.08)	7.29 (0.13)	< 0.001
PaCO ₂ (mm Hg), mean (SD) ^e	43 (11)	43 (14)	0.957
Additional physiology and laboratory values			
Twenty-four hour urine output (mL) ^f	1,833 (1,309)	1,001 (1,321)	< 0.001
Glucose (mg/dL)			
Min glucose in first day after ALI (mg/dL) ^g	103 (36)	97 (41)	0.071
Max glucose in first day after ALI (mg/dL) ^g	165 (65)	186 (78)	0.002
Creatinine (mg/dL) ^h	1.9 (1.9)	2.6 (1.7)	< 0.001
Hematocrit, % ^h	27 (5)	27 (6)	0.620
Bilirubin 2 mg/dL, %	5	22	< 0.001

Abbreviations: ALI, acute lung injury; APACHE, Acute Physiology Assessment and Chronic Health Evaluation; h, hours; IQR, interquartile range; Max, Maximum; Min, Minimum; PaCO₂, partial pressure of arterial carbon dioxide; PaO₂, partial pressure of arterial oxygen; SOFA, Sequential Organ Failure Score.

^a% May not add to 100% due to rounding; all physiologic and laboratory data were obtained from first 24 hours after ALI

^bP-value was calculated using Pearson Chi-Squared, Fisher Exact, and Wilcoxon rank-sum tests as appropriate.

^cn=507; one patient excluded because of insufficient data

^ddata provided only for patients on modes of ventilation that permit calculation of these measures; n=448 for minute ventilation, n=405 for plateau pressure, n=452 for PEEP

^en=460 due to exclusions because no arterial blood gas available for calculation

^fn=496, 12 patients excluded because they did not remain in the ICU for 24 hours

^gn=505 and n=504 for min. and max. values, respectively due to missing data

^hn=492 and n=497 for creatinine and hematocrit, respectively, due to missing data

Table 2

Predicted and Observed In-Hospital Mortality

Total Points	Predicted Mortality from Model (%)	ARMA Trial (%)	ALVEOLI Trial (%)	Observed Mortality	
				External Validation Cohort (%)	
				Entire Cohort (n=494) ^a	Sensitivity Analysis (n=402) ^b
0	8.0	8.1	12.3	10.5	10.5
1	16.5	16.0	16.3	25.3	25.7
2	31.0	30.1	27.8	40.6	38.3
3	50.6	54.4	46.5	45.7	43.9
4+	70.0	60.0	66.7	64.6	58.0

^a 14 otherwise eligible patients did not have complete data for purposes of calculating the mortality prediction score and were excluded from analysis; total number of patients in categories 0, 1, 2, 3, and 4+ were 19, 79, 133, 116 and 147 respectively.

^b Sensitivity analysis evaluated a subgroup of the external validation cohort which was matched for the eligibility criteria of the clinical trial originally used to develop the ARDSNet prediction model; 12 patients were excluded from this analysis because of incomplete data; total number of patients in categories 0, 1, 2, 3, and 4+ were 19, 70, 115, 98 and 100 respectively.

Table 3

Predicted and Observed 28-Day Mortality

Total Points	Observed Mortality				
	Predicted Mortality from Model (%)	ARMA Trial (%)	ALVEOLI Trial (%)	External Validation Cohort (%)	
				Entire Cohort (n=494) ^a	Sensitivity Analysis (n=402) ^b
0	6.6	5.4	6.2	10.5	10.5
1	13.2	13.0	10.4	19.0	18.6
2	24.6	25.2	25.3	31.6	31.3
3	41.1	42.2	41.8	39.7	36.7
4+	60.0	55.0	53.3	50.3	45.0

^a 14 otherwise eligible patients did not have complete data for purposes of calculating the mortality prediction score and were excluded from analysis; total number of patients in categories 0, 1, 2, 3, and 4+ were 19, 79, 133, 116 and 147 respectively.

^b Sensitivity analysis evaluated a subgroup of the external validation cohort which was matched for the eligibility criteria of the clinical trial originally used to develop the ARDSNet prediction model; 12 patients were excluded from this analysis because of incomplete data; total number of patients in categories 0, 1, 2, 3, and 4+ were 19, 70, 115, 98 and 100 respectively.

Box 1

ARDS Network Prediction Model for Mortality in Acute Lung Injury

Variable ^a	POINTS		
	0	1	2
Age (yrs)	39	40–64	65
Bilirubin (mg/dL)	<2.0	2.0	----
Net 24-hr volume (in-out), mL	2500	>2500	----
Hematocrit, %	26	<26	----

^aThe prediction model score is calculated as the sum of the point score for each of the 4 variables