ONLINE DATA SUPPLEMENT

Predictors of quality of life at 6 months after acute respiratory distress syndrome

Samuel M. Brown, MD MS[1,2,3], Emily Wilson, MStat[1,2], Angela P. Presson, PhD[4], Chong Zhang, MS[4], Victor Dinglas, MPH[5,6], Tom Greene, PhD[4], Ramona O. Hopkins, PhD[1,2,7], Dale M. Needham, MD PhD[5,6,8] with the National Institutes of Health NHLBI ARDS Network

[1]Center for Humanizing Critical Care, Intermountain Healthcare, Murray, UT
[2]Department of Medicine, Pulmonary and Critical Care Division, Intermountain Medical Center, Murray, UT
[3]Pulmonary and Critical Care, University of Utah School of Medicine, Salt Lake City, UT
[4]Study Design and Biostatistics Center and Division of Epidemiology, University of Utah School of Medicine, Salt Lake City, UT
[5]Outcomes After Critical Illness and Surgery Group, Johns Hopkins University, Baltimore, MD
[6]Division of Pulmonary and Critical Care Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD
[7]Department of Psychology and Neuroscience Center, Brigham Young University, Provo, UT
[8]Department of Physical Medicine and Rehabilitation, School of Medicine, Johns Hopkins University, Baltimore, MD

CONTENTS:

APPENDIX 1: Description of Statistical Methods

A. Definition of Outcome and Predictor Variables

Since EQ-5D utilities are highly negatively skewed, we applied a Box Cox transformation ((1.1 – EQ-5D)^0.3 – 1)/0.3 to EQ-5D utilities. We then standardized the transformed variable to have mean 0 and a standard deviation (SD) of 1, so that regression analysis interpretations could be made in SD increments.

Because some exposure-outcome relationships are likely to be U-shaped, we *a priori* considered quadratic terms for the following variables: sodium, temperature, systolic blood

pressure, heart rate, mean arterial blood pressure, respiratory rate, hemoglobin/hematocrit, white blood cell count, platelet count, potassium, serum glucose, alcohol use (i.e., Alcohol Use Disorders Identification Test [AUDIT] score and components), bicarbonate, magnesium, body mass index, arterial pH, $PaCO_2$.

We also *a priori* explored clinically relevant interactions between the following variables: serum creatinine and chronic hemodialysis, blood pressure and norepinephrine equivalent dosage, minute ventilation and $PaCO_2$, PaO_2:FiO_2 ratio and positive end expiratory pressure (PEEP).

For the secondary analysis evaluating patients alive at the time of hospital discharge, we summarized variables from study days 0 to 3 using 4 different summary measures: maximum, minimum, mean values, and the difference between the first and last values (to capture trends in the data). These summary variables were entered into the regression models rather than the original untransformed variables.

We considered including hospital study site as a predictor. However, the large number of participating hospitals risked over-fitting the regression model. We therefore evaluated the association of hospital site with outcome using ANOVA: because the p was > 0.05, we excluded the hospital study site variable from the model.

While the majority of patients were ventilated using volume-controlled ventilation, for patients who were ventilated using airway pressure release ventilation (APRV; 3% of patients), we estimated equivalencies in terms of PEEP, peak inspiratory pressure (PIP), and tidal volume. We calculated driving pressure as plateau pressure minus PEEP.(1)

B. Approach to Missing Data

As recommended in prior literature,(2) we imputed missing values using single imputation for variables where missingness was ≤ 15% and excluded variables with missingness > 15% (see details of imputation strategy in eTables 1 and 2). Where an imputation strategy appeared clinically relevant (e.g., missing is normal or last value carried forward), we employed that imputation strategy for a given variable. Where there was no clinically relevant imputation approach, we used a multivariate adaptive regression splines (MARS) regression(3) of the remaining variables to impute missing values for these variables. MARS employs an objective, automated feature selection strategy that tests for improvements in the squared error of a regression model and limits problems with simple forward or backward stepwise regression techniques.(4) To avoid problems with overfitting, we included predictors in the MARS regressions only where missingness was <7%. We chose the 7% threshold both because it represented a natural threshold in the data and because it was twice as conservative as the 15% threshold we employed to determine which variables had to be rejected for unduly high missingness. Single rather than multiple imputation was used due to the complexity of the analyses, including primary, secondary, and sensitivity analyses, paired with 10-fold cross validations in each case.

C. Penalized Regression

LASSO (Least Absolute Shrinkage and Selection Operator) penalized regression has been widely applied in analyses with a large number of candidate predictors, as, for example, in genetics research.(5, 6) Group LASSO was used here so that all levels of a single categorical predictor (such as degree of independent living at the time of hospital admission) would enter or leave the model together. LASSO techniques employ a penalty tuning parameter (λ) to systematically penalize coefficients for candidate predictor variables toward zero. Only the most predictive variables remain in the model after LASSO penalization. A regularization path is generated and navigated sequentially to identify feature sets. The specific value of λ has conventionally been selected using 10-fold cross validation (CV)(7) to determine the value of λ that minimizes error. We used a nested 10-fold CV to select the value of λ that provided the most parsimonious model whose CV error was within one standard error of the minimum CV error. A sample regularization path is displayed in eFigure 1 for illustration.

While LASSO penalized regression performs better than stepwise variable selection,(8, 9) all automated variable selection methods must select from among a large number of possible combinations of predictors. Because typical p values and 95% confidence intervals do not account for the uncertainty related to the choice of variables to include in the regression model ("feature selection"), we consider reported p values and confidence intervals to be *conditional upon feature selection*. We make this conditionality explicit to avoid a common error in reporting regression results after iterative feature selection.(10, 11) We used a 10-fold cross-validation technique on the selected models to estimate more conservative root mean square error (RMSE), mean absolute percentage error (MAPE), and R^2. This cross-validation to describe the final regression equation was separate from the cross validation used to select the appropriate value for λ.

