# A simple clinical predictive index for objective estimates of mortality in acute lung injury*

Colin R. Cooke, MD, MSc; Chirag V. Shah, MD, MS; Robert Gallop, PhD; Scarlett Bellamy, ScD; Marek Ancukiewicz, PhD; Mark D. Eisner, MD, MPH; Paul N. Lanken, MD; A. Russell Localio, PhD; Jason D. Christie, MD, MS; for the National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome Network

***Objective:*** **We sought to develop a simple point score that would accurately capture the risk of hospital death for patients with acute lung injury (ALI).**

***Design:*** **This is a secondary analysis of data from two randomized trials. Baseline clinical variables collected within 24 hours of enrollment were modeled as predictors of hospital mortality using logistic regression and bootstrap resampling to arrive at a parsimonious model. We constructed a point score based on regression coefficients.**

***Setting:*** **Medical centers participating in the Acute Respiratory Distress Syndrome Clinical Trials Network (ARDSnet).**

***Patients:*** **Model development: 414 patients with nontraumatic ALI participating in the low tidal volume arm of the ARDSnet Acute Respiratory Management in ARDS study. Model validation: 459 patients participating in the ARDSnet Assessment of Low tidal Volume and elevated End-expiratory volume to Obviate Lung Injury study. Model Validation: 459 patients participating in the ARDSnet Assessment of Low tidal Volume and elevated End-expiratory volume to Obviate Lung Injury trial.**

***Interventions:*** **None.**

***Measurements and Main Results:*** **Variables comprising the prognostic model were hematocrit <26% (1 point), bilirubin ≥2 mg/dL (1 point), fluid balance >2.5 L positive (1 point), and age (1 point for age 40–64 years, 2 points for age ≥65 years). Predicted mortality (95% confidence interval) for 0, 1, 2, 3, and 4+ point totals was 8% (5% to 14%), 17% (12% to 23%), 31% (26% to 37%), 51% (43% to 58%), and 70% (58% to 80%), respectively. There was an excellent agreement between predicted and observed mortality in the validation cohort. Observed mortality for 0, 1, 2, 3, and 4+ point totals in the validation cohort was 12%, 16%, 28%, 47%, and 67%, respectively. Compared with the Acute Physiology Assessment and Chronic Health Evaluation III score, areas under the receiver operating characteristic curve for the point score were greater in the development cohort (0.72 vs. 0.67, *p* = 0.09) and lower in the validation cohort (0.68 vs. 0.75, *p* = 0.03).**

***Conclusions:*** **Mortality in patients with ALI can be predicted using an index of four readily available clinical variables with good calibration. This index may help inform prognostic discussions, but validation in nonclinical trial populations is necessary before widespread use. (Crit Care Med 2009; 37:1913–1920)**

**Key Words: acute respiratory distress syndrome; acute lung injury; respiratory distress syndrome; adult; human acute respiratory distress syndrome; statistical model; logistic models; mortality determinants; mortality; in-hospital; Acute Physiology and Chronic Health Evaluation III; Bayesian prediction; prognosis**

Acute lung injury (ALI) is a devastating cause of respiratory failure associated with significant morbidity and mortality (1, 2). Despite the wealth of existing knowledge about risk factors for death in this syndrome, providers remain unable to determine which patients with ALI will ultimately die during their hospital stay. The majority of patients with ALI who die do so in the context of a decision to forgo life-sustaining treatment driven in large part by patient preferences (3–5).

Prognostication in the intensive care unit (ICU) is an important part of communication with surrogates, and often plays a role in the decision to forgo life-sustaining treatment (6, 7). Incapacitated patients rely on surrogates, such as their family members, to represent their wishes during ICU care, and surrogates often rely on clinician estimates of the likelihood of survival and functional recovery from acute illness when deciding whether to forgo life-sustaining treatment for their loved ones (7). Documented cognitive and noncognitive biases held by physicians may overly influence their prognostic estimates for a given patient and have the potential to misrepresent true risk of death (8–11). Objective prognostic models, such as the Acute Physiology Assessment and Chronic Health Evaluation (APACHE) III score (12) and Simplified Acute Physiology Score III (13), can provide estimated probabilities of death for an individual patient in the ICU. How-

***See also p. 2110.**

From the Division of Pulmonary and Critical Care Medicine (CRC), University of Washington, Seattle, WA; Division of Pulmonary, Allergy and Critical Care (CVS, PNL, JDC), University of Pennsylvania, Philadelphia, PA; Department of Biostatistics and Epidemiology (CVS, SB, ARL, JDC), Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, PA; Department of Mathematics and Applied Statistics (RG), West Chester University, West Chester, PA; Massachusetts General Hospital (MA), Boston, MA; Division of Occupational and Environmental Medicine and Division of Pulmonary and Critical Care Medicine (MDE), University of California, San Francisco, CA.

Supported, in part, by F32 HL090220, N01 HR46055, N01 HR46058 from the National Institutes of Health.

This study was conducted at the University of Pennsylvania and the University of Washington.

The authors have not disclosed any potential conflicts of interest.

For information regarding this article, E-mail: crcooke@u.washington.edu

**DOI: 10.1097/CCM.0b013e3181a009b4**

ever, experts recommend against use of these models for predicting outcomes for individual patients, in part, because of their inability to convey uncertainty in estimated probabilities of death for an individual patient and the complexity involved in their calculation (14).

The goal of this study was to develop a simple, disease-specific multivariable predictive scorecard for mortality to be used at the bedside in patients with early ALI. Given the importance of well-calibrated models for individual prognostication (15), we sought to maximize the concordance between predicted and actual probabilities of hospital death across point strata for our model, and, thus, to arrive at a system that might classify patients into groups for planning patient care.

