



Published in final edited form as:

*Crit Care Med.* 2011 December ; 39(12): 2645–2651. doi:10.1097/CCM.0b013e3182266779.

## A simple classification model for hospital mortality in patients with acute lung injury managed with lung protective ventilation

Lisa M Brown, MD, MAS<sup>1,2</sup>, Carolyn S Calfee, MD, MAS<sup>2,3</sup>, Michael A Matthay, MD<sup>2,3</sup>, Roy G Brower, MD<sup>4</sup>, B Taylor Thompson, MD<sup>5</sup>, and William Checkley, MD, PhD<sup>4,6</sup> for the National Institutes of Health Acute Respiratory Distress Syndrome Network Investigators

<sup>1</sup>Department of Surgery, University of California, San Francisco, San Francisco, CA

<sup>2</sup>Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA

<sup>3</sup>Department of Medicine, Pulmonary and Critical Care Division, University of California, San Francisco, San Francisco, CA

<sup>4</sup>Division of Pulmonary and Critical Care, School of Medicine, Johns Hopkins University, Baltimore, MD

<sup>5</sup>Pulmonary and Critical Care Unit, Massachusetts General Hospital, Boston, MA

<sup>6</sup>Program in Global Disease Epidemiology and Control, Department of International Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD

### Abstract

**Objective**—Despite improvements in the care of critically-ill patients, hospital mortality for acute lung injury remains high at approximately 40%. We developed a classification rule to stratify mechanically ventilated patients with acute lung injury according to hospital mortality and compared this rule to the APACHE III prediction.

**Patients**—We used data of 2022 participants in ARDS Network trials to build a classification rule based on 54 variables collected prior to randomization.

**Design**—We used a classification tree approach to stratify patients according to hospital mortality using a training subset of 1800 participants, and estimated expected prediction errors using tenfold cross validation. We validated our classification tree using a subset of 222 participants not included in model building, and calculated areas under the receiver operating characteristic curves (AUCs).

**Measurements and Main Results**—We identified combinations of age (>63 years), BUN (>15 mg/dl), shock, respiratory rate (>21 breaths/minute), and minute ventilation (>13.9 L/minute) as important predictors of hospital mortality at 90 days. The classification tree had a similar expected prediction error in the training set (28% versus 26%;  $p=0.18$ ) and AUC in the validation set (0.71 versus 0.73;  $p=0.71$ ) as did a model based on APACHE III.

**Conclusions**—Our tree-based classification rule performed similarly to APACHE III in stratifying patients according to hospital mortality, is simpler to use, contains risk factors that may

---

Correspondence: William Checkley, MD, PhD, Assistant Professor, Division of Pulmonary and Critical Care, School of Medicine, Johns Hopkins University, Baltimore, MD 21212, Telephone: (443) 287-8741, wcheckl@jhmi.edu, Fax: (410) 955-0036.

**Contributions:** Conception and design (LB, WC); Analysis and interpretation: (LB, WC); Drafting the manuscript for important intellectual content (LB, CC, MM, TT, RB, WC),.

Dr. Calfee consulted as a member on the Medical Advisory Boards of Ikaria and GlaxoSmithKline. Dr. Thompson holds a National Institutes of Health contract to support ARDS Net Research. The remaining authors have not disclosed any potential conflicts of interest.

be specific to acute lung injury, and identified minute ventilation as a potential novel predictor of death in patients with acute lung injury.

---

## INTRODUCTION

Acute lung injury (ALI) affects approximately 190,000 patients and is responsible for two million critical care days and 75,000 deaths in the United States each year (1). 7% of patients admitted to an intensive care unit and 11 to 23 percent of patients with respiratory failure have ALI (2). Despite improvements in critical care over the last several decades (3–6), mortality for this syndrome remains high at approximately 40% (1).

No accurate classification system specific for ALI exists to date. Several studies have attempted to identify predictors of mortality in patients with ALI to improve prognostication (7–18). Age, multiple organ failure, comorbidities, and comprehensive scoring systems such as APACHE III and SAPS 3 (19–20) have been frequently used to predict mortality. These approaches, however, were designed for use in the general population of critically ill patients and not specifically for patients with ALI. Some researchers have identified measures of lung injury severity such as minute ventilation (7), oxygenation index (11, 21), dead-space fraction (14),  $\text{PaO}_2/\text{FiO}_2$  (7, 12, 15) and plateau pressure (22, 23) as independent predictors of mortality, whereas others have identified only non-pulmonary predictors (8–10, 16). In addition, scoring systems use numerous predictors to calculate a composite score to predict mortality (19–20). These scoring systems, while relatively accurate, require many data inputs and are cumbersome.

All of the previously mentioned studies used logistic regression methods for classification (7–18). Given the complexity of critically ill patients, there are many potential predictors of mortality, and numerous interactions may exist between these predictors. When using logistic regression, investigators must decide which predictors and interactions to include in the model. Even for experts, it can be difficult to conceptualize which interactions may be clinically important.

A classification tree is a statistical learning approach that can be used to develop prediction models (24). This approach considers all potential predictors and their interactions to classify participants according to an outcome. The predictors are used to recursively partition the data into subsets according to a set of decision criteria. Classification trees may provide an advantage over traditional regression approaches by identifying hard to detect interactions. They have been successfully used to predict a number of different outcomes including the risk of major complications in patients who present to the emergency department with chest pain (25), the risk of mortality in patients hospitalized with heart failure (26), and the use of coronary angiography and revascularization procedures after thrombolysis for acute myocardial infarction (27). We sought to build a classification tree that stratified a cohort of ALI patients according to their risk of hospital mortality.

## METHODS

We used data of all participants enrolled in clinical trials of the Acute Respiratory Distress Syndrome (ARDS) Network between 1996 and 2005 (3, 4, 28). Same inclusion criteria and similar exclusion criteria were used across included trials (Online Supplement). Our primary outcome was hospital mortality at 90 days. We chose this outcome because, in critically ill patients, the risk of mortality has been shown to extend well beyond 30 days (29). We included 54 predictors for which data were collected prior to randomization (Online Supplement). We defined shock as use of vasopressors. We did not include year of

enrollment in our analysis because it was not an important predictor of hospital mortality at 90 days.