To check linear regression assumptions of linearity and homoscedasticity, we used the residual plots and fit statistics provided by Pena et al.(12)

D. Sensitivity analysis that includes death as an outcome

Following common practice, we employed available admission predictors to predict 6-month health utility outcomes among individuals who survived to 6 months. While this practice requires conditioning on survival to followup (information not available at the time of admission), we employed it because it is consistent with typical reasoning about the distribution of outcomes *among survivors*. While our primary analysis was of survivors with available 6-month EQ-5D utility score data, we performed a sensitivity analysis to explore the effect of mortality on the primary results. In this sensitivity analysis, we imputed an EQ-5D utility score of 0(13) for patients who died before 6-month follow-up, and included these patients along with survivors in a new regression model.

Because including dead patients created a bimodal distribution of outcomes (one mode at zero, largely representing dead patients, the other mode representing the most typical utility score among survivors), we transformed the composite utility (i.e., the utility score in live patients and a utility score of zero among deceased patients) into an ordinal outcome and employed penalized ordinal regression. Patients who were dead at follow-up were classified as the lowest ordinal category in this analysis. Our approach in the sensitivity analysis incorporating deceased patients is similar to the trichotomous regression approach of Pollack et al(14) and has been used in other contexts.(15, 16)

E. Exploration of Collinearity and Acute Illness Severity Variables

All feature selection methods search a vast combinatorial space of predictor sets. While feature selection methods, including the LASSO, identify meaningfully predictive predictor sets, they may exclude predictors which or collinear with identified predictors or may not incorporate meaningful predictor sets. As we observed a lack of association between acute illness severity markers and 6-month health utility, we undertook a *post hoc* analysis to evaluate collinearity among predictors and the effect of acute severity of illness predictors on our primary model.

Before performing the post hoc analysis, we defined the following variables as potentially relevant representatives of acute illness severity, all measured at the time of enrollment: Brussels organ dysfunction score, APACHE III score, PaO_2:FiO_2 ratio, and PEEP.

Pairwise correlation was calculated for candidate variables. We inspected the predictor variables which exhibited a correlation > 0.5 with the seven main predictors in our primary model. Any variables meeting those criteria were considered as potential replacements for the primary variable with which they were correlated.

In succession, each of the acute severity of illness predictors was added to the primary linear regression model to see whether the addition of that predictor changed the model substantially. We evaluated differences in log likelihood, R^2, and p values for predictors to determine whether the new variable substantially affected the primary model.

APPENDIX 2: Description of Results of Additional Analyses

A. Exploration of collinearity and acute illness severity variables

1. Collinearity

In order to assess the risks that collinearity may have posed to the final identification of features, we calculated correlations (Spearman's rho) for both the predictors in our main model and for all predictors. For the predictors in the main model, we focused on correlation with Spearman's rho > 0.5. For the general predictors, we emphasized Spearman's rho > 0.75. The primary associations of relevance among the predictors with p > 0.05 in our primary model was high correlation between ever-smoker and current-smoker (Spearman's rho 0.66) and between height and female sex (Spearman's rho -0.70). To further evaluate whether ever-smoker or current-smoker was the more important predictor, we compared univariate and multivariate associations, exchanging ever-smoker for current-smoker in the primary model. On both univariate and multivariate modeling, current-smoker had a higher coefficient, lower p value; the multivariate model incorporating current-smoker had a higher R^2 and adjusted R^2.
In eTable 7 we present the correlations among unselected predictors that were greater than 0.75. None were a part of the primary model; the vast majority were measurements of very similar entities (e.g., maximum or minimum serum creatinine or measures of alcohol abuse).

2. Acute Illness Severity

We explored the effect of prespecified acute severity of illness variables on the primary model to assure that our finding—acute severity of illness was not associated with 6-month health utility—did not represent Type 2 statistical error. We evaluated the relevance of *a priori* defined acute illness severity variables that had good face validity (i.e., PF ratio, PEEP, day 0 Brussels organ dysfunction score, and APACHE III score). We evaluated the univariate associations with 6-month health utility and evaluated whether these variables influenced the R^2 of the primary model.

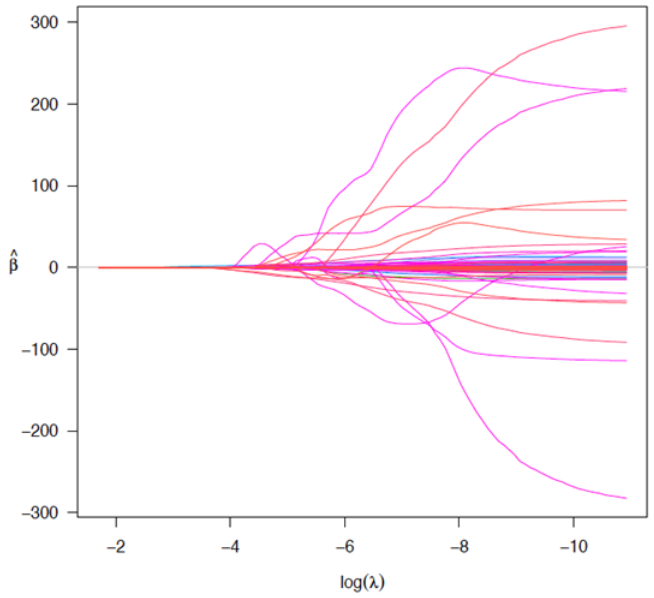
Univariate and multivariate associations were not significant, and inclusion of the elements in the model did not change the adjusted R^2 of the primary model. The results of this analysis are displayed in eTable 8.

B. Post hoc analysis of patients with retrospective "baseline" EQ5D

A subset of ALTOS patients were asked to respond to the EQ5D questionnaire with regard to their health before the hospitalization for ARDS. This occurred during a scheduled followup visit. 199 had such a visit, and 195 (98%) of them had EQ5D health utility. Among that group we first fit our primary model. We then added to that primary model the "baseline" EQ5D in order to assess, among that subgroup (which represented 32% of the study population), which predictors were significant after controlling for baseline EQ5D. Results of these two linear regression models are presented in eTable 8.

