ERJ Express. Published on February 6, 2019 as doi: 10.1183/13993003.02237-2018

EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original article

Development of a Prognostic Model of Respiratory Insufficiency or Death in Amyotrophic Lateral Sclerosis

Jason Ackrivo, John Hansen-Flaschen, E. Paul Wileyto, Richard J. Schwab, Lauren Elman, Steven M. Kawut

Please cite this article as: Ackrivo J, Hansen-Flaschen J, Wileyto EP, *et al*. Development of a Prognostic Model of Respiratory Insufficiency or Death in Amyotrophic Lateral Sclerosis. *Eur Respir J* 2019; in press (https://doi.org/10.1183/13993003.02237-2018).

Development of a Prognostic Model of Respiratory Insufficiency or Death in Amyotrophic Lateral Sclerosis

Authors

Jason Ackrivo, MD, MSCE[1]; John Hansen-Flaschen, MD[1]; E. Paul Wileyto, PhD[2]; Richard J. Schwab, MD[1]; Lauren Elman, MD*[3]; Steven M. Kawut, MD, MS*[1,2]

*Equal contribution

Authors' affiliations

[1]Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[2]Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[3]Department of Neurology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

The authors have no conflicts of interest relevant to this article.

Sources of support:

Grant funds provided by National Institutes of Health T32 HL-007891, K24 HL-103844, P01 AG032953-A1, F32 HL144145

Author for correspondence:

Jason Ackrivo, MD, MSCE

Pulmonary Allergy and Critical Care Division

839 West Gates Building

3400 Spruce Street

Philadelphia, PA 19104

Telephone: +1.215.615.3718

Fax: +1.215.614.0869

E-mail: jason.ackrivo@uphs.upenn.edu.

This article has an online data supplement.

“Take home” message: ALS has a heterogenous progression to respiratory failure. A clinician can use our clinical prognostic rule to estimate a six-month risk of respiratory failure onset, thus facilitating referrals and respiratory interventions.

Abstract

A clinically useful model to prognose onset of respiratory insufficiency in amyotrophic lateral sclerosis (ALS) would inform disease interventions, communication, and clinical trial design. We aimed to derive and validate a clinical prognostic model for respiratory insufficiency within six months of presentation to an outpatient ALS clinic.

We used multivariable logistic regression and internal cross-validation to derive a clinical prognostic model using a single-center cohort of 765 ALS patients who presented between 2006 and 2015. External validation was performed using the multicenter Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) database with 7,083 ALS patients. Predictors included baseline characteristics at first outpatient visit. The primary outcome was respiratory insufficiency within six months, defined by initiation of non-invasive ventilation, forced vital capacity < 50% predicted, tracheostomy, or death.

Of 765 patients in our center, 300 (39%) had respiratory insufficiency or death within six months. Six baseline characteristics (diagnosis age; delay between symptom onset and diagnosis; forced vital capacity; symptom onset site; ALS Functional Rating Scale-Revised (ALSFRS-R) total score; and ALSFRS-R dyspnea score) were used to prognose the risk of the primary outcome. The derivation cohort c-statistic was 0.86 (95% confidence interval (CI), 0.84 – 0.89). Internal cross validation produced a c-statistic of 0.86 (95% CI, 0.85 – 0.87). External validation of the model using the PROACT cohort produced a c-statistic of 0.74 (95% CI, 0.72 – 0.75).

We derived and externally validated a clinical prognostic rule for respiratory insufficiency in ALS. Future studies should investigate interventions on equivalent high-risk patients.

Keywords: amyotrophic lateral sclerosis, respiratory failure, prediction modeling, prognosis, non-invasive ventilation

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease with high morbidity and universal mortality. ALS is usually sporadic with an incidence of approximately 1 to 2/100,000 people in the US per year with a prevalence of

approximately 3-5/100,000.[1] Impairment of key respiratory muscles including the diaphragm, accessory muscles of respiration, and bulbar muscles leads to death through mechanisms of aspiration, diminished airway clearance due to ineffective cough, recurrent pulmonary infections, and chronic hypercapnic respiratory failure.[2-5] The cornerstone of respiratory care in ALS involves non-invasive ventilation (NIV), which has been shown to improve quality of life and potentially survival.[6] Despite the key role of respiratory failure in the morbidity and mortality associated with ALS, there remains uncertainty concerning the optimal timing of initiating respiratory care for this disease.[7]

Shortcomings in current clinical strategies for predicting the onset of respiratory insufficiency have hindered development of practice guidelines and clinical trials. Also, the absence of a reliable prognostic model has prevented clinicians from anticipating mechanically-assisted ventilation, thereby limiting the opportunity to prepare patients for shared decision making, improving timeliness of referrals for respiratory interventions, and developing clinical trial design. Accordingly, the aim of this study was to develop and validate a prognostic model for the onset of respiratory insufficiency or death within six months of presentation to an ALS center.