## MATERIALS AND METHODS

*Study Population.* The model derivation population arose from the 861 patients participating in the Acute Respiratory Distress Syndrome Clinical Trials Network (ARDSNet) low tidal volume study (Acute Respiratory Management in ARDS) (16). Briefly, intubated, mechanically ventilated patients meeting American European Consensus Conference (17) definition for ALI were randomized within 36 hours of meeting the last qualifying American European Consensus Conference criterion to receive tidal volumes of 6 or 12 mL/kg predicted body weight. Demographics, comorbidities, ALI precipitating cause, physiology, and radiographic and ventilator data were recorded within the 24 hours before change in ventilator settings for all enrolled patients. Vital status for each patient was determined at hospital discharge. We limited our development cohort to all patients randomized into the 6 mL/kg arm of the parent study to eliminate tidal volume as a predictive variable in the analysis because current best practice involves low tidal volume ventilation for this population (n = 473). Patients with trauma as the primary risk factor for ALI were excluded because of the low mortality rate in this subgroup (18).

*Model Development.* Our general strategy to develop a predictive model for death consisted of three steps. First, we identified variables previously reported as associated with mortality or severity of illness in ALI. Baseline values were selected to minimize missing data and to allow for mortality prediction at the beginning of ALI. Next, we constructed a parsimonious multivariable model based on these predictors. Finally, we validated the final predictive model in an independent sample of patients.

When deciding the covariates to be retained as candidate predictors for the multivariable model, we considered the clinical relevance and generalizability of each covariate; the amount of missing data (retaining the measure with the least missing data); and finally, the amount of spread in the covariate's scale (retaining the measure with the most variability) in that order. We assessed the collinearity among the predictors using the Pearson correlation coefficient, $\chi^2$ tests, and analysis of variance/Student's $t$ tests. When highly correlated covariates quantified the same clinical information (e.g., A-a difference and $\mathrm{Pa{\small O}_2}$), we selected the covariate that was more clinically relevant and had less missing data and more variability.

*Multivariable Modeling.* The resulting baseline clinically relevant covariates with minimal collinearity were entered into a multivariable logistic regression model. These variables included demographics (age, sex, race/ethnicity); weight; respiratory physiology ($\mathrm{Pa{\small O}_2/F{\small IO}_2}$, $\mathrm{Pa{\small CO}_2}$, positive end-expiratory pressure, number of opacified quadrants on frontal chest radiograph [19], volume/pressure-targeted ventilation, assist/control ventilation); primary ALI risk factor as coded by the clinical coordinator and physician investigator within 36 hours of ALI onset (pneumonia, sepsis, aspiration, other/none); timing of ALI onset (hospital days before ARDSnet screen, days with ALI before randomization); and physiologic and laboratory derangement (number of nonpulmonary organ failures, vasopressor use, net 24-hour fluid balance before enrollment, 24-hour urine output before enrollment, peak bilirubin, peak creatinine, lowest systolic blood pressure, lowest hematocrit). All peak and nadir values were identified during the 24-hour period before enrollment. We included continuous variables in categorical form to simplify point calculation from the final model. We determined cut points for continuous variables by assessing each variable's functional form using generalized additive models (20). We evaluated two-way multiplicative interactions for each covariate, which were excluded from the final model if they were not statistically significant.

Variable selection in the multivariable regression framework used a bootstrap algorithm (21). We generated 1000 bootstrap samples from the original dataset. Each bootstrap sample was the same size of the original derivation sample; however, patients in each bootstrap sample were randomly drawn from the original data with replacement (21). Within each bootstrap sample, we performed stepwise logistic regression with thresholds of $p = 0.10$ for selection and $p = 0.20$ for variable elimination. Predictors present in at least 600 runs (e.g., 60% of the 1000 generated bootstrap samples) were entered in a final logistic regression model using the original data (22, 23). This method determines the empirical distribution of a variable's likelihood of being included in the model, thereby quantifying the strength of evidence that a given variable is indeed a true independent predictor of death and compares favorably to more traditional cross-validation or isolated automated model development methods (23).

*Score Generation.* Point scores were assigned to each covariate by rounding the regression coefficients in the final model to integers (24). We then calculated a point score for each patient in the cohort and plotted the resulting receiver operating characteristic curve. The receiver operating characteristic curve graphically describes the overall performance of our point score (25). Discrimination of the model was summarized with area under the curve (AUC) of the receiver operating characteristic curve (25). In addition, we derived positive likelihood ratio (LR+) estimates for each level of the point score to be able to estimate how much a prior probability of death would be influenced by an observed point score. The LR+ summarizes how many more times likely patients who die are to have that particular point total than patients who survive (26, 27). Predicted probabilities of death and their respective confidence intervals for each point strata were generated from a logistic regression with mortality as the outcome and the point totals per patient as the sole predictor. Posttest probabilities of death were generated using hypothetical, provider-determined pretest probabilities of death and the LR+ for each point category as previously described (27). We calculated confidence intervals for posttest probabilities of death by incorporating the uncertainty in the LR. Pretest probabilities were assumed to have no uncertainty. We assessed calibration using the Hosmer-Lemeshow statistic with $p < 0.10$ indicating that fit was inadequate (28). Given the low power of this test in small samples, we also compared the actual and predicted mortality within each point stratum for the development and validation cohorts.

*Model Validation.* We assessed internal validity of our model by comparing the AUC of our point score to that of the predicted mortality estimated from the APACHE III score (12) using the method outlined by DeLong et al (29). APACHE probabilities of death were generated by fitting the APACHE III score in a logistic model where hospital death was the outcome. We assessed external validity by applying our model to an independent database, which consisted of the same target study population used in constructing the prediction model (participants in the ARDSnet clinical trial, Assessment of Low tidal Volume and elevated End-expiratory volume to Obviate Lung Injury [ALVEOLI]) (30). Briefly, ALVEOLI randomized 549 intubated, mechanically ventilated patients meeting the American European Consensus Conference definition for ALI or ARDS within 36 hours to receive higher or lower positive end-expiratory pressure. All patients received tidal volumes of 6 mL/kg predicted body weight. Baseline variables collected in ALVEOLI were similar to those captured in Acute Respiratory Management in ARDS. Patients were followed up until discharge. We limited our analysis of ALVEOLI to patients without trauma as the primary ALI risk factor (n = 505).