APPENDIX 3: Figures and Tables

eFigure 1. Regularization path for penalized regression



In penalized regression, the coefficients of candidate predictors are penalized (deflated), as one moves from right to left on this graph. Each line represents the "regularization path" for a specific variable. The lambda determines which set of predictors is the final set for the regression model. The lambda is chosen via 10-fold cross-validation. In this graph, the lambda is depicted as a log(lambda).

eTable 1: Summary of Baseline Candidate Variables

Variable Description	Variable Type	% Missingness	Imputation Strategy
Enrollment Information			
Age (at screening)	Continuous	0	-
Gender	Binary	0	-
Hispanic or Latino	Binary	0	-
Race is White	Binary	0	-
Race is Black or African American	Binary	0	-
Race is American Indian or Alaskan Native	Binary	0	-
Race is Asian	Binary	0	-
Race is Native Hawaiian or other Pacific Islander	Binary	0	-
ARDSNet study enrolled in	Nominal	0	-
ARDSNet study site	Nominal	0	-
ICU type	Nominal	0	-
Lung injury due to trauma	Binary	0	-
Lung injury due to sepsis	Binary	0	-
Source of sepsis	Nominal	0	-
Lung injury due to transfusion	Binary	0	-
Lung injury due to aspiration	Binary	0	-
Lung injury due to pneumonia	Binary	0	-
Other lung injury	Binary	0	-
FiO_2 at screening	Continuous	0	-
P:F ratio from screening	Continuous	0	-
Body mass index at admission	Continuous	< 1	MARS
Height (cm)	Continuous	< 1	-
Shock index	Continuous	0	-
APACHE III Demographics			
Location patient was admitted to the ICU from	Nominal	0	-
Patient residence prior to hospitalization	Nominal	0	-
Post-operative from elective surgery	Binary	0	-
ICU readmission	Binary	0	-
ICU readmission within 24 hours	Binary	0	-
Chronic health information available	Binary	0	-
Chronic or peritoneal dialysis	Binary	0	-
AIDS comorbidity	Binary	< 1	MARS
Leukemia (AML, CML, ALL, multiple myeloma)	Binary	0	-
Non-Hodgkin's lymphoma	Binary	0	-
Solid tumor with metastasis	Binary	0	-
Immune suppression	Binary	0	-
Hepatic failure with coma or encephalopathy	Binary	0	-
Cirrhosis	Binary	< 1	MARS
Diabetes Mellitus	Binary	0	-
Hypertension comorbidity	Binary	< 1	MARS
Prior myocardial infarction	Binary	< 1	MARS
Congestive heart failure	Binary	0	-
Peripheral vascular disease	Binary	0	-
Prior stroke with sequelae	Binary	< 1	MARS
Dementia	Binary	0	-
Chronic pulmonary disease	Binary	0	-
Arthritis	Binary	< 1	MARS
Peptic ulcer disease	Binary	< 1	MARS
Vasopressor use in 24 hours prior to randomization	Binary	0	-

APACHE III Physiology (values from 24 hours preceding randomization)			
Lowest temperature	Continuous	< 1	MARS
Highest temperature	Continuous	< 1	MARS
Lowest systolic blood pressure	Continuous	< 1	MARS
Highest systolic blood pressure	Continuous	< 1	MARS
Lowest mean arterial pressure	Continuous	1	MARS
Highest mean arterial pressure	Continuous	1	MARS
Lowest heart rate	Continuous	< 1	MARS
Highest heart rate	Continuous	< 1	MARS
Lowest respiratory rate	Continuous	< 1	MARS
Highest respiratory rate	Continuous	< 1	MARS
Ventilated when lowest respiratory rate occurred	Binary	< 1	MARS
Ventilated when highest respiratory rate occurred	Binary	< 1	MARS
Urine output	Continuous	1	MARS
Total fluid output	Continuous	1	MARS
Total fluid intake	Continuous	< 1	MARS
Lowest hematocrit	Continuous	< 1	MARS
Highest hematocrit	Continuous	< 1	MARS
Lowest white blood cell	Continuous	< 1	MARS
Highest white blood cell	Continuous	< 1	MARS
Lowest platelets	Continuous	< 1	MARS
Lowest serum sodium	Continuous	< 1	MARS
Highest serum sodium	Continuous	< 1	MARS
Lowest serum potassium	Continuous	< 1	MARS
Highest serum potassium	Continuous	< 1	MARS
Highest serum BUN	Continuous	< 1	MARS
Lowest serum creatinine	Continuous	< 1	MARS
Highest serum creatinine	Continuous	< 1	MARS
Lowest serum glucose	Continuous	< 1	MARS
Highest serum glucose	Continuous	< 1	MARS
Lowest serum albumin	Continuous	8	MARS
Highest serum albumin	Continuous	8	MARS
Highest serum bilirubin	Continuous	10	MARS
Lowest serum bicarbonate	Continuous	< 1	MARS
APACHE III score	Continuous	2	Components/MARS
The Alcohol Use Disorders Identification Test (AUDIT) Questionnaire and Smoking Assessment			
How often do you drink alcohol	Ordinal	6	MARS
How many drinks on a typical day when drinking	Ordinal	6	MARS
How often do you have six or more drinks	Ordinal	7	MARS
How often were you not able to stop drinking once started	Ordinal	8	MARS
How often have you failed to do what was normally expected of you because of drinking	Ordinal	8	MARS
How often do you need a drink in the morning	Ordinal	8	MARS
How often to you feel guilt or remorse after drinking	Ordinal	9	MARS
How often have you been unable to remember the previous night because of drinking	Ordinal	8	MARS
Have you or someone else been injured as a result of your drinking	Ordinal	8	MARS
Has someone been concerned about your drinking	Ordinal	9	MARS
Sum of AUDIT Questionnaire	Continuous	10	Components/MARS
Ever smoker	Binary	6	MARS
Pack years	Continuous	7	MARS
Current smoker	Binary	3	MARS