Methods

Study Design and Population

For derivation and internal validation of the prognostic model we performed a retrospective cohort study of patients at the University of Pennsylvania Comprehensive

ALS Center (Philadelphia, Pennsylvania) with first visit between January 1, 2006 and December 31, 2015.

The source population for the validation cohort was the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) database. Details on the PRO-ACT dataset are found in the online data supplement.

This study was approved by the University of Pennsylvania institutional review board and adheres to TRIPOD guidelines for transparent reporting of prediction models.[8]

Study Samples and Data Collection

The Penn cohort data were prospectively entered into a secure online data portal with follow-up until September 1, 2016. Starting in 2006, patients diagnosed with ALS by an attending neurologist using the World Federation of Neurology El Escorial Criteria were approached for consent.[9]

We excluded patients with unusable data or non-physiologic values, NIV use before diagnosis, tracheostomy before diagnosis, tracheostomy before NIV, and baseline FVC < 50%. We also excluded anyone with less than six months of follow-up time who were censored as “alive”. See the online supplement for further details regarding data collection for the Penn dataset.

The PRO-ACT database includes de-identified data from 23 phase II/III clinical trials (see online supplement). PRO-ACT inclusion criteria commonly included being 18 years of age or older; ability to provide informed consent; clinical diagnosis of ALS; FVC ≥ 50% predicted; serum creatinine < 1.5 mg/dl (133 μmol/L) and disease duration of ≤ 5 years from symptom onset. Exclusion criteria for trials from PRO-ACT included recent

exposure to the study drug; exposure to other investigational agents within the last 30 days; malnourishment; substance abuse within the last year; and active significant medical or psychiatric disease.

See online supplement for ALSFRS-R score.

Outcomes

We sought to establish a discriminating clinical prognostic rule for “respiratory insufficiency” within six months of outpatient clinic presentation. Respiratory insufficiency included any one of the following outcomes: initiation of NIV, FVC < 50% of predicted, tracheostomy placement, or death. We performed sensitivity analyses as discussed in the online supplement.

Statistical Analysis

We created a prognostic model with multivariable logistic regression to identify associations between baseline characteristics and respiratory insufficiency within six months. After deriving the model on the Penn cohort and internally cross-validating, we externally validated with the PRO-ACT cohort. We graphically represented our model performance using receiver operating characteristic curves, Kaplan-Meier curves and calibration plots. See online supplement for further details.

Results

Penn cohort

One-thousand sixty-one patients with ALS were evaluated at Penn during the study period. After excluding those with NIV or tracheostomy at baseline (n=33), baseline FVC < 50% (n=168), unusable data or non-physiologic values (n=64), or those with less than 6 months follow-up (n=31), there were 765 individuals in the Penn study sample. Mean age at diagnosis was 63 years, 58% were male, and 87% were Caucasian (**Table 1**). The average FVC at baseline was 81%, and mean baseline ALSFRS-R total score was 37. Fifty-three percent of the cohort classified themselves as “never” smokers. Median follow-up time was 2.3 years (interquartile range, 1.5-4.0 years). Thirty-nine percent (n=300) of the cohort had respiratory insufficiency (or death) by six months of observation.

We compared baseline characteristics of patients who had respiratory insufficiency by six months to those who did not (**Table 2**). Patients reaching the composite outcome had significantly older diagnosis age; shorter diagnosis delay; higher proportion of Definite ALS El Escorial criteria; higher proportion of bulbar symptom onset; lower FVC, lower ALSFRS-R total score; lower ALSFRS-R dyspnea and orthopnea scores; more coronary artery disease; and greater chance of diabetes mellitus at baseline compared to those who did not reach the composite outcome. There were no significant differences in sex or race.

In the Penn cohort, 167 (22%) initiated NIV; 246 (32%) had FVC < 50% predicted; two (0.3%) underwent tracheostomy; and 88 (12%) died by six months.

Prognosis of Respiratory Insufficiency or Death

Univariate analysis predictors are depicted in **Table 3**. After purposeful backward selection in a multivariable analysis, six predictors were retained, including age at diagnosis, diagnosis delay, symptom onset site, FVC, ALSFRS-R total score, and ALSFRS-R dyspnea score. The model was well-calibrated by both Hosmer-Lemeshow test (p=0.45, **Table E1** in the online data supplement) and standardized Pearson χ^2 test (p=0.31). The ROC curve produced from the multivariable model had a c-statistic of 0.86 (95% CI 0.84 – 0.89, **Figure 1A**). To perform internal validation, we performed a 4-fold (4 to 1) cross validation of the multivariate logistic model in the Penn cohort for the composite outcome. The ROC from internal cross validation on out-of-sample data produced a c-statistic which was almost identical to the full derivation cohort (0.86, 95% CI 0.85 – 0.87, **Figure E1**). Using the internally cross-validated ROC curve, we selected a cut-point of 0.45 for the odds of reaching respiratory insufficiency (including death) at six months, corresponding to a sensitivity of 83% and specificity of 81% (**Table 4**). The underlying risk of having respiratory insufficiency (including death) in the Penn cohort was 39%, so a "positive" prediction (odds of an event ≥ 0.45 from the model) nearly doubled the risk of reaching this end point (positive predictive value (PPV) = 77%), whereas a "negative" prediction reduced the chances of having an outcome to only 14% [negative predictive value (NPV) = 86%). In the 616-person internal derivation cohort, 239 patients had a "positive" test, identifying one third of the population at high risk of respiratory insufficiency within six months.

