Published in final edited form as:

Crit Care Med. 2011 April ; 39(4): 711–717. doi:10.1097/CCM.0b013e318207ec3c.

Use of risk reclassification with multiple biomarkers improves mortality prediction in acute lung injury

Carolyn S. Calfee, MD, MAS

Department of Medicine, University of California, San Francisco, CA

Division of Pulmonary and Critical Care Medicine, Department of Anesthesia, University of California, San Francisco, CA

Cardiovascular Research Institute, San Francisco, CA

Lorraine B. Ware, MD

Department of Medicine, Division of Allergy, Pulmonary and Critical Care, Vanderbilt University, Nashville, TN

David V. Glidden, PhD

Department of Epidemiology and Biostatistics, University of California, San Francisco, CA

Mark D. Eisner, MD, MPH

Department of Medicine, University of California, San Francisco, CA

Division of Pulmonary and Critical Care Medicine, Department of Anesthesia, University of California, San Francisco, CA

Cardiovascular Research Institute, San Francisco, CA

Polly E. Parsons, MD

Department of Medicine, Division of Pulmonary and Critical Care Medicine, University of Vermont, Burlington, VT

B. Taylor Thompson, MD

Department of Medicine, Pulmonary and Critical Care Medicine Unit, and Biostatistics Unit, Massachusetts General Hospital, Boston, MA

Michael A. Matthay, MD

Department of Medicine, University of California, San Francisco, CA

Division of Pulmonary and Critical Care Medicine, Department of Anesthesia, University of California, San Francisco, CA

Cardiovascular Research Institute, San Francisco, CA

Blood, and Lung Institute Acute Respiratory Distress Syndrome Network National Heart

Abstract

Copyright © 2011 by the Society of Critical Care Medicine and Lippincott Williams & Wilkins

Address requests for reprints to: Carolyn S. Calfee, MD, MAS, University of California, San Francisco, Pulmonary and Critical Care Division, 505 Parnassus Avenue, Box 0111, San Francisco, CA 94143-0111. carolyn.calfee@ucsf.edu.

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

The network participants, clinical coordination center, National Heart, Lung, and Blood Institute staff, Steering Committee, Data and Safety Monitoring Board, and Protocol Review Committee are acknowledged in the Appendix.

Dr. Eisner is a full-time employee Genentech. The remaining have not disclosed any potential conflicts of interest.

Objective—Multiple single biomarkers have been associated with poor outcomes in acute lung injury; however, no single biomarker has sufficient discriminating power to clearly indicate prognosis. Using both derivation and replication cohorts, we tested novel risk reclassification methods to determine whether measurement of multiple plasma biomarkers at the time of acute lung injury diagnosis would improve mortality prediction in acute lung injury.

Design—Analysis of plasma biomarker levels and prospectively collected clinical data from patients enrolled in two randomized controlled trials of ventilator therapy for acute lung injury.

Setting—Intensive care units of university hospitals participating in the National Institutes of Health Acute Respiratory Distress Syndrome Network.

Patients—Subjects enrolled in a trial of lower tidal volume ventilation (derivation cohort) and subjects enrolled in a trial of higher vs. lower positive end-expiratory pressure (replication cohort).

Interventions—None.

Measurements and Main Results—The plasma biomarkers were intercellular adhesion molecule-1, von Willebrand factor, interleukin-8, soluble tumor necrosis factor receptor-1, and surfactant protein-D. In the derivation cohort ($n = 547$), adding data on these biomarkers to clinical predictors (Acute Physiology and Chronic Health Evaluation III score) at the time of study enrollment improved the accuracy of risk prediction, as reflected by a net reclassification improvement of 22% (95% confidence interval 13% to 32%; $p < .001$). In the replication cohort ($n = 500$), the net reclassification improvement was 17% (95% confidence interval 7% to 26%; $p < .001$). A reduced set of three biomarkers (interleukin-8, soluble tumor necrosis factor receptor-1, and surfactant protein-D) had nearly equivalent prognostic value in both cohorts.

Conclusions—When combined with clinical data, plasma biomarkers measured at the onset of acute lung injury can improve the accuracy of risk prediction. Combining three or more biomarkers may be useful for selecting a high-risk acute lung injury population for enrollment in clinical trials of novel therapies.

Keywords

acute respiratory distress syndrome; biological marker; pulmonary edema; risk prediction; risk reclassification

---

Over the past decade, multiple single plasma biomarkers have been associated with poor clinical outcomes in patients with acute lung injury (ALI) and its more severe form, the acute respiratory distress syndrome (ARDS) (1). These biomarker studies have provided insight into the pathogenesis, confirming the importance of endothelial injury (2), disordered coagulation and fibrinolysis (3), acute inflammation (4, 5), and lung epithelial injury in ALI/ARDS (6, 7). However, no single biomarker has been sufficiently strongly associated with outcomes to provide discriminating power for either diagnosis or prognosis. As a result, the role of plasma biomarkers in clinical and clinical research settings has been limited, despite the failure of clinical data alone to accurately predict outcomes.