Baseline Ventilator Parameters (most recent values prior to randomization)			
Calculated delivered tidal volume	Continuous	13	Derived[1]; MARS if all missing
Total respiratory rate	Continuous	1	Derived[2]; MARS if all missing
Total minute ventilation	Continuous	3	Derived[3]; MARS if all missing
PEEP	Continuous	1	MARS
FiO_2 prior to randomization	Continuous	0	-
SpO_2 prior to randomization	Continuous	< 1	Ellis equation from PaO_2
Peak inspiratory pressure	Continuous	8	MARS
Mean airway pressure	Continuous	9	MARS
Volume control vent mode	Binary	0	-
Calculated delivered tidal volume after initial vent change	Continuous	51	Set to tidal (original value)
PEEP after initial vent change	Continuous	50	Set to PEEP (original value)
Baseline Vital Signs and Labs (values closest to the time preceding randomization)			
Heart rate	Continuous	0	-
Systolic blood pressure	Continuous	0	-
Diastolic blood pressure	Continuous	0	-
Mean arterial pressure	Continuous	12	Derived[4]
Temperature	Continuous	< 1	MARS
Hemoglobin	Continuous	< 1	MARS
Sodium	Continuous	< 1	MARS
Potassium	Continuous	< 1	MARS
Glucose	Continuous	< 1	MARS
Serum bicarbonate	Continuous	< 1	MARS
Serum phosphorous	Continuous	11	MARS
Serum magnesium	Continuous	11	MARS
Albumin	Continuous	14	MARS
Number of quadrants with opacities	Continuous	0	-
Corticosteroids	Binary	0	-
Acute renal failure or receiving renal replacement therapy	Binary	0	-
Neosynephrine dose (weight adjusted)	Continuous	0	-
Epinephrine dose (weight adjusted)	Continuous	0	-
Dopamine dose (weight adjusted)	Continuous	0	-
Norepinephrine dose (weight adjusted)	Continuous	0	-
Total vasopressor dosage (in Norepinephrine equivalent)	Continuous	0	-
Arterial Blood Gas Summary			
Maximum P:F ratio	Continuous	4	MARS
Minimum P:F ratio	Continuous	4	MARS
Mean P:F ratio	Continuous	4	MARS
Difference of max P:F ratio and minP:F ratio	Continuous	4	Difference of MARS values
Maximum PaO_2	Continuous	3	MARS
Minimum PaO_2	Continuous	3	MARS
Mean PaO_2	Continuous	3	MARS
Difference of max PaO_2 and min PaO_2	Continuous	3	Difference of MARS values
Maximum arterial pH	Continuous	3	MARS
Minimum arterial pH	Continuous	3	MARS
Mean arterial pH	Continuous	3	MARS
Difference of max arterial pH and min arterial pH	Continuous	3	Difference of MARS values
Maximum $PaCO_2$	Continuous	3	MARS
Minimum $PaCO_2$	Continuous	3	MARS
Mean $PaCO_2$	Continuous	3	MARS
Difference of max $PaCO_2$ and min $PaCO_2$	Continuous	3	Difference of MARS values

Glasgow Coma Scale			
Total Glasgow Coma Scale (GCS) score	Continuous	< 1	Sum of MARS values
GCS Eye Opening	Continuous	< 1	MARS
GCS Motor Response	Continuous	< 1	MARS
GCS Verbal Response	Continuous	< 1	MARS
Length of Stay			
Hospital length of stay*	Continuous	1	MARS
First ICU length of stay*	Continuous	1	Derived using the ratio of HospLOS to ICULOS
Cumulative ICU length of stay*	Continuous	1	Derived using the ratio of HospLOS to ICULOS2

Components/MARS: Calculated from components; MARS used to impute missing components
[1] Derived from minute ventilation / respiratory rate
[2] Derived from minute ventilation / tidal volume
[3] Derived from respiratory rate * tidal volume
[4] Derived from 2/3 * diastolic blood pressure + 1/3 * systolic blood pressure
* Excluded from the primary analysis

eTable 2: Summary of "On Study" (Repeated Measures) Candidate Variables.
For each candidate variable, summary statistics of the Day 0 to Day 3 data, as well as the Day 7 data, were included as possible predictors. If Day 0 data was missing, the baseline data was used.

Description	Variable Type	% Missingness		Imputation Strategy
		Day 0-3[&]	Day 7[&&]	
Brussels Data (24 Hour Worst Value)				
Systolic blood pressure	Continuous	0	6	LVCF
Platelets	Continuous	< 1	12	LVCF
Creatinine	Continuous	< 1	10	LVCF
Vasopressor	Binary	0	5	LVCF
Daily Fluid Data (total for last 24 hours)				
Total fluid intake*	Continuous	0	27	LVCF
Packed red blood cells	Continuous	18	42	Assumed zero
Fresh frozen plasma	Continuous	22	43	Assumed zero
Total fluid out*	Continuous	0	28	LVCF
Total urine output*	Continuous	0	28	LVCF
On Study Ventilator Parameters (values closest to 8 AM)				
Calculated delivered tidal volume*	Continuous	14	58	LVCF
Total respiratory rate*	Continuous	2	45	LVCF
Total minute ventilation*	Continuous	2	46	LVCF
PEEP*	Continuous	2	46	LVCF
Volume control vent mode	Binary	0	0	-
P:F ratio*	Continuous	14	54	LVCF
PaO_2*	Continuous	14	54	LVCF
$PaCO_2$*	Continuous	14	54	LVCF
Arterial pH*	Continuous	14	54	LVCF
On Study Vital Signs and Labs (values closest to 8 AM)				
Heart rate*	Continuous	< 1	26	LVCF
Systolic blood pressure*	Continuous	< 1	26	LVCF
Diastolic blood pressure*	Continuous	< 1	26	LVCF
Temperature*	Continuous	< 1	27	LVCF
Vasopressors in 12 hours prior*	Binary	1	35	LVCB
MAP below 60 in 12 hours prior*	Binary	1	35	LVCB
Neosynephrine dose (weight adjusted)	Continuous	0	0	-
Epinephrine dose (weight adjusted)	Continuous	0	0	-
Dopamine dose (weight adjusted)	Continuous	0	0	-
Norepinephrine dose (weight adjusted)	Continuous	0	0	-
Total vasopressor dosage (in Norepinephrine equivalent)	Continuous	0	0	-
Hemoglobin*	Continuous	< 1	22	LVCF
Sodium*	Continuous	< 1	19	LVCF
Potassium*	Continuous	< 1	19	LVCF
Glucose*	Continuous	< 1	20	LVCF
Serum bicarbonate*	Continuous	1	22	LVCF
Serum phosphorus*	Continuous	8	39	LVCF
Serum magnesium*	Continuous	8	40	LVCF
Albumin*	Continuous	14	42	LVCF
Glasgow Coma Scale (measured on Day 7 only)				
Total Glasgow Coma Scale (GCS) score	Continuous	-	6	Sum of MARS values
GCS Eye Opening	Continuous	-	6	MARS
GCS Motor Response	Continuous	-	6	MARS
GCS Verbal Response	Continuous	-	6	MARS