## Biostatistical Methods

Our objective was to build a classification tree according to the risk of hospital mortality at 90 days (24). We then determined how our model performed against a logistic regression model of APACHE III score on hospital mortality at 90 days.

We used a conditional inference tree to develop our prediction rule (30, 31). To build our classification tree, we used a three-step approach. First, we examined pairwise correlations to identify highly correlated predictors. In this first step, we reduced the number of predictors to 49. Second, we used a random forest with 500 classification trees to identify the ten most important predictors (32). By creating numerous classification trees, the random forest indicates which predictors are used most frequently to partition the data with the goal of optimizing discrimination between participants who died in the hospital at 90 days and those who were alive. The random forest produces a statistical measure known as variable importance for each predictor included in the model, which is scaled against the variable with the highest degree of discrimination. Third, we built a classification tree using the ten predictors with the highest variable importance. We classified participants as dead in the hospital at 90 days if the predicted mortality was  $\geq 0.5$  and alive if otherwise.

We used 1800 participants for the training set (approximately 90% of all the data) and 222 participants (approximately 10%) as a test set. We used ten-fold cross-validation to identify the stopping criterion that minimized the expected prediction error (EPE) of our model. The EPE is a statistical measure of how well the predicted outcomes agree with the observed outcomes. We compared the EPE of the tree model with the EPE for the logistic regression model of APACHE III score using a t-test. We validated our classification tree using the test set and compared the areas under the ROC curve with that of the APACHE III score using a Mann-Whitney test (33). See Online Supplement for additional details.

We used the Cochran-Armitage test for trend (34) and logistic regression to determine if the proportion of patients who died in the hospital at 90 days decreased over time between 1996 and 2005. We used R ([www.r-project.org](http://www.r-project.org)) for statistical analysis.

This study was approved by the Internal Review Board of the Johns Hopkins School of Medicine, Baltimore, MD.

## RESULTS

### Descriptive statistics

2022 participants received lung protective ventilation. We provide summary statistics for all 49 predictors in Table 1. Median age was 49 years and 50% were male. The most common causes of ALI were pneumonia (41%), sepsis (24%) and aspiration (15%). 33% of participants required vasopressors before randomization. Hospital mortality at 90 days did not change between 1996 and 2005 (Figure 1;  $p=0.31$ ). We summarized differences in baseline characteristics across trials in Table 2.

### Relative importance of predictors

The ten most important variables, in decreasing order of importance, were age, blood urea nitrogen (BUN), barotrauma, albumin, cancer, bilirubin, total respiratory rate, minute ventilation, shock and urine output (Figure 2). We assigned age a relative variable importance of 100% and compared all other predictors with age.

### Expected prediction error from cross-validation

The EPE for our classification tree was 27.8%. That is, our classification tree misclassified participants who died in the hospital at 90 days as survivors and vice versa approximately one quarter of the time. However, the EPE for our classification tree was similar to that of a model based on the APACHE III score (Table 3,  $p=0.18$ , t-test).

### Classification model

Our final classification tree identified five predictors to stratify participants according to hospital mortality at 90 days (Figure 3). Age was the first split criterion. Participants aged  $>63$  years had a higher hospital mortality than those who were aged  $\leq 63$  years (48% versus 22%, respectively). Total respiratory rate and minute ventilation further split participants after age. For patients aged  $\leq 63$  years, the absolute difference in hospital mortality for participants with a minute ventilation  $>13.9$  L/minute versus  $\leq 13.9$  L/minute was 15% (32% versus 17%, respectively). Among those  $\leq 63$  years of age and with a minute ventilation  $\leq 13.9$  L/minute, the difference in hospital mortality with a BUN  $>15$  mg/dl versus  $\leq 15$  mg/dl was 13% (23% versus 10%, respectively). The presence of shock provided further discrimination in two instances. At those splits, the difference in hospital mortality between patients with shock versus those without shock was approximately two to three fold. Among those  $>63$  years of age, total respiratory rate was better than minute ventilation at partitioning participants age according to hospital mortality.

### Areas under the curves on the test set

We then applied our classification tree to the 222 participants who were placed aside as a test set. We did not find a difference between the areas under the ROC curves for the prediction model based on the classification tree and the logistic regression model based on the APACHE III score (Figure 4;  $p=0.71$ ; Mann-Whitney test).

## DISCUSSION

Using a classification tree approach, we derived a simple rule to stratify ALI patients according to the risk of hospital mortality by age, minute ventilation or respiratory rate, BUN and shock. Our classification tree performed as well at discriminating between survivors and non-survivors as did a model based on APACHE III, but is simpler to use and does not require any calculations. It also identified minute ventilation as a potential novel marker of disease severity in ALI.

Age was the most important predictor of mortality. This finding is consistent with previous studies of mortality in ALI patients (7–8, 10, 12–13, 17, 35). Age is also a component of comprehensive scoring systems for critically-ill patients (19, 20). Minute ventilation and respiratory rate were also important predictors. Minute ventilation incorporates both tidal volume and respiratory rate; however, since our participants received lung protective ventilation with low tidal volumes, respiratory rate may have contributed more than tidal volume to the importance of minute ventilation for predicting mortality. A high minute ventilation or respiratory rate may indicate either severe respiratory failure or an attempt to compensate for acidosis secondary to hypoperfusion or to an elevated pulmonary dead space. Severe respiratory failure is the cause of death in  $<20\%$  of ALI patients (35, 36). However, sepsis with multiple organ failure is the most common cause of death (35, 36). Therefore, the importance of minute ventilation or respiratory rate for predicting mortality may be associated with hypoperfusion, metabolic acidosis and elevated pulmonary dead space. However, we cannot ascertain which mechanisms led to an increased minute ventilation nor ascertain the prognostic importance of these mechanisms. Additional

mechanisms that may have lead to increased minute ventilation in ALI patients include anxiety, pain, and setting the respiratory rate at a higher level than needed.