At the same time, mortality from ALI/ARDS, particularly in clinical trials of novel therapies, continues to decline, likely as a result of improvements in supportive treatments. In the 1980s and 1990s, most studies reported mortality associated with ARDS of 40% to 60% (8, 9). More recently, Rubenfeld et al (10) found a mortality rate of 38.5% in a population-based cohort of patients in King County, WA. Even more striking is the decline in mortality in patients enrolled in clinical trials of novel therapies for ALI/ARDS (11); in the most recently presented trial from the National Heart, Blood, and Lung Institute's ARDS Network, the Albuterol for the Treatment of Acute Lung Injury trial, mortality had dropped to 21% overall (13). While this improvement in patient outcomes in ALI is gratifying, it also

creates a challenge for researchers designing clinical trials, because larger sample sizes are now required to demonstrate an effect on mortality. Selection of a higher risk subgroup of ALI patients may therefore be important in future trials of new therapies.

Predictive biomarkers have frequently been evaluated with receiver operating characteristic (ROC) analyses, which evaluate the discrimination of a test or marker (i.e., the accuracy with which the marker is associated with a dichotomous outcome) (14). Because improvements in the C-statistic (commonly termed the “area under the curve”) require large odds ratios with little overlap in values between cases and controls, sophisticated analyses have demonstrated that ROC-based approaches may be overly insensitive to the added value of a new marker. As an example, Cook and colleagues (14) demonstrated that such an approach would have led to discarding high-density lipoprotein, low-density lipoprotein, and total cholesterol as significant predictors of cardiovascular disease in the Women's Health Study. For this reason, novel methods of evaluating prognostic biomarkers that assess the biomarkers' relevant clinical value have been developed, including risk reclassification analysis (15). This analytic approach has been applied to cardiovascular diseases, for instance, to measure the additive value of measuring multiple plasma biomarkers in patients at risk for coronary artery disease (16), but risk reclassification has not been previously tested in ALI or in critical illness to our knowledge.

Given the demonstrated associations between single plasma biomarkers and prognosis in ALI, and our previous finding that data on multiple plasma biomarkers confer only a modest prognostic benefit using traditional ROC-based methods (17), we hypothesized that a panel of plasma biomarkers, each individually shown to be associated with poor outcomes in ALI/ARDS (2–5, 7, 18), would improve the accuracy of risk prediction when analyzed with novel risk reclassification methods (14, 16). To test the robustness of these findings, we conducted these analyses in both a derivation cohort and a replication cohort of patients with ALI/ARDS, drawn from two large randomized controlled trials of ventilator management.

METHODS

Clinical data and biological samples were obtained from patients enrolled in the National Heart, Blood, and Lung Institute ARDS Network's randomized controlled trials of lower tidal volume ventilation (12) and higher vs. lower positive end-expiratory pressure (19). Details of the original trials have been previously published in full. The lower tidal volume ventilation trial was used as the derivation cohort (n = 902; full biomarker data available for 547), and the higher vs. lower positive end-expiratory pressure trial was used as a replication cohort (n = 549; full biomarker data available for 500). All clinical data and biological samples used for this analysis were collected at the study baseline (prerandomization). Patients were followed until discharge home or death or for at least 90 days in both studies. Both trials were approved by the institutional review boards at each participating hospital; informed consent was obtained from the patients or surrogates at all but one hospital in the derivation cohort, where this requirement was waived. Additional information on the procedures for collecting and verifying clinical data in the original trials is available in the online supplement (see Supplemental Digital Content 1, <http://links.lww.com/CCM/A215>).

Assay Procedures

Plasma biomarkers were measured using enzyme-linked immunoassay techniques. Details of the methods used to perform the plasma measurements have been previously published (2–5, 7, 18).

Statistical Methods

We considered for inclusion in the model eight plasma biomarkers previously associated with poor clinical outcomes in ALI/ARDS and previously measured in both cohorts: surfactant protein-D (SP-D) (7), von Willebrand factor antigen (2), soluble intercellular adhesion molecule-1 (18), interleukin-6 and interleukin-8 (IL-6, IL-8) (4), soluble tumor necrosis factor receptor-1 (sTNFr-1) (5), plasminogen activator inhibitor-1 (3), and protein C (3). Box plots depicting the range of values for each biomarker in the derivation cohort are included in supplemental Figure E1 (see Supplemental Digital Content 1, <http://links.lww.com/CCM/A215>). Plasma measurements were made on the basis of plasma availability and have been previously included in other analyses (17, 20, 21). Patients with any missing biomarker data were excluded. To reduce the number of biomarkers included and ensure that each biomarker contributed independent and orthogonal information to the predictive model, we constructed a multivariable logistic regression model for prediction of death in the derivation cohort including all eight biomarkers (as well as the use of lower tidal volume ventilation) (supplemental Table E1; see Supplemental Digital Content 1, <http://links.lww.com/CCM/A215>); using manual stepwise backward selection, biomarkers with nonsignificant p values ($p > .05$) were sequentially removed with serial iterations of the likelihood ratio test. Given their non-normal distributions, all biomarker values were natural log-transformed for inclusion in regression models. Two patients had undetectable levels of one biomarker each (one for plasminogen activator inhibitor-1 and one for von Willebrand factor antigen); each was replaced by the value 1 to use log-transformed values.