LVCF: Last value carried forward
LVCB: Last value carried backward
*Day 7 data was not included as a candidate predictor for these variables due to high missingness (> 15%) on Day 7
[&]Percentage of patients missing data for all days 0 through 3
[&&]Excludes patients who died or were discharged prior to Day 7

eTable 3: Secondary ordinal regression, in which death is considered a utility of 0

Variable	OR (95% CI)	p value
Baseline characteristics		
Resides at home with informal help	0.25 (0.16-0.38)	<0.001
Age	0.98 (0.97-0.99)	<0.001
Hospital admission characteristics		
APACHE III score (per 1 unit increase)	0.99 (0.98-0.99)	<0.001

eTable 4: Secondary regression model, based on patients alive at hospital discharge

Variable	LASSO Coefficient*	OLS Coefficient* (95% CI)	p value
Baseline characteristics			
Resides at home with informal help	-0.24	-0.58 (-0.25, -0.92)	**<0.001**
Requires professional help at home OR residence in health care facility	-0.21	-0.44 (-0.89, 0.01)	0.05
Latino Ethnicity	-0.05	-0.31 (-0.05, -0.57)	**0.02**
Female sex	-0.19	-0.31 (-0.16, -0.46)	**<0.001**
AIDS comorbidity	0.35	0.77 (0.15, 1.38)	**0.02**
Age	-0.004	-0.007 (-0.002, -0.013)	**0.01**
Chronic pulmonary disease	-0.004	-0.14 (-0.38, 0.10)	0.25
Ever smoker	-0.006	-0.08 (-0.29, 0.14)	0.48
Hospital admission characteristics			
Nadir respiratory rate on day before enrollment (per 1 breath/min)	0.003	0.01 (0.002, 0.03)	**0.02**
Difference in platelets between day three and day one of admission	0.00	-0.001 (-0.002, 0.00)	0.10
Minimum glucose from first three days of admission	0.00	0.003 (0.00, 0.006)	0.05
Glasgow Coma Scale verbal score on day 7 of admission	0.006	0.02 (-0.03, 0.07)	0.45
Hospital LOS (per day)	-0.003	-0.007 (-0.002, -0.013)	**0.009**
Lifestyle factors			
Current smoker status (at baseline)	-0.18	-0.29 (-0.09, -0.5)	**0.005**
Body mass index (at baseline)	-0.01	-0.02 (-0.008, -0.03)	**<0.001**

*Because of the Box Cox transformation, we inverted the sign of coefficients to aid in interpretation. Positive coefficients are associated with better outcome.
LASSO: Least Absolute Shrinkage and Selection Operator (refers to the penalty used in penalized regression); OLS: ordinary least squares

eTable 5: Characteristics of ARDS survivors included in secondary analysis, based on patients alive at hospital discharge

	Values (N=645)
Baseline characteristics	
Age (years); mean ± SD	49 ± 14
Female sex; n (%)	331 (51%)
Latino or Hispanic; n (%)	56 (9%)
Residence prior to hospitalization; n (%) Home independently Home with informal help Requiring professional help at home OR residence at healthcare facility	 589 (91%) 36 (6%) 20 (3%)
AIDS comorbidity; n (%)	10 (2%)
Hospital admission characteristics	
Nadir respiratory rate on day before enrollment; mean ± SD	18 ± 6
Hospital LOS; median (IQR)	17 (12 – 26)
Lifestyle factor	
Current smoker status (at baseline); n (%)	269 (42%)
Body mass index (at baseline); mean ± SD	30 ± 8

eTable 6: Characteristics of ARDS survivors included in secondary analysis at time of hospital discharge, in which death is considered a utility of 0

	Values (N=978)
Baseline characteristics	
Age (years); mean ± SD	52 ± 16
Residence prior to hospitalization; n (%) Home independently Home with informal help Requiring professional help at home OR residence at healthcare facility	 834 (85%) 87 (9%) 57 (6%)
Hospital admission characteristics	
APACHE III score; mean ± SD	92 ± 28

eTable 7. Correlation among unselected predictors

Variable 1	Variable 2	Spearman's rho
Lowest hematocrit	Highest hematocrit	0.85
Lowest white blood cell	Highest white blood cell	0.89
Lowest serum sodium	Highest serum sodium	0.89
Lowest serum creatinine	Highest serum creatinine	0.95
Lowest serum albumin	Highest serum albumin	0.87
How many drinks on a typical day when drinking?	How often do you have six or more drinks?	0.85
How often were you not able to stop drinking once started?	How often have you failed to do what was normally expected of you because of drinking?	0.75
How often were you not able to stop drinking once started?	How often do you need a drink in the morning?	0.75
How often were you not able to stop drinking once started?	How often have you been unable to remember the previous night because of drinking?	0.76
How often have you failed to do what was normally expected of you because of drinking?	How often have you been unable to remember the previous night because of drinking?	0.76
How often do you need a drink in the morning?	How often have you been unable to remember the previous night because of drinking?	0.78
How often to you feel guilt or remorse after drinking?	How often have you been unable to remember the previous night because of drinking?	0.80
PEEP at enrollment	Mean airway pressure	0.80
Calculated delivered tidal volume	Calculated delivered tidal volume after initial vent change	0.80
PEEP at enrollment	PEEP after initial vent change	0.86
Lowest heart rate	Heart rate (baseline)	0.76
Systolic blood pressure	Mean arterial pressure	0.80
Diastolic blood pressure	Mean arterial pressure	0.87
Lowest hematocrit	Hemoglobin (baseline)	0.80
Highest hematocrit	Hemoglobin (baseline)	0.81
Lowest serum sodium	Sodium (baseline)	0.89
Highest serum sodium	Sodium (baseline)	0.90
Lowest serum bicarbonate	Serum bicarbonate (baseline)	0.84
Lowest serum albumin	Albumin (baseline)	0.80
Heart rate (baseline)	Shock index	0.76
Maximum P:F ratio	Mean P:F ratio	0.86
Minimum P:F ratio	Mean P:F ratio	0.82
Maximum PaO_2	Mean PaO2	0.84
Minimum arterial pH	Mean arterial pH	0.89
Maximum $PaCO_2$	Mean $PaCO_2$	0.88
Minimum $PaCO_2$	Mean $PaCO_2$	0.84
Difference of max arterial pH and min arterial pH	Difference of max $PaCO_2$ and min $PaCO_2$	0.78
Maximum $PaCO_2$	Difference of max $PaCO_2$ and min $PaCO_2$	0.76
Difference of max P:F ratio and min P:F ratio	Difference of max PaO_2 and min PaO_2	0.76
Maximum PaO_2	Difference of max PaO_2 and min PaO_2	0.93