Shock before randomization was also an important predictor of death. This is consistent with a recent finding by our group in which mortality was 31% in those with shock versus 12% in those without shock (37). The most common cause of elevated BUN is decreased renal function. One study demonstrated that the mortality rate of those who develop acute kidney injury (defined as a 50% rise in creatinine) was 58% compared with 28% in those who do not develop acute kidney injury (38). However, BUN may also be elevated if renal excretion of urea is temporarily impaired due to hypoperfusion or shock and highlights the importance of shock as a predictor of ALI mortality. This is consistent with prior studies that have demonstrated higher mortality in patients with ALI from sepsis compared to ALI from other causes (35, 36, 41, 42).

Our analytical approach identified a different set of predictors from that of a previously proposed model (39). In that study, the investigators conducted a retrospective analysis of three multicenter clinical studies in patients with ALI, and identified predictors of death or ventilator dependence from variables prospectively recorded during the first three days of mechanical ventilation. They found that age, oxygenation index, and cardiovascular failure three days after intubation predicted death or prolonged mechanical ventilation (39). Additional markers of nonpulmonary organ failures such as creatinine, platelet count, bilirubin, and Glasgow Coma Scale score were not important. In contrast, our model identified both blood urea nitrogen and bilirubin as important predictors of mortality. There are several reasons why our findings may differ. First, previous the study used observational data to derive their predictive model whereas we used data from clinical trials collected prior to randomization. Baseline differences persist when using data of patients enrolled into clinical trials because once randomization occurs patients are treated according to a standardized protocol. On the other hand, observational studies are subject to variation in clinical practice that affects patient outcomes. Therefore, when predicting mortality using baseline patient data, data from clinical trials may lead to an unbiased estimation when compared to data from observational studies. Second, the previous investigators used logistic regression, whereas we used classification trees. Third, our analysis uses five times more patients in the derivation cohort to build the model than did the previous investigators, but our validation cohort was substantially smaller. Finally, the previous study only used data of ALI patients who were alive and mechanically ventilated on day three, whereas we used baseline data of all enrolled patients. By excluding those patients who died prior to day three, the previous investigators may have overlooked important predictors of early mortality.

Our classification model may also be used to stratify ALI patients according to the risk of hospital mortality. A recent study developed a simple prediction score for mortality in patients with ALI that included age, bilirubin, 24-hour fluid balance, and hematocrit (8). While the ability of this previous score to discriminate between survivors and non-survivors was not as good as APACHE III, the score was well calibrated; the predicted and observed mortalities were reasonably similar. Calibration may be more important than discrimination because calibration involves the absolute risk of death for a patient rather than determining whether that patient is more or less likely to die when compared with other patients. Similarly, the expected prediction error in our model was no different from that of APACHE III.

Another important use of classification models is risk stratification for enrollment into clinical trials. For the results of a trial to be applicable to a general population of patients with ALI, it is important to include as many patients as possible. With this approach, the

population of enrolled patients is heterogeneous and includes subsets of patients with highly variable risks of death. Without this approach, increasing numbers of patients will need to be enrolled into future clinical trials to demonstrate a reduction in mortality due to new therapies. In addition, mortality in ALI patients may be attributable to non-pulmonary organ failure rather than just ALI. One strategy for reducing sample size would be to selectively recruit patients with a high risk of mortality attributable to ALI and a high likelihood of response to therapy (40). If the mortality attributable to ALI is low and a therapy that is directed towards reducing mortality from ALI is being tested, a large sample would be required to demonstrate a difference in mortality between the control group and the treatment group (40). Although we validated our classification tree in a test set, it needs to be tested prospectively in another group of patients to see how well it can predict hospital mortality. If accurate, our classification tree would be simple to use to screen ALI patients for enrollment into clinical trials.

We used classification trees to predict hospital mortality because these methods can identify both important predictors and complex interactions between them. The automatic ability to identify interactions is an important advantage over traditional methods. Using standard regression methods, an interaction must first be considered by the investigator and then tested in the model to determine whether it is significant. For example, in our classification tree, the combination of age greater than 63 years and total respiratory rate greater than 21 breaths/minute yielded a predicted hospital mortality of 61%. This is one example of an interaction that may not be considered by an investigator because the combination of age and total respiratory rate may not seem like a clinically important interaction. Interestingly, some of the most important predictors identified by the random forest were not present in the final classification tree. This may be because the interactions between some of the other predictors are more effective at predicting mortality than some of the individual variables. One disadvantage to this approach is that there may be an important loss of statistical power when using continuous variables because the tree selects the optimal cut-point for each continuous predictor. This allows for a classification rule that is easy to use, at the cost of statistical power (43). Classification trees have been used successfully to predict clinical outcomes in cardiology (25–27), but are just starting to be applied to critically-ill patients (44).

Our model performed as well as a model based on the APACHE III score; however, both models had an expected predictor errors of approximately 25%. The classification tree may perform better at predicting ICU mortality, rather than hospital mortality, because the time of death is closer to when the predictor data were collected. Furthermore, our model and APACHE III use only baseline data to predict hospital mortality. Several studies have indicated that longitudinal scores may be more accurate for predicting mortality than using baseline scores (45–48). One study demonstrated that only 6% of patients with a higher acute physiology score on Day 3 compared to Day 1 survived to hospital discharge (45). This was in contrast to 43% survival to hospital discharge for those patients whose acute physiology scores were lower on Day 3 compared to Day 1. In addition, the change in acute physiology score had a positive predicted value of 97% for predicting 100-day mortality, whereas that of APACHE III was 80%. If change in predictors is more accurate for predicting mortality, our classification tree would be much easier to use on a daily basis than calculating an APACHE III score.

Our study has some limitations. First, we used data from randomized controlled trials conducted in tertiary medical centers. Moreover, participants had to meet eligibility criteria to be enrolled, and excluded patients with chronic liver disease (Child-Pugh Class C) and patients with malignancy or any other irreversible condition for which six-month mortality was estimated to be 50% or greater. Since patients with these conditions are not represented



in our cohort, we could not determine if these conditions were more or less important for predicting mortality than the predictors used to create our classification tree. Thus, the generalizability of our prediction model may be limited to patients at tertiary medical centers and to eligibility criteria of ARDS Network trials. Thus, our prediction model requires additional external validation. A second limitation of our study is sample size. When using tree-based methods, it is optimal to have a large number of observations. However, this is one of the largest cohorts of ALI patients and we included more participants than most other studies (8–17) that attempted to predict mortality in patients with ALI. Finally, APACHE has recently undergone a revision (APACHE IV) due to poor calibration of APACHE III; however, APACHE III is still widely used.