We used the Acute Physiology and Chronic Health Evaluation (APACHE) III score as our clinical prediction model in light of recent data showing that APACHE scores predict clinical outcomes in ALI/ARDS patients with accuracy equivalent to that of a hand-selected group of clinical predictors (22). Details of the calculation of APACHE III scores have been previously published (23); components include both acute physiologic variables and information on comorbidities. In the derivation cohort, assignment to lower tidal volume ventilation was included as a covariate in all models, given its beneficial effect on outcomes; in the replication cohort, in which all patients were ventilated with lower tidal volumes, the level of positive end-expiratory pressure had no significant impact on mortality and was not included as a covariate.

Unlike cardiovascular disease, risk prediction guidelines intended to direct therapy and/or research strategy do not exist for ALI/ ARDS. Thus, we empirically selected risk cutoffs with the goal of selecting a group of patients at higher than average risk of death but not at such high risk that future research interventions would be unlikely to have benefit. We tested the ability of the clinical prediction model to accurately classify patients into one of three risk categories (<30% risk of death, 30% to 70% risk, >70% risk) on the basis of hospital survival at 90 days. We then added the selected biomarkers as predictors to the model. The accuracies of these two models (clinical prediction model alone vs. an expanded model including clinical and biomarker data) were compared using the risk reclassification approach of Pencina et al (15). In this approach, the proportion of subjects whose risk is more accurately classified using the expanded model is balanced against the proportion of subjects whose risk is less accurately classified with the expanded model, generating a net reclassification improvement and an associated p value (derived from a standard z score). The formula used to calculate the net reclassification improvement and a sample calculation are shown in supplemental Figure E2 (see Supplemental Digital Content 1, <http://links.lww.com/CCM/A215>). In addition, we calculated the integrated discrimination improvement (IDI), which evaluates reclassifications that occur at all levels of risk by treating predicted risk as a continuous variable (15). Both the net reclassification improvement and IDI are reported as a percent improvement, which compares the proportion of patients correctly classified in the new model with that correctly classified in

the old model. We assessed model calibration for the logistic regression model including biomarkers using the standardized mortality ratio, the Hosmer-Lemeshow test, and a graphical examination of observed vs. predicted risk, all of which demonstrated excellent calibration in both the derivation and replication cohorts. We also evaluated the incremental benefit to the C-statistic observed with the addition of the biomarkers to the clinical prediction model in both cohorts. Some of the results of these studies have been previously reported in the form of an abstract (24).

RESULTS

Clinical Characteristics of the Subjects

The clinical characteristics of the patients in the derivation and replication cohorts are shown in (Table 1). While the demographic characteristics of the two populations were fundamentally similar, the severity of illness and clinical outcomes differed significantly between the two cohorts. Specifically, patients enrolled in the replication cohort had a higher severity of illness, as measured by the APACHE III score, but better clinical outcomes when compared with patients in the derivation cohort.

Selection of Biomarkers for Inclusion

Biomarkers were selected for inclusion in the predictive model on the basis of their independent contributions to predicting death at 90 days in the derivation cohort, as described in more detail in the statistical methods. Plasminogen activator inhibitor-1, IL-6, and protein C were sequentially removed from the original model (supplemental Table E1; see Supplemental Digital Content 1, <http://links.lww.com/CCM/A215>) due to their lack of independent statistical contribution, leaving SP-D, IL-8, soluble intercellular adhesion molecule-1, von Willebrand factor antigen, and sTNF α -1 (Table 2).

Additive Value of Biomarkers Using Risk Reclassification

In the derivation cohort, the addition of data on the five selected biomarkers improved risk prediction significantly, compared with the clinical prediction model alone (Table 3). Among 183 patients who died, 41 were predicted to be at higher risk using the biomarker model, while 16 were predicted to be at lower risk. Among 359 patients who survived, 67 were reclassified into a lower risk category using the biomarker model, while 36 were reclassified into a higher risk category. The net reclassification improvement for the biomarker model, compared with a clinical prediction model including the APACHE III score and the use of lower tidal volume ventilation, was 22% (95% confidence interval 13% to 32%; $p < .001$). Calculation of the net reclassification improvement is shown in supplemental Figure E2 (see Supplemental Digital Content 1, <http://links.lww.com/CCM/A215>).

The IDI considers reclassifications that occur at all levels of risk by treating predicted risk as a continuous variable. In the derivation cohort, the IDI demonstrated significant improvement in risk classification across all levels of risk (IDI = 0.099; $p < .001$) (Fig. 1a).

Findings were similar in the replication cohort (supplemental Table E2 [see Supplemental Digital Content 1, <http://links.lww.com/CCM/A215>]; Fig. 1b). Specifically, among 132 patients who died, 26 were estimated to be at higher risk after reclassification, while 6 were estimated to be at lower risk. Among 362 patients who survived, 24 were estimated to be at lower risk, while 19 were estimated to be at higher risk. The net reclassification improvement was 17% (95% confidence interval 7% to 26%; $p < .001$), while the IDI was 0.039 ($p < .001$).

Sensitivity Analysis Using Different Risk Cutoffs

To test the robustness of these results, we conducted sensitivity analyses in which we evaluated the effect of altering the risk cutoffs. We tested cutoffs of 25% and 75%, as well as cutoffs of 33% and 66%. In both cases, the net reclassification improvement remained highly significant in both cohorts (supplemental Table E3; see Supplemental Digital Content 1, <http://links.lww.com/CCM/A215>). When more extreme risk cutoffs of 20% and 80% were selected, resulting in a more heterogeneous range of risks in the middle subgroup, the improvements in risk prediction were weaker.