We excluded variables that were constituents of other variables (e.g., the individual AUDIT components vs. final score).
"Baseline" refers to the most recent values prior to randomization
"Lowest" and "Highest" refer to values from 24 hours preceding randomization
PEEP: Positive End Expiratory Pressure.

eTable 8. Univariate and multivariate adjusted linear regression, incorporating 4 acute illness variables.

Variable	Coefficient* (95% CI)	p value	R^2	Coefficient*,adj (95% CI)	p value^{adj}	Adjusted $R^{2.adj}$
Minimum P:F ratio	0.000 (-0.002, 0.001)	0.847	0.15	0.000 (-0.001, 0.001)	0.852	0.14
PEEP	0.007 (-0.014, 0.028)	0.495	0.15	0.002 (-0.018, 0.022)	0.822	0.14
Baseline Brussels Score	-0.008 (-0.097, 0.080)	0.852	0.15	-0.013 (-0.096, 0.071)	0.766	0.14
APACHE III Score	0.002 (-0.001, 0.005)	0.278	0.15	0.001 (-0.002, 0.004)	0.404	0.14

*Because of the Box Cox transformation, we inverted the sign of coefficients to aid in interpretation. Positive coefficients are associated with better outcome.
adj Adjusted for the predictors in our primary model
PEEP: Positive End Expiratory Pressure; APACHE: Acute Physiology and Chronic Health Evaluation

eTable 9. Regression model for post hoc analysis among 195 patients with "baseline" EQ5D health utility

Predictor	Primary model		Primary model + "baseline" EQ5D	
	Coefficient* (95% CI)	p value	Coefficient* (95% CI)	p value
Baseline characteristics				
Age	-0.009 (0.000, -0.019)	**0.047**	-0.006 (-0.014, 0.002)	0.14
Female Sex	-0.302 (-0.042, -0.562)	**0.023**	-0.154 (-0.387, 0.079)	0.19
Hispanic/Latino Ethnicity	-0.181 (-0.805, 0.442)	0.567	-0.147 (-0.698, 0.404)	0.6
Resides at home with informal help**	-0.301 (-1.011, 0.408)	0.403	0.112 (-0.525, 0.750)	0.73
Require professional help at home OR residence in health care facility**	-0.632 (-1.459, 0.195)	0.133	-0.690 (-1.421, 0.041)	0.06
AIDS comorbidity	1.499 (-0.331, 3.329)	0.108	1.038 (-0.584, 2.660)	0.21
Pulmonary comorbidity	-0.347 (-0.752, 0.058)	0.093	-0.391 (-0.033, -0.749)	**0.032**
Nadir respiratory rate	0.022 (0.002, 0.042)	**0.031**	0.020 (0.037, 0.002)	**0.031**
Lifestyle factors				
Current smoker status (at baseline)	-0.491 (-0.212, -0.771)	**0.001**	-0.250 (-0.505, 0.006)	0.06
Body mass index (at baseline)	-0.019 (-0.003, -0.035)	**0.02**	-0.019 (-0.005, -0.033)	**0.008**
Baseline EQ-5D	-	-	2.001 (1.457, 2.545)	**<0.001**
	R^2=0.19 Adjusted R^2=0.15		R^2=0.37 Adjusted R^2=0.33	

*Because of the Box Cox transformation, we inverted the sign of coefficients to aid in interpretation. Positive coefficients are associated with better outcome.
**Reference category for this variable is "Resides at home with no help"

APPENDIX 4: Eligibility criteria for EDEN, OMEGA, ALTA, and ALTOS

A. Eligibility criteria for EDEN and OMEGA

EDEN and OMEGA Inclusion Criteria

Patients were eligible for inclusion if they met all of the below criteria. Criteria 1-3 must have been present within a 24-hour time period:

Acute onset (defined below) of:

1. PaO_2/FiO_2 ≤300 or equivalent adjusted for altitude
2. Bilateral infiltrates consistent with pulmonary edema on frontal chest radiograph. The infiltrates could be patchy, diffuse, homogeneous, or asymmetric.
3. Requirement for positive pressure ventilation via endotracheal tube, and
4. No clinical evidence of left-sided cardiac failure to account for bilateral pulmonary infiltrates,

and

5. Intention of primary medical team to provide enteral nutrition to the patient

The 48-hour enrollment time window began when criteria 1-3 were met. If a patient met the first three inclusion criteria but had a PAOP (Pulmonary Arterial Wedge Pressure) >18 mmHg, then the first four criteria had to persist for >12 hours after the PAOP had declined to ≤18 mmHg, and still be within the 48-hour enrollment window.

"Acute onset" was defined as the duration of the hypoxemia criterion (#1) and the chest radiograph criterion (#2) must have been present for ≤ 28 days at the time of randomization. Opacities considered "consistent with pulmonary edema" included any opacities not fully explained by mass, atelectasis, or effusion or opacities known to be chronic (greater than 28 days). Vascular redistribution, indistinct vessels, and indistinct heart borders alone were not considered "consistent with pulmonary edema" and thus did not count as qualifying opacities for this study.