In conclusion, we created a simple classification rule that stratified patients according to risk of hospital mortality that performed as well as the APACHE III score. In addition, our classification tree is simple to use and contains risk factors for mortality that may be specific to patients with ALI. Therefore, it may be valuable to use for risk stratification and screening ALI patients for enrollment into clinical trials.

## Acknowledgments

**Funding:** Supported by NHLBI Contracts NO1-HR-46054 through 46064 and NO1-HR 56165 through 56179 with the National Institutes of Health, National Heart, Lung, and Blood Institute. Lisa Brown was supported by a National Institutes of Health Training Grant (T32GM008258-21). Carolyn Calfee was supported by NHLBI grant HL090833 and by a grant of the Flight Attendant Medical Research Institute. Dr. Matthay was supported by NHLBI HL51856. William Checkley was supported by a Clinician Scientist Award of the Johns Hopkins University and a Pathway to Independence Award from the National Institutes of Health, National Heart, Lung, and Blood Institute (K99HL096955).

The authors would like to acknowledge Dr. Hector Corrada-Bravo, Assistant Professor, Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD, and Dr. Jerome Friedman, Professor in the Department of Statistics, Stanford University, Palo Alto, CA, for helpful suggestions in the statistical analyses and software code.

## References

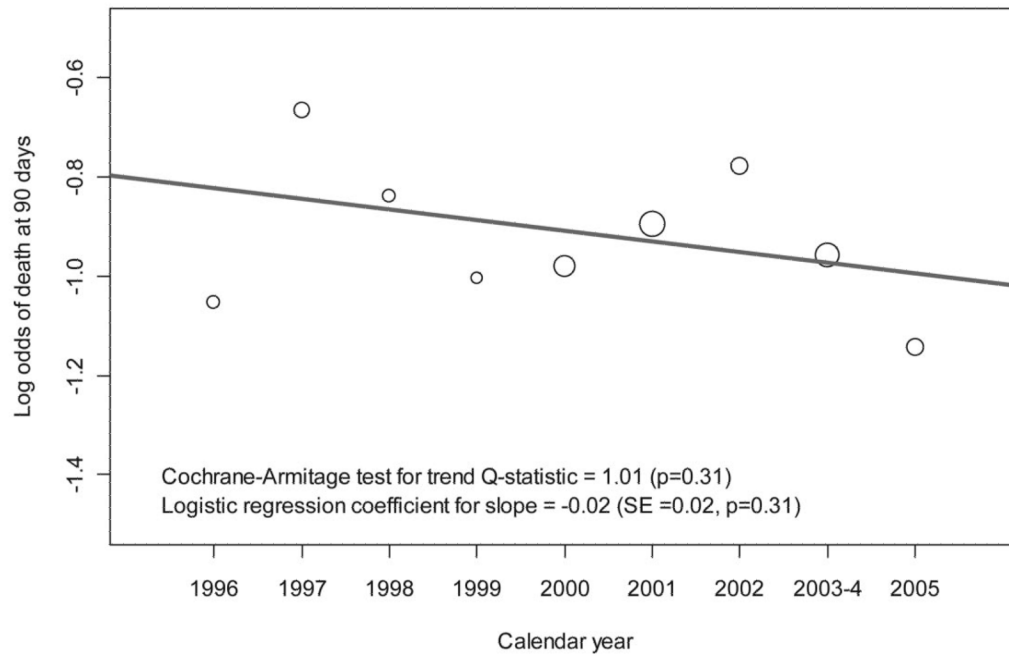
1. Rubenfeld GD, Caldwell E, Peabody E, Weaver J, Martin DP, Neff M, et al. Incidence and outcomes of acute lung injury. *N Engl J Med.* 2005; 353:1685–93. [PubMed: 16236739]
2. Ware LB. Prognostic determinants of acute respiratory distress syndrome in adults: impact on clinical trial design. *Crit Care Med.* 2005; 33:S217–22. [PubMed: 15753731]
3. 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; 342:1301–8. [PubMed: 10793162]
4. Wiedemann HP, Wheeler AP, Bernard GR, Thompson BT, Hayden D, deBoisblanc B, et al. Comparison of two fluid-management strategies in acute lung injury. *N Engl J Med.* 2006; 354:2564–75. [PubMed: 16714767]
5. Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, Lopez-Rodriguez A, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med.* 2001; 344:699–709. [PubMed: 11236773]
6. Kress JP, Pohlman AS, O'Connor MF, Hall JB. Daily interruption of sedative infusions in critically ill patients undergoing mechanical ventilation. *N Engl J Med.* 2000; 342:1471–7. [PubMed: 10816184]
7. Cooke CR, Kahn JM, Caldwell E, Okamoto VN, Heckbert SR, Hudson LD, et al. Predictors of hospital mortality in a population-based cohort of patients with acute lung injury. *Crit Care Med.* 2008; 36:1412–20. [PubMed: 18434894]
8. Cooke CR, Shah CV, Gallop R, Bellamy S, Ancukiewicz M, Eisner MD, et al. A simple clinical predictive index for objective estimates of mortality in acute lung injury. *Crit Care Med.* 2009; 37:2110–1. [PubMed: 19448459]