Comparison With the C-Statistic

We also analyzed the discriminating accuracy of the clinical prediction model compared with the model including the biomarkers using a traditional, ROC curve approach (Table 4). In the derivation cohort, the addition of the panel of five biomarkers to the clinical prediction model significantly improved the C-statistic (p value for comparison of ROC curves of $<.001$). In the replication cohort, the improvement in the C-statistic did not meet criteria for statistical significance, although it was close to significant ($p = .051$).

Reduced Biomarker Model Using Risk Reclassification

To minimize the number of biomarkers needed and increase the practicality of the measurement, we tested the net reclassification improvement of all combinations of two or three biomarkers in the derivation cohort (supplemental Table E4; see Supplemental Digital Content 1, <http://links.lww.com/CCM/A215>). The best combination of three biomarkers was IL-8, SP-D, and sTNFr-1; this combination yielded a net reclassification improvement of 19% in the derivation cohort ($p < .001$) and 16% in the replication cohort ($p < .001$).

DISCUSSION

With mortality continuing to decline in clinical trials in ALI (25), new approaches to the selection of high-risk patients for enrollment in clinical trials may help reduce the sample size needed to test the effect of new therapies. We found that data on plasma levels of five biomarkers (soluble intercellular adhesion molecule-1, von Willebrand factor antigen, IL-8, SP-D, and sTNFr-1) significantly improved risk prediction when compared to a robust clinical prediction model. This improvement in risk prediction was demonstrated in both derivation and replication cohorts and was maintained in sensitivity analyses using different risk cutoffs designed to capture a moderate risk subgroup of patients. Reduction of the biomarker model to include only three biomarkers (IL-8, sTNFr-1, and SP-D) resulted in only a minimal decrement in accuracy. Of note, analysis of the predictive value of the biomarkers using risk reclassification methods was much more sensitive to changes in predictive accuracy than analysis using traditional ROC curve methods, a finding that has been demonstrated in other disease states (14, 16).

While the declining mortality rate in randomized controlled trials of ALI is cause for celebration for those who care for patients with lung injury, such improvements in outcome also generate serious consequences for prospective clinical trials in ALI. As an example, if a new therapy as effective as the use of lower tidal volume ventilation (relative risk reduction of 22%) were tested on a cohort of patients with a mortality rate similar to that of patients enrolled in the Albuterol for the Treatment of Acute Lung Injury trial (20%), 2,462 patients would need to be enrolled to have adequate statistical power to demonstrate an effect, nearly triple the number of patients enrolled in the original lower tidal volume trial (12). Increasingly large trials are more costly and require both more complex organizational effort and considerably longer time windows to complete. Of more concern, trials of smaller sample size then become increasingly underpowered and prone to type II error (that is,

missing a true positive effect). Finally, enrolling patients at low risk of death into trials of potentially risky new treatments may be ethically questionable if the risk of the novel therapy being tested exceeds its potential mortality benefits. Thus, selecting a higher risk group of patients for enrollment in ALI clinical trials is urgently needed on scientific, logistic, financial, ethical, and clinical grounds.

From a practical standpoint, how could a risk prediction strategy such as this be used to select patients for enrollment in clinical trials? As described above, the number of biomarkers measured could be reduced to three (IL-8, SP-D, and sTNFr-1) without a significant loss in predictive accuracy; these assays would require only 250 μ L of plasma if performed individually and markedly less if measured in a multiplex system. This quantity of sample could likely be obtained from excess plasma drawn for clinical purposes, thus allowing patients to be screened for eligibility without additional blood sampling. Once the biomarkers are measured, a process that could reasonably be performed in 6–8 hrs, the results could then be used in combination with APACHE score to calculate a predicted probability of death for each patient. A sample calculation is provided in supplemental Figure E3 (see Supplemental Digital Content 1, <http://links.lww.com/CCM/A215>). Patients at moderate risk of death could then be preferentially enrolled in a large-scale clinical trial of a higher risk novel therapy; alternatively, for phase I or equivalent trials, patients at higher risk of death could be preferentially enrolled, as is currently the practice in phase I oncology trials of new and potentially toxic agents. Furthermore, as more biomarkers are measured in routine clinical practice, the improvement in substantively predictive clinical prediction models such as APACHE III may become more clinically relevant.

Since no established categories for risk prediction exist in ALI, in contrast to coronary artery disease, our choice of risk cutoffs was derived from our observation of prior studies in ALI and the need to select a moderately high risk group of patients for clinical trials of future therapies. The risk cutoffs presented here were selected empirically to generate useful subgroups—specifically to enable the selection of a group of patients at moderate risk of mortality in whom the risk of death may be more modifiable than in those at either extreme of risk. However, as both the IDI and the data in Figure 1 demonstrate, meaningful reclassifications occurred at all levels of risk in both survivors and nonsurvivors and in both the derivation and replication cohorts. This finding underscores the result of our sensitivity analyses by demonstrating that the improvement in meaningful reclassification is robust to changes in the precise risk categories selected.