Table 1. Baseline characteristics of patients eligible for model development by vital status

| Variable[a] | Vital Status at Hospital Discharge[b] |  | $p$ |
|---|---|---|---|
|  | Alive | Dead |  |
| Cases | 275 | 139 |  |
| Age, years, median (IQR) | 48 (37–61) | 60 (45–72) | <0.001 |
| Male (%) | 59 | 61 | 0.61 |
| Race (%) |  |  | 0.27 |
| White | 75 | 70 |  |
| Black | 17 | 19 |  |
| Hispanic | 4 | 4 |  |
| Other/unknown | 4 | 7 |  |
| Timing of ALI |  |  |  |
| Hospital days before ALI, median (IQR) | 2 (1–5) | 4 (1–8) | 0.001 |
| ALI days before randomization, median (IQR) | 1 (0–1) | 1 (0–1) | 0.97 |
| Primary ALI risk factor (%) |  |  | 0.59 |
| Pneumonia | 36 | 33 |  |
| Sepsis | 28 | 34 |  |
| Aspiration | 17 | 19 |  |
| Multiple transfusion | 3 | 2 |  |
| None/other | 16 | 12 |  |
| Severity of illness |  |  |  |
| APACHE III score | 79 (27) | 96 (30) | <0.001 |
| Number of organ failures, median (IQR) | 1 (0–1) | 1 (0–2) | 0.01 |
| Net volume during preceding 24 hrs (mL) | 2276 (3616) | 3361 (4546) | 0.01 |
| Vasopressor use (%) | 38 | 49 | 0.03 |
| Respiratory physiology |  |  |  |
| Minute ventilation (L/min) | 13 (4) | 14 (4) | 0.1 |
| Plateau pressure (mm Hg) | 29 (7) | 31 (8) | 0.01 |
| PEEP, mm Hg, median (IQR) | 8 (5–10) | 10 (5–10) | 0.11 |
| Pao$_2$/Fio$_2$ ratio | 152 (71) | 135 (61) | 0.02 |
| pH | 7.37 (0.1) | 7.36 (0.1) | 0.67 |
| Paco$_2$ (mm Hg) | 36 (8) | 36 (8) | 0.98 |
| Additional physiology and laboratories[c] |  |  |  |
| Systolic blood pressure (mm Hg) | 89 (19) | 83 (19) | 0.003 |
| Twenty-four–hour urine output (mL) | 2400 (1539) | 2068 (1612) | 0.05 |
| Glucose (mg/dL) | 177 (100) | 184 (90) | 0.48 |
| Creatinine (mg/dL) | 1.6 (1.5) | 1.8 (1.4) | 0.22 |
| Hematocrit (%) | 30 (6) | 29 (5) | 0.04 |
| Bilirubin (mg/dL) | 1.6 (1.9) | 2.4 (3.3) | 0.02 |

IQR, interquartile range; ALI, acute lung injury; APACHE, Acute Physiology Assessment and Chronic Health Evaluation; Pao$_2$, partial pressure of arterial oxygen; Paco$_2$, partial pressure of arterial carbon dioxide.

[a]Data were missing for plateau pressure in 87 (21%) patients, bilirubin, 38 (9%); Paco$_2$, 30 (7%); Pao$_2$/Fio$_2$, 30 (7%); fluid balance, 27 (7%); glucose, 26 (6%); creatinine, 23 (6%); urine output, 22 (5%); minute ventilation, 4 (<1%); pH, 3 (<1%); hematocrit, 3 (<1%); APACHE III in two; vasopressor in two; primary ALI risk factor in one; systolic blood pressure in one; [b]numbers reflect mean (SD) unless otherwise noted. Percentages may not add to 100 due to rounding; [c]numbers represent worst values over the 24-hr period surrounding enrollment day.

Table 2. Model-based points for each cut point in predictive variables in the final multivariable model

| Variable | 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 | — |

As a sensitivity analysis, we determined the influence of missing data on our model by performing multiple imputation (SAS PROC MI) for each incomplete covariate as described by Rubin and Schenker (31). The imputed model and mortality estimates derived from the imputed model were identical to those from complete case analysis. We also used the same variables and cut points to determine model performance for predicting 28-day mortality.

The Institutional Review Board for each center participating in ARDSnet approved the parent studies. All statistical analyses were conducted with SAS 9.1 (Statistical Analysis Systems, Cary, NC) and Stata 9.2 (StataCorp, College Station, TX). All tests of significance used a two-sided $\alpha = 0.05$.

## RESULTS

Of the 902 patients participating in the ARDSnet low tidal volume study, 429 were randomized to the 12 mL/kg tidal volume arm and excluded. Of the remaining 473 patients, 59 (12%) were excluded because of trauma as the primary risk factor for ALI, leaving 414 patients (88% of patients in the 6 mL/kg arm) available for analysis. Demographics, ALI risk factor, severity of illness, and laboratory and physiology data for the cohort are shown in Table 1. Of the 414 patients in the development cohort, 139 (33%) died at hospital discharge, similar to the 31% mortality reported in the 6 mL/kg arm of the parent study (16). In general, patients who were dead at hospital discharge were older and had a greater severity of physiologic and laboratory derangement.