9. Doyle RL, Szaflarski N, Modin GW, Wiener-Kronish JP, Matthay MA. Identification of patients with acute lung injury. Predictors of mortality. *Am J Respir Crit Care Med.* 1995; 152:1818–24. [PubMed: 8520742]
10. Zilberberg MD, Epstein SK. Acute lung injury in the medical ICU: comorbid conditions, age, etiology, and hospital outcome. *Am J Respir Crit Care Med.* 1998; 157:1159–64. [PubMed: 9563734]
11. Monchi M, Bellenfant F, Cariou A, Joly LM, Thebert D, Laurent I, et al. Early predictive factors of survival in the acute respiratory distress syndrome. A multivariate analysis. *Am J Respir Crit Care Med.* 1998; 158:1076–81. [PubMed: 9769263]
12. Luhr OR, Antonsen K, Karlsson M, Aardal S, Thorsteinsson A, Frostell CG, et al. Incidence and mortality after acute respiratory failure and acute respiratory distress syndrome in Sweden, Denmark, and Iceland. The ARF Study Group. *Am J Respir Crit Care Med.* 1999; 159:1849–61. [PubMed: 10351930]
13. Rocco TR Jr, Reinert SE, Cioffi W, Harrington D, Buczko G, Simms HH. A 9-year, single-institution, retrospective review of death rate and prognostic factors in adult respiratory distress syndrome. *Ann Surg.* 2001; 233:414–22. [PubMed: 11224631]
14. Nuckton TJ, Alonso JA, Kallet RH, Daniel BM, Pittet JF, Eisner MD, et al. Pulmonary dead-space fraction as a risk factor for death in the acute respiratory distress syndrome. *N Engl J Med.* 2002; 346:1281–6. [PubMed: 11973365]
15. Estenssoro E, Dubin A, Laffaire E, Canales H, Saenz G, Moseinco M, et al. Incidence, clinical course, and outcome in 217 patients with acute respiratory distress syndrome. *Crit Care Med.* 2002; 30:2450–6. [PubMed: 12441753]
16. Venet C, Guyomarc'h S, Pingat J, Michard C, Laporte S, Bertrand M, et al. Prognostic factors in acute respiratory distress syndrome: a retrospective multivariate analysis including prone positioning in management strategy. *Intensive Care Med.* 2003; 29:1435–41. [PubMed: 12827238]
17. Brun-Buisson C, Minelli C, Bertolini G, Brazzi L, Pimentel J, Lewandowski K, et al. Epidemiology and outcome of acute lung injury in European intensive care units. Results from the ALIVE study. *Intensive Care Med.* 2004; 30:51–61. [PubMed: 14569423]
18. Raurich JM, Vilar M, Colomar A, Ibanez J, Ayestaran I, Perez-Barcena J, et al. Prognostic value of the pulmonary dead-space fraction during the early and intermediate phases of acute respiratory distress syndrome. *Respir Care.* 2010; 55:282–7. [PubMed: 20196876]
19. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest.* 1991; 100:1619–36. [PubMed: 1959406]
20. Moreno RP, Metnitz PG, Metnitz B, Bauer P, Afonso de Carvalho S, Hoeftl A. SAPS 3 Investigators. Modeling in-hospital patient survival during the first 28 days after intensive care unit admission: a prognostic model for clinical trials in general critically ill patients. *J Crit Care.* 2008; 23:339–48. [PubMed: 18725038]
21. Seeley E, McAuley DF, Eisner M, Miletin M, Matthay MA, Kallet RH. Predictors of mortality in acute lung injury during the era of lung protective ventilation. *Thorax.* 2008; 63:994–8. [PubMed: 18566110]
22. Hager DN, Krishnan JA, Hayden DL, Brower RG. ARDS Clinical Trials Network. Tidal volume reduction in patients with acute lung injury when plateau pressures are not high. *Am J Respir Crit Care Med.* 2005; 172:1241–5. [PubMed: 16081547]
23. Checkley W, Brower R, Korpak A, Thompson BT. Acute Respiratory Distress Syndrome Network Investigators. Effects of a clinical trial on mechanical ventilation practices in patients with acute lung injury. *Am J Respir Crit Care Med.* 2008; 177:1215–22. [PubMed: 18356562]
24. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees.* Belmont, California: Wadsworth; 1984.
25. Goldman L, Cook EF, Johnson PA, Brand DA, Rouan GW, Lee TH. Prediction of the need for intensive care in patients who come to the emergency departments with acute chest pain. *N Engl J Med.* 1996; 334:1498–504. [PubMed: 8618604]