This study has important implications for future biomarker research in critically ill patients. ROC curve analysis is frequently used to test the utility of novel biomarkers for diagnosis and prognosis both in critical illness (acute kidney injury, ventilator-associated pneumonia) (26, 27) and in other disease states (28) and has considerable value in these analyses, in part due to its easy interpretability. As Cook and colleagues (14) elegantly demonstrated, however, this analytic approach may be insensitive to changes in prognostic accuracy, and its use could lead to discarding as useless biomarkers that have since been demonstrated to have major utility for both pathogenesis and prognosis (e.g., high-density lipoprotein cholesterol in coronary artery disease). Our analysis confirms that an ROC-based approach may be insensitive to major changes in predictive accuracy, as the contribution of biomarkers to risk prediction that was both statistically and clinically significant in the replication cohort (17%) was not clearly significant using ROC-based analysis. These findings suggest that alternatives to ROC-based analyses should be included in future studies of the prognostic contribution of novel biomarkers.

One of the strengths of this study is the availability of both a derivation and a replication cohort. While inclusion and exclusion criteria were similar for the two cohorts, mortality and

other important clinical outcomes such as ventilator-free days differed considerably between the two groups (12, 19), suggesting that the reclassification approach is robust to changes in the rate of outcomes in the population studied. Likewise, the detailed protocols and data standardization employed by the ARDS Network result in data that are methodologically consistent across both the derivation and validation cohorts, thus rendering differences in outcome potentially more meaningful. Another strength of the study is the diversity of lung injury patients included in the trial. Patients were drawn from multiple centers around the United States, represented a range of racial and ethnic groups, and were well-balanced for gender. Furthermore, the most common etiologies of ALI (sepsis, pneumonia, aspiration, trauma) were well represented in both cohorts.

We considered other clinical prediction models before selecting the APACHE score as the comparison group against which to judge the utility of the biomarker panel. Specifically, we considered using a hand-selected set of clinical predictors with demonstrated associations with mortality in the derivation cohort. However, we ultimately selected APACHE III scores as the optimal clinical predictive model for several reasons. First, APACHE III scores performed better in tests of discrimination in the derivation cohort than hand-selected clinical prediction models (data not shown). Second, Cooke and colleagues recently reported that APACHE scores perform equally well compared to carefully constructed multivariable models in patients with ALI in both population-based cohorts and in the Acute Respiratory Management in Acute Respiratory Distress Syndrome data set (this study's derivation cohort) (22). Third, APACHE IV (a recently developed update of the APACHE III scoring system) was the most accurate predictor of risk-adjusted intensive care unit mortality in a recently published large population-based cohort (29).

This study has some limitations. First, plasma biomarkers were not measured on all patients in both cohorts, although the number of patients with missing biomarker data in the replication cohort was small ($n = 49$). Plasma was unavailable in the majority of cases due to the fact that many patients in the derivation cohort were co-enrolled in trials of ketoconazole and lisofylline, which required measurement of plasma drug levels. There were no differences in mortality, ventilator-free days, APACHE III score, gender, race-ethnicity, $\text{PaO}_2/\text{FIO}_2$ ratio, or cause of ALI between subjects who had missing biomarker data and those who were included in the derivation cohort (data not shown). Second, we considered for inclusion in our models only those plasma biomarkers previously measured by our group in both the derivation and replication cohorts. Other plasma and urine biomarkers have prognostic value in patients with ALI, both in patients from the derivation cohort (6, 30, 31) and in patients from other ALI cohorts (32, 33). Due to the very limited volume of plasma and urine remaining from these cohorts, measurement of additional biomarkers in these patients was not feasible. Thus, it is possible that additional biomarkers could add additional predictive value beyond the improvements demonstrated with these five biomarkers. Third, patients enrolled in randomized controlled trials may not be representative of the broader population of patients with ALI/ARDS, which may limit the generalizability of our findings. Finally, close inspection of the predictive value of both the clinical and biomarker models reveals that both predictive models skew toward predicting lower risk. Specifically, some patients who died were predicted by both models in both cohorts to be low risk ($<30\%$), while relatively few patients were predicted by either model to be high risk ($>70\%$). Similarly, in the example calculation provided in the online supplement (see Supplemental Digital Content 1, <http://links.lww.com/CCM/A215>), while the biomarker model improves the accuracy of mortality prediction, the predicted risk of death for the patient (who ultimately dies) remains lower than 50%. This discrepancy suggests that one or several important contributors to mortality have not been included in either predictive model, and thus, there remains room for improvement in both the clinical prediction model and the risk reclassification models we have developed.

In summary, the addition of data on five plasma biomarkers to previously validated clinical prediction models significantly improved risk prediction in two cohorts of patients with ALI, both a derivation and a replication cohort. In an era of declining mortality and increasingly large sample size requirements for clinical trials in ALI (25), the use of selected plasma biomarkers to risk stratify patients before enrollment in trials of novel therapies should be further investigated.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported, in part, by contracts National Heart, Lung, and Blood Institute grant NO1-HR 46054, 46055, 46056, 46057, 46058, 46059, 46060, 46061, 46062, 46063, and 46064 with the National Heart, Lung, and Blood Institute, Bethesda, MD. Dr. Calfee was supported by grant HL090833, by the Flight Attendant Medical Research Institute, Miami, FL, and by grant KL2RR024130 from the National Center for Research Resources, a component of the National Institutes of Health, Bethesda, MD. Dr. Matthay was supported by National Heart, Lung, and Blood Institute grant HL 51856. Dr. Ware was supported by grant HL081332.

APPENDIX

National Institutes of Health, National Heart, Lung, and Blood Institute ARDS Network.