*Multivariable Modeling.* During multivariable modeling, 64 additional patients were excluded because of missing data for bilirubin (n = 38, 9%), fluid balance (n = 24, 6%), and hematocrit (n = 2). Variables retained in the final regression of the covariates present in >60% of the bootstrap iterations included age, hematocrit, 24-hour fluid balance, and bilirubin. The model derived from imputed data was identical to that derived by complete case analysis. For simplicity, we report only the results of the complete case analysis. Point values generated from the regression coefficients for each of these covariates are shown in Table 2. The resulting point total for each patient was incorporated in a regression with hospital mortality as the outcome. We refer to this model as the custom model. Predicted mortality by point total for the development cohort and observed mortality in the development and validation cohorts are presented in Table 3. The mean predicted mortality for each point strata was very close to the observed mortality in both the development and the validation cohorts. In all strata, observed mortality in the validation cohort fell within the confidence bounds of the predicted mortality.

LR+ and 95% confidence intervals for each point total in the combined cohorts are also shown in Table 3. By using the LR+s from Table 3, we calculated the hypothetical posttest probability of death as a function of point total from our model over a range of pretest probabilities of death (Table 4).

The comparison between predicted mortality estimated from the APACHE III score and the mortality rate predicted by the custom model is illustrated in Figure 1. Overall, there was considerable spread in the predicted mortality estimated from

Table 3. Predicted and observed hospital mortality, and positive likelihood ratios in the derivation set (ARMA) and the validation set (ALVELOLI)

| | Predicted Mortality | | Observed Mortality | | Diagnostic Likelihood Ratio |
|---|---|---|---|---|---|
| Total Points | % | 95% CI | ARMA | ALVEOLI | + (95% CI)[a] |
| 0 | 8.0 | (4.6–13.7) | 8.1 | 12.3 | 0.30 (0.16–0.54) |
| 1 | 16.5 | (11.9–22.5) | 16.0 | 16.3 | 0.47 (0.34–0.64) |
| 2 | 31.0 | (26.0–36.6) | 30.1 | 27.8 | 0.98 (0.79–1.23) |
| 3 | 50.6 | (42.7–58.4) | 54.4 | 46.5 | 2.50 (1.94–3.22) |
| 4+ | 70.0 | (58.1–79.5) | 60.0 | 66.7 | 4.13 (2.12–8.07) |

CI, confidence interval; ARMA, Acute Respiratory Management in Acute Respiratory Distress Syndrome; ALVEOLI, Assessment of Low Tidal Volume and Elevated End-Expiratory Pressure to Obviate Lung Injury.

[a] Pooled likelihood ratios for ARMA and ALVEOLI. Positive likelihood ratio (LR+) can be multiplied by the pretest odds of outcome to get the posttest odds of outcome. Pretest odds can be calculated as $p/1-p$, where $p$ = pretest probability of disease. Posttest probability is calculated as (posttest odds/1 + posttest odds).

Table 4. Estimated posttest percent hospital mortality (95% confidence interval) for a range of pretest rates of death

| Pretest Estimated Mortality (%) | Calculated Point Total for Patient | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4+ |
| 5 | 2 (1–3) | 2 (2–3) | 5 (4–6) | 12 (9–14) | 18 (10–30) |
| 10 | 3 (2–6) | 5 (4–7) | 10 (8–12) | 22 (18–26) | 31 (19–47) |
| 25 | 9 (5–15) | 14 (10–18) | 25 (21–29) | 45 (39–52) | 58 (41–73) |
| 50 | 23 (14–35) | 32 (26–39) | 50 (44–55) | 71 (66–76) | 81 (68–89) |
| 75 | 47 (33–62) | 59 (51–66) | 75 (70–79) | 88 (85–91) | 93 (86–96) |
| 90 | 73 (60–83) | 81 (76–85) | 90 (88–92) | 96 (95–97) | 97 (95–99) |
| 95 | 85 (76–91) | 90 (87–92) | 95 (94–96) | 98 (97–98) | 99 (98–99) |

[a] Posttest percents calculated using likelihood ratios reported in Table 3 and Bayes' Theorem. Confidence intervals incorporate the uncertainty in the estimated likelihood ratios; [b] numbers represent a hypothetical, bedside assessment of the chance of dying before calculation of point score.

the APACHE III score within each point total. Hosmer-Lemeshow goodness of fit test for the custom model in the development and validation cohort showed no evidence of inadequate fit ($\chi^2_{df=3} = 1.5$, $p = 0.67$ and $\chi^2_{df=3} = 1.0$, $p = 0.79$, respectively).

Receiver operating characteristic curves for the custom model in the development and validation cohorts are compared with APACHE III in Figure 2. The custom model outperformed APACHE III in the development cohort and performed worse than APACHE III in the validation cohort. The AUC for the custom model in the derivation set was 0.72 compared with 0.67 for APACHE III ($p = 0.09$). When applied to the validation cohort, the AUC for the custom model was 0.68, whereas the AUC for APACHE III was 0.75 ($p = 0.03$).

*Twenty-Eight–Day Mortality.* At 28 days, 90 (26%) patients in the development cohort died. Predicted 28-day mortality, observed 28-day mortality, and LR+ for the development and validation cohorts are present in Table 5. In general, 28-day mortality was lower than hospital mortality for each point total; however, there was good agreement between predicted and observed mortality for each point total in the validation cohort. LR+s for each point total were similar to those reported for hospital mortality. Discrimination of the custom model in the development cohort was similar to discrimination in the validation cohort (AUC 0.71 vs. 0.71, respectively). Hosmer-Lemeshow goodness of fit test for the custom model for 28-day mortality in the development and validation cohort showed no evidence of inadequate fit ($\chi^2_{df=3} = 0.37$, $p = 0.95$ and $\chi^2_{df=3} = 1.04$, $p = 0.79$, respectively).

## DISCUSSION

We developed and validated a simple, easily calculable scoring model that accurately predicts hospital mortality for patients with ALI. Our simple point score, incorporating age, 24-hour fluid balance, hematocrit, and bilirubin, is able to discriminate patients with high mortality from those with a lower mortality. Importantly, observed mortality in the validation dataset fell within predicted mortality ranges for the point total strata, indicating good model calibration. Furthermore, the accuracy of the model's prediction for 28-day mortality was similar to that predicting hospital mortality. These results support the use of this model as a useful clinical tool for prognostication, classification, and counseling.