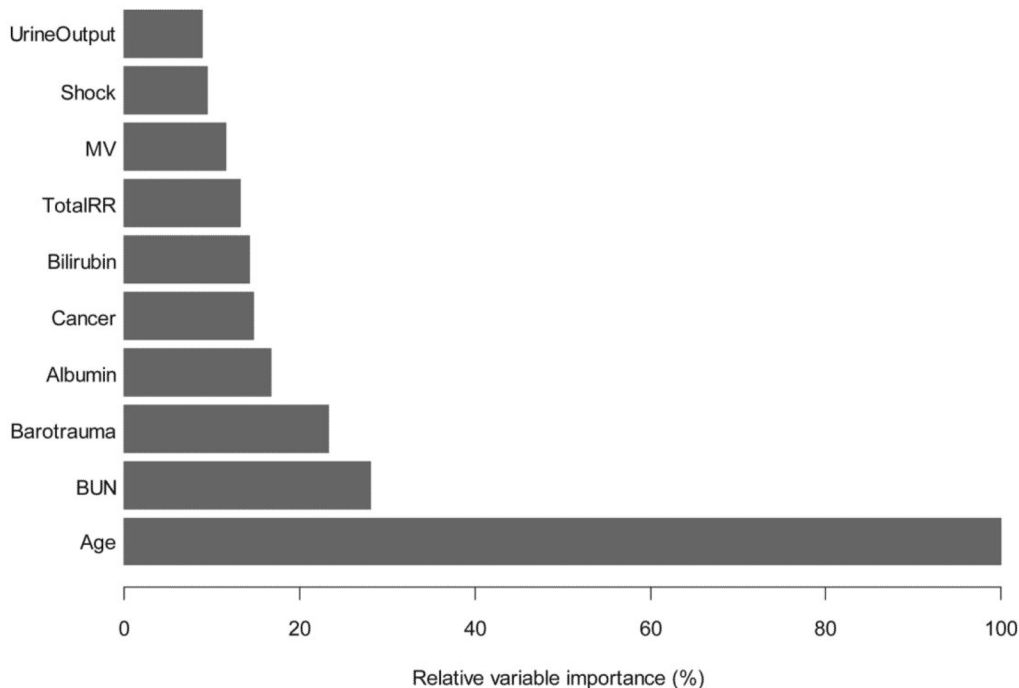
26. Fonarow GC, Adams KF Jr, Abraham WT, Yancy CW, Boscardin WJ. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. *JAMA*. 2005; 293:572–80. [PubMed: 15687312]
27. Pilote L, Miller DP, Califf RM, Rao JS, Weaver WD, Topol EJ. Determinants of the use of coronary angiography and revascularization after thrombolysis for acute myocardial infarction. *N Engl J Med*. 1996; 335:1198–205. [PubMed: 8815943]
28. Brower RG, Lanken PN, MacIntyre N, Matthay MA, Morris A, Ancukiewicz M, Schoenfeld D, Thompson BT. National Heart, Lung, and Blood Institute ARDS Clinical Trials Network. Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome. *N Engl J Med*. 2004; 351:327–36. [PubMed: 15269312]
29. Angus DC, Musthafa AA, Clermont G, Griffin MF, Linde-Zwirble WT, Dremsizov TT, et al. Quality-adjusted survival in the first year after the acute respiratory distress syndrome. *Am J Respir Crit Care Med*. 2001; 163:1389–94. [PubMed: 11371406]
30. Hothorn T, Hornik K, Zeileis A. Unbiased Recursive Partitioning: A Conditional Inference Framework. *J Computational Graphical Stat*. 2006; 15:651–74.
31. Hothorn T, Hornik K, Van de Wiel M, Zeileis A. A Lego System for Conditional Inference. *The American Statistician*. 2006; 60:257–63.
32. Breiman L. Random Forests. *Machine Learning*. 2001; 45:5–32.
33. Zar, J. *Biostatistical Analysis*. Englewood Cliffs: Prentice & Hall; 1984.
34. Armitage, P.; Berry, G. *Methods in Medical Research*. 3. Oxford: Blackwell Science Ltd; 1994.
35. Ferring M, Vincent JL. Is outcome from ARDS related to the severity of respiratory failure? *Eur Respir J*. 1997; 10(6):1297–300. [PubMed: 9192932]
36. Stapleton RD, Wang BM, Hudson LD, Rubenfeld GD, Caldwell ES, Steinberg KP. Causes and timing of death in patients with ARDS. *Chest*. 2005; 128:525–32. [PubMed: 16100134]
37. Matthay MA, Brower R, Thompson BT, Schoenfeld D, Eisner M, Carson S, et al. Randomized, placebo controlled trial of an aerosolized beta-2 adrenergic agonist (albuterol) for the treatment of acute lung injury. *Amer J Respir Crit Care Med*. 2009; 179:A2166.
38. Liu KD, Glidden DV, Eisner MD, Parsons PE, Ware LB, Wheeler A, et al. Predictive and pathogenetic value of plasma biomarkers for acute kidney injury in patients with acute lung injury. *Crit Care Med*. 2007; 35:2755–61. [PubMed: 18074478]
39. Gajic O, Afessa B, Thompson BT, Frutos-Vivar F, Malinchoc M, Rubenfeld GD, Esteban A, Anzueto A, Hubmayr RD. Second International Study of Mechanical Ventilation and ARDS-net Investigators. Prediction of death and prolonged mechanical ventilation in acute lung injury. *Crit Care*. 2007; 11(3):R53. [PubMed: 17493273]
40. Wood KA, Huang D, Angus DC. Improving clinical trial design in acute lung injury. *Crit Care Med*. 2003; 31:S305–11. [PubMed: 12682457]
41. Calfee CS, Eisner MD, Ware LB, Thompson BT, Parsons PE, Wheeler AP, et al. Trauma-associated lung injury differs clinically and biologically from acute lung injury due to other clinical disorders. *Crit Care Med*. 2007; 35:2243–50. [PubMed: 17944012]
42. Eisner MD, Thompson T, Hudson LD, Luce JM, Hayden D, Schoenfeld D, et al. Efficacy of low tidal volume ventilation in patients with different clinical risk factors for acute lung injury and the acute respiratory distress syndrome. *Am J Respir Crit Care Med*. 2001; 164:231–6. [PubMed: 11463593]
43. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 2006; 25:127–41. [PubMed: 16217841]
44. Trujillano J, Badia M, Servia L, March J, Rodriguez-Pozo A. Stratification of the severity of critically ill patients with classification trees. *BMC Med Res Methodol*. 2009; 9:83. [PubMed: 20003229]
45. Afessa B, Keegan MT, Mohammad Z, Finkielman JD, Peters SG. Identifying potentially ineffective care in the sickest critically ill patients on the third ICU day. *Chest*. 2004; 126:1905–9. [PubMed: 15596691]
46. Gartman EJ, Casserly BP, Martin D, Ward NS. Using serial severity scores to predict death in ICU patients: a validation study and review of the literature. *Curr Opin Crit Care*. 2009; 15:578–82. [PubMed: 19855272]

47. Esserman L, Belkora J, Lenert L. Potentially ineffective care. A new outcome to assess the limits of critical care. *JAMA*. 1995; 274:1544–51. [PubMed: 7474223]
48. Wagner DP, Knaus WA, Harrell FE, Zimmerman JE, Watts C. Daily prognostic estimates for critically ill adults in intensive care units: results from a prospective, multicenter, inception cohort analysis. *Crit Care Med*. 1994; 22:1359–72. [PubMed: 8062557]