Network Participants

Cleveland Clinic Foundation, Herbert P. Wiedemann, MD,* Alejandro C. Arroliga, MD, Charles J. Fisher Jr, MD, John J Komara Jr, BA, RRT, Patricia Periz-Trepichio, BS, RRT; *Denver Health Medical Center*, Polly E. Parsons, MD, *Denver VA Medical Center*, Carolyn Welsh, MD; *Duke University Medical Center*, William J. Fulkerson Jr, MD,* Neil MacIntyre, MD, Lee Mallatratt, RN, Mark Sebastian, MD, John Davies, RRT, Elizabeth Van Dyne, RN, Joseph Govert, MD; *Johns Hopkins Bayview Medical Center*, Jonathan Sevransky, MD, Stacey Murray, RRT; *Johns Hopkins Hospital*, Roy G. Brower, MD, David Thompson, MS, RN, Henry E. Fessler, MD; *LDS Hospital*, Alan H. Morris, MD,* Terry Clemmer, MD, Robin Davis, RRT, James Orme Jr, MD, Lindell Weaver, MD, Colin Grissom, MD, Frank Thomas, MD, Martin Gleich, MD (posthumous); *McKay-Dee Hospital*, Charles Lawton, MD, Janice D'Hulst, RRT; *MetroHealth Medical Center of Cleveland*, Joel R. Peerless, MD, Carolyn Smith, RN; *San Francisco General Hospital Medical Center*, Richard Kallet, MS, RRT, John M. Luce, MD; *Thomas Jefferson University Hospital*, Jonathan Gottlieb, MD, Pauline Park, MD, Aimee Girod, RN, BSN, Lisa Yannarell, RN, BSN; *University of California, San Francisco*, Michael A. Matthay, MD,* Mark D. Eisner, MD, MPH, Brian Daniel, RCP, RRT; *University of Colorado Health Sciences Center*, Edward Abraham, MD,* Fran Piedalue, RRT, Rebecca Jagusch, RN, Paul Miller, MD, Robert McIntyre, MD, Kelley E. Greene, MD; *University of Maryland*, Henry J. Silverman, MD,* Carl Shanholtz, MD, Wanda Corral, BSN, RN, *University of Michigan*, Galen B. Toews, MD,* Deborah Arnoldi, MHSA, Robert H. Bartlett, MD, Ron Dechert, RRT, Charles Watts, MD; *University of Pennsylvania*, Paul N. Lanken, MD,* Harry Anderson III, MD, Barbara Finkel, MSN, RN, C. William Hanson III, MD; *University of Utah Hospital*, Richard Barton, MD, Mary Mone, RN; *University of Washington/Harborview Medical Center*, Leonard D. Hudson, MD,* Greg Carter, RRT, Claudette Lee Cooper, RN, Annemieke Hiemstra, RN, Ronald V. Maier, MD, Kenneth P. Steinberg, MD; *Utah Valley Regional Medical Center*, Tracy Hill, MD, Phil Thaut, RRT; *Vanderbilt University*, Arthur P. Wheeler, MD,* Gordon Bernard, MD,* Brian Christman, MD, Susan Bozeman, RN, Linda Collins, Teresa Swope, RN, Lorraine B. Ware, MD.

Clinical Coordinating Center

Massachusetts General Hospital, Harvard Medical School, David A. Schoenfeld, PhD,* B. Taylor Thompson, MD, Marek Ancukiewicz, PhD, Douglas Hayden, MA, Francine Molay, MSW, Nancy Ringwood, BSN, RN, Gail Wenzlow, MSW, MPH, Ali S. Kazerooni, BS.

National Heart, Blood, and Lung Institute Staff

Dorothy B. Gail, PhD, Andrea Harabin, PhD,* Pamela Lew, Myron Waclawiw, PhD.

*Steering Committee

Gordon R. Bernard, MD, Chair; Principal Investigator from each center as indicated by an asterisk.

Data and Safety Monitoring Board

Roger G. Spragg, MD, Chair, James Boyett, PhD, Jason Kelley, MD, Kenneth Leeper, MD, Marion Gray Secundy, PhD, Arthur Slutsky, MD.

Protocol Review Committee

Joe G. N. Garcia, MD, Chair, Scott S. Emerson, MD, PhD, Susan K. Pingleton, MD, Michael D. Shasby, MD, William J. Sibbald, MD.