Our results are notable for the excellent concordance or calibration between our custom model's predicted mortality rate and the observed mortality in each point strata within the validation cohort. Although the AUC of our model in the validation cohort was worse than in the development cohort, calibration remained intact. Discrimination refers to a model's ability to distinguish survivors from nonsurvivors. The AUC represents the probability that a patient who died had a greater predicted probability of dying than a patient who survived. Calibration refers to the agreement between predicted probabilities and the actual, observed probabilities. Ideally, a predictive model should have excellent discrimination (AUC >0.9) and calibration (observed rates = predicted rates). Maximizing calibration is of primary importance when a model is used to counsel patients or their families about prognosis (15), because patients and their families are more interested in accurate assessment of the probability of death (calibration), not necessarily how sick the patient is relative to other patients (discrimination) (15).

This model can be used to inform prognosis (e.g., in counseling patients or families) but should not be used for decision making (e.g., withdrawal of support). The literature documenting the presence of cognitive biases in physician decision making is extensive (8, 10). Confronted with the task of prognosticating in the complex environment of the ICU, physicians must assess the probability of an uncertain event. Physicians often use heuristics or simple rules of thumb in place of explicit analysis of probabilities to reduce these complex tasks to simpler judgments (8). Although often useful when used by experienced ICU attending physicians (32), these heuristics can lead to severe errors in assessing the probability of an event. For example, the availability of recent memories (e.g., "the last patient I cared for…") (8, 10), an aversion

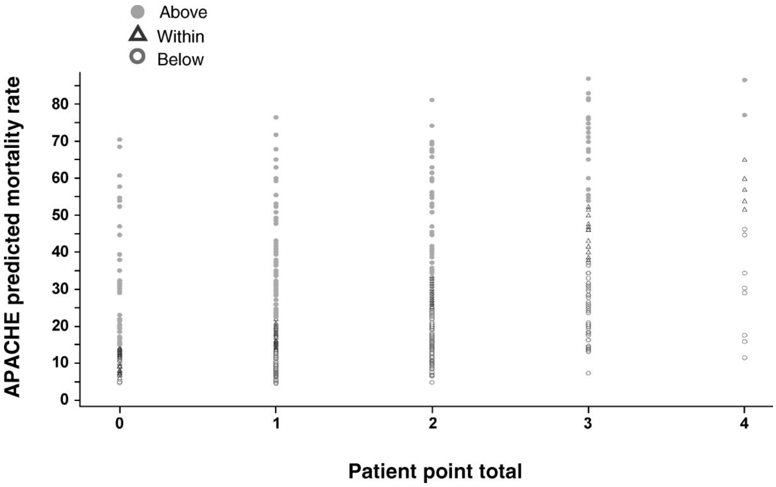


Figure 1. Calibration plot. For each patient in the validation dataset, the predicted mortality estimated from the Acute Physiology Assessment and Chronic Health Evaluation (*APACHE*) III score is plotted against the point total from the custom model developed from the Acute Respiratory Distress Syndrome Network low tidal volume study. Patients with an APACHE III predicted mortality overlapping with the custom model predicted mortality in the validation cohort are shown using triangles (within). Closed circles (above) indicate patients where APACHE III predicts a greater rate of death than predicted by the simple model. Open circles (below) indicate patients where APACHE III predicts a lower rate of death than predicted by the simple model.

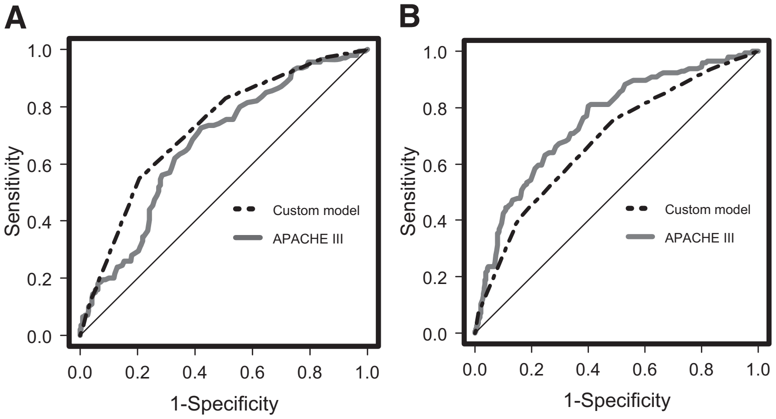


Figure 2. Receiver operating characteristic (ROC) curves for the custom model. *A*, comparison of the ROC curves for the custom model and Acute Physiology Assessment and Chronic Health Evaluation (*APACHE*) III score (area 0.72 vs. 0.67, $p = 0.09$) in the development cohort. *B*, comparison between custom model and APACHE III (area 0.68 vs. 0.75, $p = 0.03$) in the validation cohort.

to change therapeutic course (status-quo bias) (11, 33), or the potential to feel more responsible for an adverse outcome because of active treatment compared with inaction (regret/outcome bias) (10) can unduly influence a physician's estimates of prognosis in the ICU. There are often additional factors that ought not play a role in prognostic decision making, such as physician age, experience and religion, patient age and race, and other conscious or unconscious biases that impede rational and compassionate decision making in critically ill patients (9, 34–37). These biases may contribute to the discrepancy between an attending physician's predicted outcome and the patient's actual outcome (38).