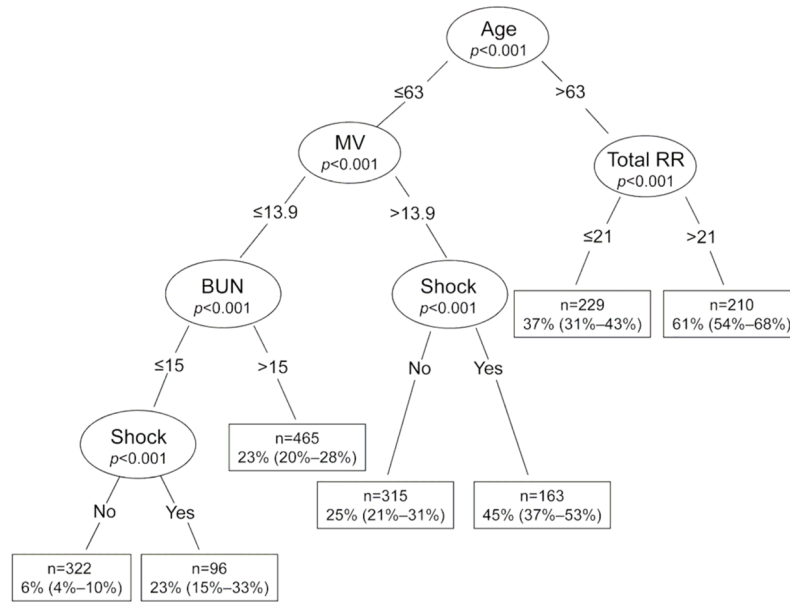
**Figure 1. Lack in trend in hospital mortality at 90 days in participants of ARDS Network trials who received lung protective ventilation; 1996–2005**

The scale of the x-axis is calendar year and that of the y-axis is log odds of hospital mortality. Each circle represents the log odds of mortality at each calendar year. The size of the circle is proportional to sample size. The red line is the logistic regression fit of log odds mortality on calendar year.



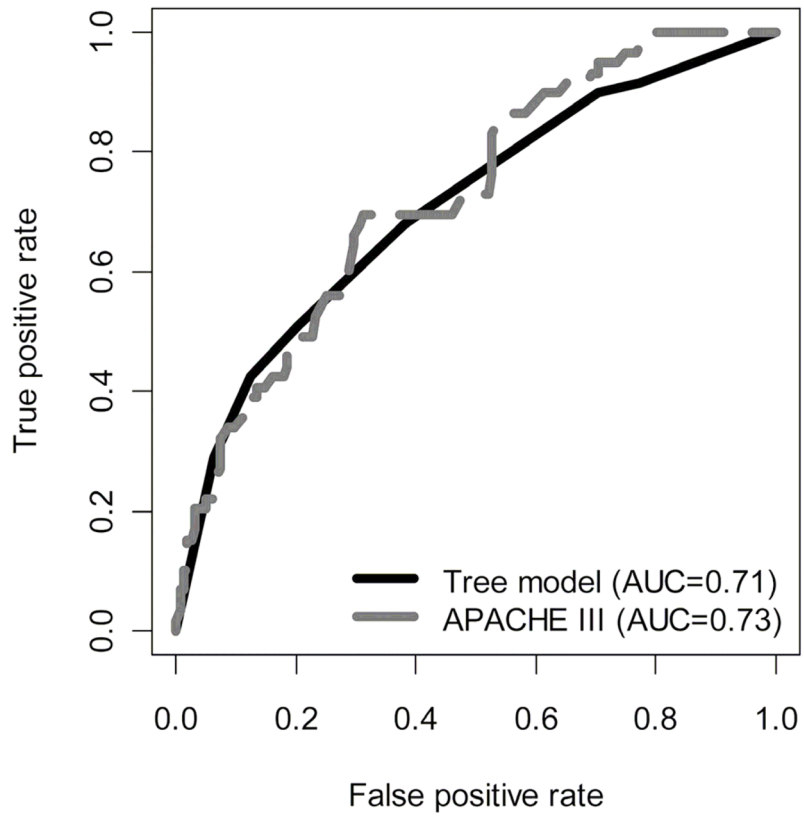
**Figure 2. Ten predictors with the highest relative variable importance as determined by a random forest using baseline data from ARDS Network clinical trials; 1996–2005**

We constructed a random forest consisting of 500 conditional classification trees. By creating numerous classification trees, the random forest indicates which predictors are used most frequently to partition the data with the goal of optimizing discrimination between participants who died in the hospital at 90 days and those who were alive. Therefore, this approach provides a metric that can be used to compare the relative importance of the predictors in terms of their ability to discriminate between survivors and non-survivors. We display the ten predictors with the highest variable importance. Age has a variable importance of 100% and all other variables are compared to this reference.



**Figure 3. Classification tree of hospital mortality at 90 days in 1800 participants of ARDS Network trials; 1996–2005**

The p-values represent the significance level of the partition as evaluated by the conditional inference tree. The boxes outlined in dark are the terminal nodes. We present the sample size, estimated probability of hospital mortality at 90 days and corresponding 95% CI in parentheses.



**Figure 4. Receiver operating characteristic curves of tree model versus APACHE III using data from ARDS Network trials; 1996–2005**

Determined by applying our classification tree and the APACHE III score to the 222 participants in the test set.



**Table 1**

Baseline Characteristics for 2022 Patients with Acute Lung Injury; 1996–2005.

Characteristic	Median (IQR) or n (%)	Missing Data n (%)
90-Day Mortality	575 (28.4)	0
<b>Demographics</b>		
Age (years)	49 (38–63)	0
Male (%)	1,121 (55.4)	0
Hispanic (%)	56 (2.8)	0
Black (%)	369 (18.3)	0
<b>Respiratory Predictors</b>		
PaO <sub>2</sub> (mmHg)	80 (66–102)	1 (0.05)
FiO <sub>2</sub> (%)	0.8 (0.55–1.0)	1 (0.05)
PaO <sub>2</sub> /FiO <sub>2</sub> (mmHg)	119 (83–170)	1 (0.05)
SpO <sub>2</sub> (%)	95 (93–98)	39 (1.9)
PEEP (cmH <sub>2</sub> O)	10 (5–12)	7 (0.3)
Oxygenation Index	13.0 (7.8–20.6)	277 (13.7)
PaCO <sub>2</sub> (mmHg)	38 (33–44)	98 (4.8)
Tidal Volume (ml/kg)	7.8 (6.4–9.4)	310 (15.3)
Total Respiratory Rate (breaths/min)	22 (18–29)	6 (0.3)
Minute Ventilation (L/min)	12 (9.6–14.9)	36 (1.8)
Arterial pH	7.39 (7.32–7.44)	98 (4.8)
Peak Pressure (cmH <sub>2</sub> O)	33 (27–39)	178 (8.8)
Mean Airway Pressure (cmH <sub>2</sub> O)	15 (12–19)	276 (13.6)
Plateau Pressure (cmH <sub>2</sub> O)	27 (22–31)	527 (26.1)
<b>Lung Injury Category</b>		
Trauma	178 (8.8)	0
Sepsis	478 (23.6)	0
Multiple Transfusion	47 (2.3)	0
Aspiration	305 (15.1)	0
Pneumonia	837 (41.4)	0
Other	176 (8.7)	0
Chest Radiograph Quadrants		11 (0.5)
Two	123 (6.1)	
Three	330 (16.3)	
Four	1,558 (77.1)	
Barotrauma	121 (6.0)	949 (46.9)
<b>Cardiovascular Predictors</b>		
Mean Arterial Pressure (low) (mmHg)	60 (54–68)	22 (1.1)
Mean Arterial Pressure (high) (mmHg)	97 (88–110)	22 (1.1)
Heart Rate (low) (beats/min)	89 (76–102)	3 (0.1)
Heart Rate (high) (beats/min)	125 (110–140)	3 (0.1)
Central Venous Pressure (mmHg)	12 (8–15)	932 (46.1)