EDEN and OMEGA Exclusion Criteria

1. Age <13 years
2. Greater than 48 hours since all inclusion criteria met
3. Neuromuscular disease that impairs ability to ventilate without assistance, such as:
   a. cervical spinal cord injury at level C5 or higher
   b. amyotrophic lateral sclerosis
   c. Guillain-Barré Syndrome
   d. myasthenia gravis
   e. kyphoscoliosis or chest wall deformity resulting in severe exercise restriction, secondary polycythemia, or respirator dependence
4. Pregnant or breast-feeding
5. Severe chronic respiratory disease, demonstrated by any of:
   a. FEV1 <20 ml/kg PBW
   b. FEV1/FVC <50% predicted
   c. Chronic hypercapnea with $PaCO_2$ >45 mm Hg
   d. Chronic hypoxemia with PaO_2 <55 mm Hg on FiO_2 = 0.21
   e. Radiographic x-ray evidence of any chronic over-inflation or chronic interstitial infiltration
   f. Hospitalization within the past 6 months for respiratory failure
   g. Chronic restrictive, obstructive, neuromuscular, chest wall, or pulmonary vascular disease resulting in severe exercise restriction, secondary polycythemia, severe pulmonary hypertension with mean PAP >40 mm Hg, or respirator dependency
6. Burns greater than 40% total body surface area
7. Malignancy or other irreversible disease or condition for which 6-month mortality is estimated to be greater than 50%:
   a. Poorly controlled neoplasms
   b. Known HIV positive with known end stage process and known CD4 count <50/mm3
   c. Prior cardiac arrest requiring cardiopulmonary resuscitation without fully demonstrated neurologic recovery
   d. New York Heart Association Class IV exercise restriction
   e. Chronic condition making patient respirator dependent
8. Allogeneic bone marrow transplant in the last 5 years
9. Patient, surrogate, or physician not committed to full support (exception: a patient was not excluded if he/she would receive all supportive care except for attempts at resuscitation from cardiac arrest).
10. Severe chronic liver disease (Child-Pugh Score of 11-15)
11. Diffuse alveolar hemorrhage from vasculitis
12. Morbid obesity (a body weight of greater than 1 kg/cm of body height) [this represents a body mass index of approximately 55-65 kg/m^2]
13. No consent/inability to obtain consent
14. Unwillingness or inability to utilize the ARDS network 6 ml/kg PBW lung protective ventilation protocol
15. Moribund patient not expected to survive 24 hours
16. No intent to obtain central venous access for monitoring intravascular pressures
17. Greater than 72 hours since initiation of mechanical ventilation
18. Refractory shock, defined by any of the following:
   a. Dopamine infusion at rate >15 mcg/kg/min
   b. Dobutamine infusion at rate >15 mcg/kg/min
   c. Epinephrine or Norepinephrine infusion at rate >30 mcg/min
   d. Phenylephrine infusion at rate >50 mcg/min
   e. Milrinone infusion at rate >0.5 mcg/kg /min
   f. Vasopressin infusion at rate >0.04 U/min
   g. Intra-aortic Balloon Pump
19. Unable to obtain enteral access
20. Presence of partial or complete mechanical bowel obstruction, or ischemia, or infarction
21. Current TPN use or intent to use TPN within 7 days
22. Severe malnutrition with BMI <18.5 or loss of >30% total body weight in the previous 6 months
23. Laparotomy expected within 7 days
24. Unable to raise head of bed 30 degrees
25. Short-bowel syndrome or absence of gastrointestinal tract
26. Presence of high-output (>500 cc/day) enterocutaneous fistula
27. INR >5.0 or platelet count <30,000/mm3 or history of bleeding disorder
28. Intracranial hemorrhage within the previous month
29. Allergy to enteral formula, n-3 fatty acids, gamma-linolenic acid, vitamin E, vitamin C, betacarotene, taurine,

or L-carnitine
30. Requirement for, or physician insistence on, enteral formula supplemented with n-3 fatty acids (ex: Oxepa®, Impact®) or providing n-3 fatty acid, GLA, or anti-oxidant supplementation

From: Rice TW, Wheeler AP, Thompson BT, Steingrub J, Hite RD, Moss M et al. Initial trophic vs full enteral feeding in patients with acute lung injury: the EDEN randomized trial. JAMA 2012; 307(8):795-803.

B. ALTA Eligibility Criteria

ALTA Inclusion Criteria

1. PaO_2/FiO_2 ≤300 or equivalent adjusted for altitude where appropriate
2. Bilateral infiltrates consistent with pulmonary edema on frontal chest radiograph. The infiltrates could be patchy, diffuse, homogeneous, or asymmetric.
3. Requirement for positive pressure ventilation via endo tracheal tube, and
4. No clinical evidence of left-sided cardiac failure to ac count for bilateral pulmonary infiltrates