REFERENCES

1. Ware LB. Prognostic determinants of acute respiratory distress syndrome in adults: Impact on clinical trial design. *Crit Care Med.* 2005; 33:S217–S222. [PubMed: 15753731]
2. Ware LB, Eisner MD, Thompson BT, et al. Significance of von Willebrand factor in septic and nonseptic patients with acute lung injury. *Am J Respir Crit Care Med.* 2004; 170:766–772. [PubMed: 15201135]
3. Ware LB, Matthay MA, Parsons PE, et al. Pathogenetic and prognostic significance of altered coagulation and fibrinolysis in acute lung injury/acute respiratory distress syndrome. *Crit Care Med.* 2007; 35:1821–1828. [PubMed: 17667242]
4. Parsons PE, Eisner MD, Thompson BT, et al. Lower tidal volume ventilation and plasma cytokine markers of inflammation in patients with acute lung injury. *Crit Care Med.* 2005; 33:1–6. discussion 230–232. [PubMed: 15644641]
5. Parsons PE, Matthay MA, Ware LB, et al. Elevated plasma levels of soluble TNF receptors are associated with morbidity and mortality in patients with acute lung injury. *Am J Physiol Lung Cell Mol Physiol.* 2005; 288:L426–L431. [PubMed: 15516488]
6. Calfee CS, Ware LB, Eisner MD, et al. Plasma receptor for advanced glycation end products and clinical outcomes in acute lung injury. *Thorax.* 2008; 63:1083–1089. [PubMed: 18566109]
7. Eisner MD, Parsons P, Matthay MA, et al. Plasma surfactant protein levels and clinical outcomes in patients with acute lung injury. *Thorax.* 2003; 58:983–988. [PubMed: 14586055]
8. Fowler AA, Hamman RF, Good JT, et al. Adult respiratory distress syndrome: Risk with common predispositions. *Ann Intern Med.* 1983; 98:593–597. [PubMed: 6846973]
9. Doyle RL, Szaflarski N, Modin GW, et al. Identification of patients with acute lung injury. Predictors of mortality. *Am J Respir Crit Care Med.* 1995; 152:1818–1824. [PubMed: 8520742]
10. Rubenfeld GD, Caldwell E, Peabody E, et al. Incidence and outcomes of acute lung injury. *N Engl J Med.* 2005; 353:1685–1693. [PubMed: 16236739]
11. Erickson SE, Martin GS, Davis JL, et al. Recent trends in acute lung injury mortality: 1996–2005. *Crit Care Med.* 2009; 37:1574–1579. [PubMed: 19325464]

12. 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; 342:1301–1308. [PubMed: 10793162]
13. Matthay M, Brower R, Thompson B, et al. Randomized, placebo-controlled trial of an aerosolized beta-2 adrenergic agonist (albuterol) for the treatment of acute lung injury. *Am J Respir Crit Care Med*. 2009; 179:A2166.
14. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007; 115:928–935. [PubMed: 17309939]
15. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, et al. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Stat Med*. 2008; 27:157–172. discussion 207–112. [PubMed: 17569110]
16. Zethelius B, Berglund L, Sundström J, et al. Use of multiple biomarkers to improve the prediction of death from cardiovascular causes. *N Engl J Med*. 2008; 358:2107–2116. [PubMed: 18480203]
17. Ware LB, Koyama T, Billheimer DD, et al. Prognostic and pathogenetic value of combining clinical and biochemical indices in patients with acute lung injury. *Chest*. 2010; 137:288–296. [PubMed: 19858233]
18. Calfee CS, Eisner MD, Parsons PE, et al. Soluble intercellular adhesion molecule-1 and clinical outcomes in patients with acute lung injury. *Intensive Care Med*. 2009; 35:248–257. [PubMed: 18670758]
19. 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; 351:327–336. [PubMed: 15269312]
20. Calfee CS, Eisner MD, Ware LB, et al. Trauma-associated lung injury differs clinically and biologically from acute lung injury due to other clinical disorders. *Crit Care Med*. 2007; 35:2243–2250. [PubMed: 17944012]
21. Ware LB, Koyama T, Billheimer DD, et al. Prognostic and pathogenetic value of combining clinical and biochemical indices in patients with acute lung injury. *Chest*. 2010; 137:268–296.
22. Cooke CR, Kahn JM, Caldwell E, et al. Predictors of hospital mortality in a population-based cohort of patients with acute lung injury. *Crit Care Med*. 2008; 36:1412–1420. [PubMed: 18434894]
23. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*. 1991; 100:1619–1636. [PubMed: 1959406]
24. Calfee CS, Ware LB, Glidden DV, et al. Combining multiple plasma biomarkers improves risk prediction in ALI in both derivation and replication cohorts. *Am J Respir Crit Care Med*. 2009; 179:A4626.
25. Spragg RG, Bernard GR, Checkley W, et al. Beyond mortality: Future clinical research in acute lung injury. *Am J Respir Crit Care Med*. 2010; 181:1121–1127. [PubMed: 20224063]
26. Molitoris BA, Melnikov VY, Okusa MD, et al. Technology insight: Biomarker development in acute kidney injury—What can we anticipate? *Nat Clin Pract Nephrol*. 2008; 4:154–165. [PubMed: 18227821]
27. Luyt CE, Guérin V, Combes A, et al. Procalcitonin kinetics as a prognostic marker of ventilator-associated pneumonia. *Am J Respir Crit Care Med*. 2005; 171:48–53. [PubMed: 15447947]
28. Wang TJ, Gona P, Larson MG, et al. Multiple biomarkers for the prediction of first major cardiovascular events and death. *N Engl J Med*. 2006; 355:2631–2639. [PubMed: 17182988]
29. Kuzniewicz MW, Vasilevskis EE, Lane R, et al. Variation in ICU risk-adjusted mortality: Impact of methods of assessment and potential confounders. *Chest*. 2008; 133:1319–1327. [PubMed: 18403657]
30. McClintock DE, Ware LB, Eisner MD, et al. Higher urine nitric oxide is associated with improved outcomes in patients with acute lung injury. *Am J Respir Crit Care Med*. 2007; 175:256–262. [PubMed: 17082495]
31. McClintock DE, Starcher B, Eisner MD, et al. Higher urine desmosine levels are associated with mortality in patients with acute lung injury. *Am J Physiol Lung Cell Mol Physiol*. 2006; 291:L566–L571. [PubMed: 16698854]

32. Bajwa EK, Boyce PD, Januzzi JL, et al. Biomarker evidence of myocardial cell injury is associated with mortality in acute respiratory distress syndrome. *Crit Care Med.* 2007; 35:2484–2490. [PubMed: 18084839]
33. Karpaliotis D, Kirtane AJ, Ruisi CP, et al. Diagnostic and prognostic utility of brain natriuretic peptide in subjects admitted to the ICU with hypoxic respiratory failure due to noncardiogenic and cardiogenic pulmonary edema. *Chest.* 2007; 131:964–971. [PubMed: 17426196]

Figure 1.