Characteristic	Median (IQR) or n (%)	Missing Data n (%)
Shock	626 (33.2)	136 (6.7)
<b>Other Predictors</b>		
Temperature (Low) (C)	36.8 (36.1–37.4)	3 (0.1)
Temperature (High) (C)	38.5 (37.8–39.1)	3 (0.1)
24 Hour Urine Output (ml)	1,803 (1,125–2,850)	77 (3.8)
<b>Chronic Health Information</b>		
Dialysis	28 (1.4)	41 (2.0)
Diabetes	324 (16.0)	41 (2.0)
AIDS	131 (6.5)	41 (2.0)
Immunosuppression	215 (10.6)	41 (2.0)
Cancer	93 (4.6)	41 (2.0)
Liver Failure	73 (3.6)	41 (2.0)
<b>Laboratory Values</b>		
WBC (mm <sup>3</sup> )	11.8 (7.7–16.7)	14 (0.7)
Hematocrit (%)	31 (28–35)	6 (0.3)
Platelets (low) (×1000/mm <sup>3</sup> )	161 (91–242)	18 (0.9)
Sodium (mEq/L)	138 (135–142)	4 (0.2)
Potassium (low) (mEq/L)	3.7 (3.4–4.1)	6 (0.3)
Potassium (high) (mEq/L)	4.1 (3.8–4.6)	29 (1.4)
Bicarbonate (low) (mEq/L)	22 (18–25)	20 (1.0)
BUN (high) (mg/dL)	19 (13–33)	13 (0.6)
Creatinine (mg/dL)	1.0 (0.8–1.6)	5 (0.2)
Glucose (low) (mg/dL)	116 (94–146)	34 (1.7)
Glucose (high) (mg/dL)	142 (116–193)	57 (2.8)
Bilirubin (high) (mg/dL)	0.8 (0.5–1.7)	345 (17.1)
Albumin (g/dL)	2.2 (1.8–2.6)	298 (14.7)

Selected baseline characteristics and hospital mortality at 90 days in six clinical trials conducted by the National Institutes of Health ARDS Network investigators; 1996–2005.

**Table 2**

	Tidal volume, Ketoconazole and Lisofylline trials	PEEP trial	Fluid management and catheter treatment trials
Sample size	473	549	1000
Mean age in years (SD)	51.1 (17.2)	51.2 (16.9)	49.7 (16.0)
Males (%)	60.2%	55.0%	53.4%
Mean APACHE III score (SD)	82.3 (29.0)	94.0 (31.8)	94.1 (30.9)
Mean plateau pressure cm H <sub>2</sub> O (SD)	29.7 (7.7)	26.8 (7.1)	26.1 (7.0)
Mean PaO <sub>2</sub> /FIO <sub>2</sub> mmHg (SD)	131.9 (60.5)	127.6 (57.5)	131.6 (62.6)
Mean minute ventilation L/min (SD)	13.3 (4.2)	12.0 (3.8)	12.3 (4.0)
Mean total respiratory rate breaths/min (SD)	19.8 (6.7)	23.0 (7.7)	25.5 (7.9)
Mean BUN mg/dL (SD)	27.7 (22.8)	29.4 (24.7)	24.9 (19.2)
Shock (%)	42.9	39.3	46.2
Hospital mortality at 90 days (%)	30.2	26.9	28.4

**Table 3**  
Expected Prediction Errors (EPE) for the Ten Classification Trees and APACHE III using Cross Validation.

Training Fold	Participants per fold	Observed death rate in fold	Predicted death rate for tree model	EPEs for tree model	Predicted death rate for APACHE III	EPEs for APACHE III
<b>1</b>	180	47/180 (26%)	19/180 (11%)	24.4	21/180 (12%)	21.1
<b>2</b>	180	53/180 (29%)	10/180 (6%)	26.1	17/180 (9%)	25.6
<b>3</b>	180	51/180 (28%)	22/180 (7%)	28.3	21/180 (12%)	26.7
<b>4</b>	180	50/180 (28%)	33/180 (18%)	33.9	30/180 (17%)	27.8
<b>5</b>	180	49/180 (27%)	15/180 (8%)	26.7	23/180 (13%)	27.8
<b>6</b>	180	45/180 (25%)	13/180 (9%)	25.6	25/180 (14%)	23.3
<b>7</b>	180	45/180 (25%)	23/180 (14%)	30.0	21/180 (12%)	23.3
<b>8</b>	180	60/180 (33%)	31/180 (14%)	32.8	28/180 (16%)	33.3
<b>9</b>	180	49/180 (27%)	8/180 (4%)	22.8	21/180 (12%)	24.4
<b>10</b>	180	59/180 (33%)	16/180 (9%)	28.3	20/180 (11%)	23.9
<b>Mean EPE</b>				27.8(95% CI25.3 to 30.4)		25.7 (95% CI23.3 to 28.1)