For these reasons, there is a great need for objective measures to facilitate prognostication in critically ill patients who are immune to bias and subjectivity. To date, however, experts advocate against using traditional severity-of-illness measures (e.g., APACHE, Simplified Acute Physiology Score) for decision making at the end of life for multiple reasons (32, 39, 40). There is little evidence to suggest that prognostication systems influence the physician's decisions on caring for patients at the end of life (41). Additional objections stem from the inability of severity scores to convey uncertainty in estimated probabilities of death, the poor concordance between individual predictions among different severity models (39), the poor performance of such models at the extremes of estimated probabilities (e.g., close to zero or to one), and the complexity involved in their calculation (42). On the basis of the above limitations, we caution physicians in the solitary use of our model purely for decision making in individual patients; ICU severity of illness scores, including our point score, will never predict patient outcomes with 100% certainty. Although accurate for populations of patients, such models can never truly account for all uncertainty when applied to individuals. Nonetheless, families value prognostic discussions and use mortality estimates to prepare emotionally for the possibility that a patient may not survive even when they appreciate that prognostic estimates may not be correct (43, 44). Providing stratum-specific estimates of mortality, such as those provided by our point score, to patients and their families has been recommended by many risk communication experts (45, 46).

Although the use of scoring systems as a sole guide to making decisions about whether to initiate or continue to provide intensive care is inappropriate (40), they can provide an objective means for providers to inform their own assessment of prognosis. Combining clinician estimates of mortality with model estimates of mortality improves one's overall ability to discriminate patients who live from those who die compared with either estimate alone (41, 47). Given physicians' pessimistic estimates of mortality, whether combining physician and model estimates improves agreement between the expected and actual mortality is still unclear (47, 48).

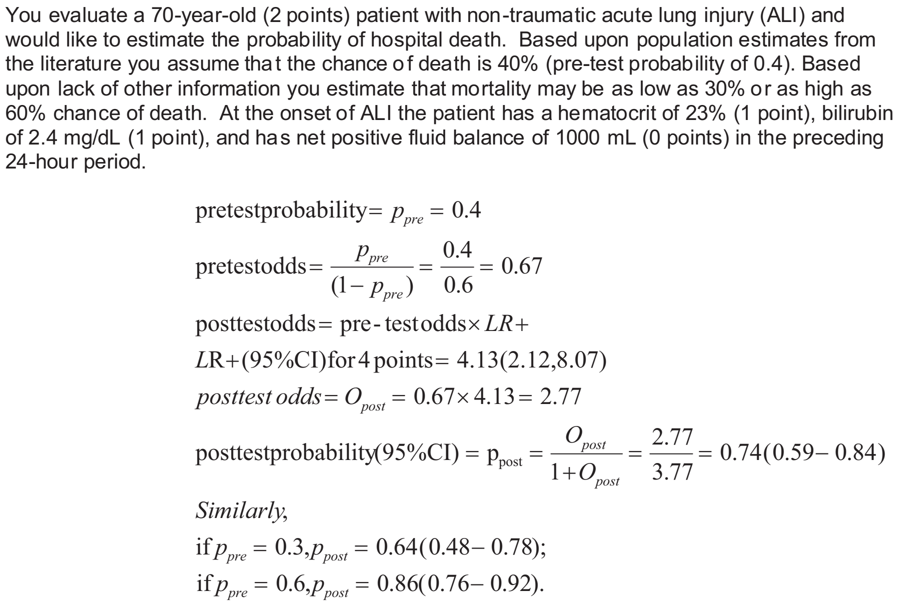
Providers can use the LRs from our model at the bedside similarly to a diagnostic test to estimate the posttest probability of death. Figure 3 illustrates a hypothetical "case study" examining how a prior probability of death of 0.4 (based on population estimates from the literature) is updated to a probability of 0.74 with the knowledge that the patient's point score is 4. It is important to note that population-based data support a pretest mortality in all patients with ALI of approximately 40% (49). Given this estimate, most patients with ALI will have posttest mortalities, indicating a significant chance of surviving to hospital dis-

Table 5. Predicted and observed 28-day mortality in the derivation set (ARMA) and the validation set (ALVELOLI)

| Total Points | Predicted Mortality % | Predicted Mortality 95% CI | Observed Mortality ARMA | Observed Mortality ALVEOLI | Diagnostic Likelihood Ratio + (95% CI)[a] |
|---|---|---|---|---|---|
| 0 | 6.6 | (3.6–11.8) | 5.4 | 6.2 | 0.20 (0.09–0.45) |
| 1 | 13.2 | (9.1–18.7) | 13.0 | 10.4 | 0.42 (0.29–0.60) |
| 2 | 24.6 | (20.0–29.8) | 25.2 | 25.3 | 1.08 (0.87–1.37) |
| 3 | 41.1 | (33.7–49.0) | 42.2 | 41.8 | 2.33 (1.81–3.00) |
| 4+ | 60.0 | (47.3–71.6) | 55.0 | 53.3 | 3.82 (2.00–7.27) |

CI, confidence interval; ARMA, Acute Respiratory Management in Acute Respiratory Distress Syndrome; ALVEOLI, Assessment of Low Tidal Volume and Elevated End-Expiratory Pressure to Obviate Lung Injury.

[a] Pooled likelihood ratios for ARMA and ALVEOLI. Positive likelihood ratio (LR+) can be multiplied by the pretest odds of outcome to get the posttest odds of outcome. Pretest odds can be calculated as $p/1-p$, where $p$ = pretest probability of disease. Posttest probability is then calculated as (posttest odds/1 + posttest odds).