Risk reclassification of all patients enrolled in derivation (*a*) and replication (*b*) cohorts. The figure represent mortality prediction in the clinical model alone (*x*-axis) vs. the clinical model plus biomarkers (*y*-axis); patients who died and patients who did not die are represented separately, since the direction of improved classification differs between these two groups. The *line of equivalence* is marked in *blue*; *green dots* represent subjects with more accurate risk prediction in the biomarker model, and *red dots* represent subjects with less accurate risk prediction in the biomarker model. Patients above the *line of equivalence* had higher risk predicted from the biomarker model than the clinical prediction model (APACHE III plus lower tidal volume ventilation); patients below the *line of equivalence* had higher risk predicted from the clinical prediction model than the biomarker model.

Table 1

Clinical characteristics of patients included from derivation and replication cohorts

Characteristic	Derivation Cohort (ARMA, n = 547)	Replication Cohort (ALVEOLI, n = 500)	p
Age, yrs	50 ± 16	51 ± 17	.46
Female gender	320 (59)	273 (55)	.20
Race-ethnicity			.80
White	412 (75)	381 (76)	
Black	83 (15)	65 (13)	
Latino	32 (6)	30 (6)	
Asian	13 (2)	18 (4)	
Other	7 (1)	6 (1)	
Cause of acute lung injury			
Sepsis	138 (25)	110 (22)	
Trauma	53 (10)	42 (8)	.47
Pneumonia	206 (38)	201 (40)	
Transfusions	16 (3)	75 (15)	
Aspiration	79 (14)	75 (15)	
Acute Physiology and Chronic Health Evaluation III score	82 ± 29	94 ± 32	<.001
PaO ₂ /FiO ₂ ratio	132 ± 61	129 ± 58	.37
Lower tidal volume ventilation, %	49	100	<.001
Mortality at 90 days, %	34	27	.01
Ventilator-free days	12 (0, 22)	18 (0, 24)	<.001
Organ-failure-free days	15 (0, 25)	19 (4, 25)	.02

ARMA, Acute Respiratory Management in Acute Respiratory Distress Syndrome; ALVEOLI, Assessment of Low tidal Volume and elevated End-expiratory volume to Obviate Lung Injury.

Data expressed as mean ± SD, n (%), or median (interquartile range) as appropriate. Excluded patients were missing one or more biomarker measurements.

Table 2

Selection of biomarkers for inclusion in the predictive model (derivation cohort)

Biomarker	Odds Ratio for Death at 90 days (95% Confidence Interval)	p
Surfactant protein-D	1.55 (1.27, 1.88)	<.001
Interleukin-8	1.36 (1.15, 1.62)	<.001
Soluble tumor necrosis factor receptor-1	1.58 (1.20, 2.09)	.001
Von Willebrand factor	1.57 (1.16, 2.12)	.003
Soluble intercellular adhesion molecule-1	0.74 (0.59, 0.95)	.02

The table depicts results from a multivariable logistic regression with the outcome of death at 90 days. Odds ratios represent the changes in the odds of death per one natural log increase in the biomarker level. Use of lower tidal volume ventilation was also included as a predictor in this model, since the data were derived from the Acute Respiratory Distress Syndrome Network's trial of lower tidal volume ventilation, as was the Acute Physiology and Chronic Health Evaluation III score.

Table 3

Reclassification using the clinical predictor plus biomarker model compared with the clinical prediction model alone (derivation cohort)

	Model With Clinical Predictors and Biomarkers			
Model With Clinical Predictors Alone	<30% Risk	30% to 70% Risk	>70% Risk	Total Number
In 183 patients who died				
<30% risk	32	15	1	48
30% to 70% risk	16	93	25	134
>70% risk	0	0	1	1
Total number	48	108	27	183
In 359 patients who did not die				
<30% to 70% risk	167	32	0	199
30% to 70% risk	66	88	4	158
>70% risk	1	0	1	2
Total number	234	120	5	359

Clinical predictors: Acute Physiology and Chronic Health Evaluation III score, use of lower tidal volume ventilation. Net reclassification improvement of 22% ($p < .001$). Five patients not included due to missing data on the Acute Physiology and Chronic Health Evaluation score.

Table 4

Receiver operating characteristic curve analysis comparing the performance of the model with clinical predictors alone vs. the model with clinical predictors plus biomarkers

Model	Area Under Curve	<i>p</i> for Comparison
Derivation cohort clinical prediction model	0.680	<.001
Derivation cohort clinical plus biomarkers model	0.747	
Replication cohort clinical prediction model	0.767	.051
Replication cohort clinical plus biomarker model	0.793	