ALTA Exclusion Criteria

1. Age <13 years
2. Greater than 48 hours since all inclusion criteria met
3. Neuromuscular disease that impairs ability to ventilate without assistance, such as:
    a. cervical spinal cord injury at level C5 or higher
    b. amyotrophic lateral sclerosis
    c. Guillain-Barré Syndrome
    d. myasthenia gravis
    e. kyphoscoliosis or chest wall deformity resulting in severe exercise restriction, secondary polycythemia, or respirator dependence
    4. Pregnant or breast-feeding
5. Severe chronic respiratory disease, demonstrated by any of:
    a. FEV1 <20 ml/kg PBW
    b. FEV1/FVC <50% predicted
    c. Chronic hypercapnea with $PaCO_2$ >45 mm Hg
    d. Chronic hypoxemia with PaO_2 <55 mm Hg on FiO_2 = 0.21
    e. Radiographic x-ray evidence of any chronic over-inflation or chronic interstitial infiltration
    f. Hospitalization within the past 6 months for respiratory failure
    g. Chronic restrictive, obstructive, neuromuscular, chest wall, or pulmonary vascular disease resulting in severe exercise restriction, secondary polycythemia, severe pulmonary hypertension with mean PAP >40 mm Hg, or respirator dependency
6. Burns greater than 40% total body surface area
7. Malignancy or other irreversible disease or condition for which 6-month mortality is estimated to be greater than 50%:
    a. Poorly controlled neoplasms
    b. Known HIV positive with known end stage process and known CD4 count <50/mm3
    c. Prior cardiac arrest requiring cardiopulmonary resuscitation without fully demonstrated neurologic recovery
    d. New York Heart Association Class IV exercise restriction
    e. Chronic condition making patient respirator dependent
8. Allogeneic bone marrow transplant in the last 5 years
9. Patient, surrogate, or physician not committed to full support (exception: a patient was not excluded if he/she would receive all supportive care except for attempts at resuscitation from cardiac arrest).
10. Severe chronic liver disease (Child-Pugh Score of 11-15)
11. Diffuse alveolar hemorrhage from vasculitis
12. Morbid obesity (a body weight of greater than 1kg/cm of body height)
13. No consent/inability to obtain consent
14. Unwillingness or inability to utilize the ARDS network 6 ml/kg PBW lung protective ventilation protocol
15. Moribund patient not expected to survive 24 hours.
16. No intent to obtain central venous access for monitoring intravascular pressures
17. Contraindication to aerosolized albuterol.
18. Daily use (prior to study hospitalization) of inhaled beta agonist, corticosteroid, or oral leukotriene modifier for reactive airway disease.
19. Unwillingness of primary physician to discontinue inpatient beta agonist use.
20. Acute myocardial infarction or acute coronary syndrome within 30 days.
21. Severe congestive heart failure.
22. Participation in other experimental medication trial with in 30 days with the exception of EDEN/OMEGA.
23. Heart rate greater than 85% of maximal predicted heart rate (MGR85) as calculated by MHR85 = 85% x (220-age) or 140 beats per minute (whichever is lower).
24. Patients receiving high frequency ventilation.
25. Atrial fibrillation (new since hospital admission) requiring anticoagulation.
26. Greater than 5 PVCs per minute in the four hour period prior to randomization.

From: National Heart Lung and Blood Institute Acute Respiratory Distress Syndrome Clinical Trials Network, Matthay MA, Brower RG, Carson S, Douglas IS, Eisner M, Hite D, Holets S, Kallet RH, Liu KD, MacIntyre N, Moss M, Schoenfeld D, Steingrub J, Thompson BT. Randomized, placebo-controlled clinical trial of an aerosolized beta(2)-agonist for treatment of acute lung injury. Am J Respir Crit Care Med 2011; 184: 561-568.

C. ALTOS Eligibility Criteria

Patients were excluded from evaluation of post-discharge outcomes in ALTOS if (a) age <18 years old, (b) did not speak English, (c) had no fixed address (i.e., homeless), (d) had baseline cognitive impairment (based on chart review and/or interview with the patient and/or surrogate when necessary), or (e) failed to give consent for long-term follow-up.

From: Needham DM, Dinglas VD, Bienvenu OJ, Colantuoni E, Wozniak AW, Rice TW, Hopkins RO, Network NNA. One year outcomes in patients with acute lung injury randomised to initial trophic or full enteral feeding: prospective follow-up of EDEN randomised trial. BMJ 2013; 346: f1532.

REFERENCES

1. Amato MB, Meade MO, Slutsky AS, Brochard L, Costa EL, Schoenfeld DA, Stewart TE, Briel M, Talmor D, Mercat A, Richard JC, Carvalho CR, Brower RG. Driving pressure and survival in the acute respiratory distress syndrome. N Engl J Med 2015; 372: 747-755.
2. van der Heijden GJ, Donders AR, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. J Clin Epidemiol 2006; 59: 1102-1109.
3. Friedman JH. Multivariate adaptive regression splines. The annals of statistics 1991: 1-67.
4. Friedman JH, Roosen CB. An introduction to multivariate adaptive regression splines. Statistical Methods in Medical Research 1995; 4: 197-217.
5. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B (Methodological) 1996; 58: 267-288.
6. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2011; 73: 273-282.
7. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York: Springer; 2001.
8. Moons KG, Donders AR, Steyerberg EW, Harrell FE. Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: a clinical example. J Clin Epidemiol 2004; 57: 1262-1270.
9. Flom PL, Cassell DL. Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use. NorthEast SAS Users Group Inc 20th Annual Conference: 11-14th November 2007; Baltimore, Maryland; 2007.
10. Freedman DA. A Note on Screening Regression Equations. The American Statistician 1983; 37: 152-155.
11. Ellison AM, Gotelli NJ, Inouye BD, Strong DR. P values, hypothesis testing, and model selection: it's deja vu all over again. Ecology 2014; 95: 609-610.
12. Pena EA, Slate EH. Global Validation of Linear Model Assumptions. J Am Stat Assoc 2006; 101: 341.
13. Naglie G, Krahn MD, Naimark D, Redelmeier DA, Detsky AS. Primer on medical decision analysis: Part 3--Estimating probabilities and utilities. Med Decis Making 1997; 17: 136-141.
14. Pollack MM, Holubkov R, Funai T, Berger JT, Clark AE, Meert K, Berg RA, Carcillo J, Wessel DL, Moler F, Dalton H, Newth CJ, Shanley T, Harrison RE, Doctor A, Jenkins TL, Tamburro R, Dean JM, Eunice Kennedy Shriver National Institute of Child H, Human Development Collaborative Pediatric Critical Care Research N. Simultaneous Prediction of New Morbidity, Mortality, and Survival Without New Morbidity From Pediatric Intensive Care: A New Paradigm for Outcomes Assessment. Crit Care Med 2015; 43: 1699-1709.
15. Lachin JM. Worst-rank score analysis with informatively missing observations in clinical trials. Control Clin Trials 1999; 20: 408-422.
16. Chen YH, Gould AL, Nessly ML. Treatment comparisons for a partially categorical outcome applied to a biomarker with assay limit. Stat Med 2005; 24: 211-228.