You evaluate a 70-year-old (2 points) patient with non-traumatic acute lung injury (ALI) and would like to estimate the probability of hospital death. Based upon population estimates from the literature you assume that the chance of death is 40% (pre-test probability of 0.4). Based upon lack of other information you estimate that mortality may be as low as 30% or as high as 60% chance of death. At the onset of ALI the patient has a hematocrit of 23% (1 point), bilirubin of 2.4 mg/dL (1 point), and has net positive fluid balance of 1000 mL (0 points) in the preceding 24-hour period.

$$\text{pretestprobability} = p_{pre} = 0.4$$

$$\text{pretestodds} = \frac{p_{pre}}{(1-p_{pre})} = \frac{0.4}{0.6} = 0.67$$

$$\text{posttestodds} = \text{pre-testodds} \times LR+$$

$$LR+(95\%\text{CI})\text{for 4 points} = 4.13(2.12,8.07)$$

$$\textit{posttest odds} = O_{post} = 0.67 \times 4.13 = 2.77$$

$$\text{posttestprobability}(95\%\text{CI}) = \text{p}_{\text{post}} = \frac{O_{post}}{1+O_{post}} = \frac{2.77}{3.77} = 0.74(0.59-0.84)$$

*Similarly,*

$$\text{if } p_{pre} = 0.3, p_{post} = 0.64(0.48-0.78);$$

$$\text{if } p_{pre} = 0.6, p_{post} = 0.86(0.76-0.92).$$

Figure 3. Example calculation of posttest probability of death. Confidence intervals for the posttest probability integrate uncertainty in the likelihood ratio. *LR+*, positive likelihood ratio; *CI*, confidence interval.

charge. We also stress that, in practice, providers often have an uncertainty in their estimated pretest probability of death. Our analyses do not incorporate this uncertainty and, thus, confidence intervals around the posttest probabilities are too narrow.

There are several strengths to our analysis. We used a well-defined cohort of patients with ALI cared for in hospitals throughout the United States. We subsequently validated our model using an independent cohort of patients arising from a similar patient population. Finally, our score, using only four readily available clinical variables, is considerably easier to calculate than the APACHE III predicted probability of death or Simplified Acute Physiology Score III predicted probability of death, yet maintaining excellent discrimination and calibration.

We also recognize several limitations to our analysis. First, our model was derived on data from the ARDSNet low tidal volume study, a study conducted over 10 years ago. The mortality of ALI has decreased over time as implementation of evidence-based therapy in this disease improved (50). We attempted to address this limitation by validating the model in a more contemporary population of patients (ALVEOLI); nevertheless, our model may perform differently in more current ALI cohorts. Second, our derivation population had a small number of deaths, limiting our ability to evaluate all potential predictors of death without overfitting the model (51). Third, in contrast to development of APACHE III, our model development was limited to variables available in the dataset; we were unable to evaluate some potentially important predictors, such as pulmonary dead space and positive end-expiratory pressure responsiveness, because they were not collected routinely in this cohort (52–54). We were also unable to evaluate the predictive ability of other comorbidities, such as chronic liver disease and metastatic cancer (55), because patients with these underlying illnesses were excluded from the parent study. Fourth, in addition to excluding trauma patients, we excluded 15% (64 of 414) of the cohort because of missing data to maximize the utility of our model in practice. This may have influenced the variables selected for our model and may bias the mortality within each strata when applied. Validation of our model in populations with complete data is important before its routine use. Fifth, our model was derived in a cohort collected from multiple academic tertiary care hospitals participating in a randomized trial with specific exclusion criteria. Documented differences between academic-based and community-based patients with ALI and patients enrolled vs. not enrolled in randomized trials may prevent generalization to the broader community (49). Furthermore, our inclusion of fluid balance, a treatment-dependent variable, may influence the performance of our model under different practice patterns. Further validation of this model in a contemporary, large, multicenter study should be performed before widespread adoption. Finally, APACHE III was developed to predict mortality using data during the first 24 hours of ICU stay; therefore, our use of APACHE III scores generated at the time of enrollment may have resulted in underperformance of APACHE III.

## CONCLUSIONS

We have developed a simple prognostic score that accurately identifies groups of patients with ALI at high risk of death. This model can facilitate a provider's assessment of prognosis when informing patients and their families about the possible outcomes of ALI. Before widespread use, this model should be validated in contemporary nonclinical trial populations.

## REFERENCES

1. Herridge MS, Cheung AM, Tansey CM, et al: One-year outcomes in survivors of the acute respiratory distress syndrome. *N Engl J Med* 2003; 348:683–693
2. Rubenfeld GD, Herridge MS: Epidemiology and outcomes of acute lung injury. *Chest* 2007; 131:554–562
3. Prendergast TJ, Claessens MT, Luce JM: A national survey of end-of-life care for critically ill patients. *Am J Respir Crit Care Med* 1998; 158:1163–1167
4. Cook D, Rocker G, Marshall J, et al: Withdrawal of mechanical ventilation in anticipation of death in the intensive care unit. *N Engl J Med* 2003; 349:1123–1132
5. Stapleton RD, Wang BM, Hudson LD, et al: Causes and timing of death in patients with ARDS. *Chest* 2005; 128:525–532
6. Luce JM, White DB: The pressure to withhold or withdraw life-sustaining therapy from critically ill patients in the United States. *Am J Respir Crit Care Med* 2007; 175: 1104–1108
7. White DB, Engelberg RA, Wenrich MD, et al: Prognostication during physician-family discussions about limiting life support in intensive care units. *Crit Care Med* 2007; 35: 442–448
8. Tversky A, Kahneman D: Judgment under uncertainty: Heuristics and biases. *Science* 1974; 185:1124–1131
9. Christakis NA, Asch DA: Biases in how physicians choose to withdraw life support. *Lancet* 1993; 342:642–646
10. Bornstein BH, Emler AC: Rationality in medical decision making: A review of the literature on doctors' decision-making biases. *J Eval Clin Pract* 2001; 7:97–107
11. Aberegg SK, Haponik EF, Terry PB: Omission bias and decision making in pulmonary and critical care medicine. *Chest* 2005; 128: 1497–1505
12. 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
13. Moreno RP, Metnitz PG, Almeida E, et al: SAPS 3—From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med* 2005; 31:1345–1355
14. Herridge MS: Prognostication and intensive care unit outcome: The evolving role of scoring systems. *Clin Chest Med* 2003; 24: 751–762
15. Justice AC, Covinsky KE, Berlin JA: Assessing the generalizability of prognostic information. *Ann Intern Med* 1999; 130:515–524
16. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. The Acute Respiratory Distress Syndrome Network. *N Engl J Med* 2000; 342: 1301–1308
17. Bernard GR, Artigas A, Brigham KL, et al: The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. *Am J Respir Crit Care Med* 1994; 149: 818–824
18. 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
19. Murray JF, Matthay MA, Luce JM, et al: An expanded definition of the adult respiratory distress syndrome. *Am Rev Respir Dis* 1988; 138:720–723
20. Hastie T, Tibshirani R: Generalized additive-models—Some applications. *J Am Stat Assoc* 1987; 82:371–386
21. Efron B, Tibshirani R: An Introduction to the Bootstrap. New York, Chapman & Hall, 1993
22. Altman DG, Andersen PK: Bootstrap investigation of the stability of a Cox regression model. *Stat Med* 1989; 8:771–783
23. Austin PC, Tu JV: Bootstrap methods for developing predictive models. *Am Statistician* 2004; 58:131–137
24. Moons KG, Harrell FE, Steyerberg EW: Should scoring rules be based on odds ratios or regression coefficients? *J Clin Epidemiol* 2002; 55:1054–1055
25. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29–36
26. Sackett DL, Haynes RB, Tugwell P: Clinical Epidemiology: A Basic Science for Clinical Medicine. First Edition. Boston, Little, Brown, 1985
27. Deeks JJ, Altman DG: Diagnostic tests 4: Likelihood ratios. *BMJ* 2004; 329:168–169
28. Lemeshow S, Hosmer DW Jr: A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982; 115:92–106
29. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 1988; 44:837–845
30. 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
31. Rubin DB, Schenker N: Multiple imputation in health-care databases: An overview and some applications. *Stat Med* 1991; 10: 585–598
32. Sinuff T, Adhikari NK, Cook DJ, et al: Mortality predictions in the intensive care unit: Comparing physicians with scoring systems. *Crit Care Med* 2006; 34:878–885
33. Redelmeier DA, Shafir E: Medical decision making in situations that offer multiple alternatives. *JAMA* 1995; 273:302–305
34. Cook DJ, Guyatt GH, Jaeschke R, et al: Determinants in Canadian health care workers of the decision to withdraw life support from the critically ill. Canadian Critical Care Trials Group. *JAMA* 1995; 273:703–708
35. Christakis NA, Asch DA: Medical specialists prefer to withdraw familiar technologies when discontinuing life support. *J Gen Intern Med* 1995; 10:491–494
36. Christakis NA, Asch DA: Physician characteristics associated with decisions to withdraw life support. *Am J Public Health* 1995; 85: 367–372
37. Hinkka H, Kosunen E, Metsanoja R, et al: Factors affecting physicians' decisions to forgo life-sustaining treatments in terminal care. *J Med Ethics* 2002; 28:109–114
38. Detsky AS, Stricker SC, Mulley AG, et al: Prognosis, survival, and the expenditure of hospital resources for patients in an intensive-care unit. *N Engl J Med* 1981; 305: 667–672
39. Lemeshow S, Klar J, Teres D: Outcome prediction for individual intensive care patients: Useful, misused, or abused? *Intensive Care Med* 1995; 21:770–776
40. Consensus statement of the Society of Critical Care Medicine's Ethics Committee regarding futile and other possibly inadvisable treatments. *Crit Care Med* 1997; 25:887–891
41. Knaus WA, Harrell FE Jr, Lynn J, et al: The SUPPORT prognostic model. Objective estimates of survival for seriously ill hospitalized adults. Study to understand prognoses and preferences for outcomes and risks of treatments. *Ann Intern Med* 1995; 122:191–203
42. Metnitz PG, Moreno RP, Almeida E, et al: SAPS 3—From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. *Intensive Care Med* 2005; 31:1336–1344
43. Zier LS, Burack JH, Micco G, et al: Doubt and belief in physicians' ability to prognosticate during critical illness: The perspective of surrogate decision makers. *Crit Care Med* 2008; 36:2341–2347
44. Evans LR, Boyd EA, Malvar G, et al: Surrogate decision-makers' perspectives on discussing prognosis in the face of uncertainty. *Am J Respir Crit Care Med* 2008; 179:48–53
45. Gigerenzer G, Edwards A: Simple tools for understanding risks: From innumeracy to insight. *BMJ* 2003; 327:741–744
46. Thomson R, Edwards A, Grey J: Risk communication in the clinical consultation. *Clin Med* 2005; 5:465–469
47. Rocker G, Cook D, Sjokvist P, et al: Clinician predictions of intensive care unit mortality. *Crit Care Med* 2004; 32:1149–1154
48. Wildman MJ, Sanderson C, Groves J, et al: Implications of prognostic pessimism in patients with chronic obstructive pulmonary disease (COPD) or asthma admitted to intensive care in the UK within the COPD and asthma outcome study (CAOS): Multicentre observational cohort study. *BMJ* 2007; 335: 1132
49. Rubenfeld GD, Caldwell E, Peabody E, et al: Incidence and outcomes of acute lung injury. *N Engl J Med* 2005; 353:1685–1693
50. Zambon M, Vincent JL: Mortality rates for

patients with ALI/ARDS have decreased over time. *Chest* 2008; 133:1120–1127

51. Peduzzi P, Concato J, Kemper E, et al: A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; 49:1373–1379
52. Nuckton TJ, Alonso JA, Kallet RH, et al: Pulmonary dead-space fraction as a risk factor for death in the acute respiratory distress syndrome. *N Engl J Med* 2002; 346: 1281–1286
53. Ware LB: Prognostic determinants of acute respiratory distress syndrome in adults: Impact on clinical trial design. *Crit Care Med* 2005; 33:S217–S222
54. Gattinoni L, Caironi P, Cressoni M, et al: Lung recruitment in patients with the acute respiratory distress syndrome. *N Engl J Med* 2006; 354:1775–1786
55. 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