Patients with a positive-predicted odds of respiratory insufficiency had significantly higher risk of the primary outcome within six months (**Figure 1B**) and at one year

(**Figure E2**, both p<0.001). The median time to respiratory insufficiency for the group with a positive prediction (odds of an event ≥ 0.45) was 123 days (interquartile range, 77 – 239 days), and for those with a negative prediction was 469 days (interquartile range, 259 – 889 days).

A calculator for prognosticating the risk of respiratory insufficiency at six months is included in the online supplementary material.

Mortality

We also used the multivariate logistic model in the Penn cohort to prognose death alone. The ROC produced from the multivariable model had a c-statistic of 0.84 (95% CI 0.80 – 0.89, **Figure E3**) for death. Four-fold internal cross-validation produced a c-statistic of 0.83 (95% CI 0.82 – 0.84, **Figure E4**). For internal cross-validation, a cut-off of 0.11 had sensitivity 83% and specificity 81%, similar to the respiratory insufficiency model (**Table 4**).

PRO-ACT

The PRO-ACT cohort contained 10,723 subjects. We excluded 2,473 patients with less than six months of follow-up time and those with prior tracheostomy or prior NIV (n=108), as well as anyone with FVC < 50 at start of observation (n=1,059). The final PRO-ACT cohort contained 7,083 individuals. The mean age was 56 years, 62% were male, and 96% were Caucasian (**Table 5**). The average FVC at baseline was 88%, and mean baseline ALSFRS-R total score was 37. The median follow-up time was 0.98

years (interquartile range, 0.65-1.32 years). Thirty-five percent (n=2453) had respiratory insufficiency (including death) by six months of observation.

We compared baseline characteristics of individuals who did and did not meet our composite outcome within six months of diagnosis (**Table E2**). Subjects reaching the composite endpoint were significantly older, were less likely to be male, had a shorter diagnosis delay, had bulbar-onset symptoms, had lower FVC, and had lower ALSFRS-R total score compared to those who did not reach the composite outcome. There was no significant difference in race.

At six months, 2453 (35%) individuals had respiratory insufficiency; 360 (5%) were initiated on NIV; 1398 (20%) had an FVC less than 50% predicted; and 1,168 (16%) died. No one received a tracheostomy.

Prognosis of Respiratory Insufficiency or Death

We applied the prognostic model and cutoff from the Penn cohort to PRO-ACT. The model yielded a c-statistic of 0.74 (95% CI, 0.72 – 0.75) (**Figure 2A**). Table 4 shows the model performance using the outcome probability cut-off of ≥ 0.45, which produced a sensitivity of 53% and specificity of 82%. The Hosmer-Lemeshow goodness-of-fit test was ($p<0.001$) (**Table E3**) and standardized Pearson χ^2 test was ($p=0.001$). The calibration plot (**Figure E5**) illustrates excellent precision estimates, with somewhat higher than expected events in the lower-risk groups and lower than expected events in the higher-risk groups.

Patients with a high probability of respiratory insufficiency had an increased risk of respiratory insufficiency at six months ($p<0.001$, **Figure 2B**). Median time to respiratory

insufficiency for the group with a positive prediction (odds of an event ≥ 0.45) was 182 days (interquartile range, 91 – 344 days), and for those with a negative prediction was 381 days (interquartile range, 204 – 581 days).

Mortality

Applying the clinical prognostic rule to death in PRO-ACT produced a c-statistic of 0.72 (95% CI, 0.71 – 0.74) (**Figure E6**). Applying a cut-point of 0.11 produced a sensitivity of 47% (95% CI, 44 – 50%), specificity of 82% (95% CI, 81 – 83%), PPV of 36% (95% CI, 34 – 38%), and NPV of 88% (95% CI, 87 – 89%) (**Table 4**).

Sensitivity Analyses

See online supplement.

Discussion

We found that younger age, less diagnostic delay, lower FVC, bulbar symptom onset site, lower ALSFRS-R total, and ALSFRS-R dyspnea ≤ 2 at baseline were significantly associated with a higher risk of respiratory insufficiency or death at six months in a large single-center cohort and a dataset of multiple clinical trials in ALS. The model had high sensitivity, specificity, PPV and NPV in the derivation cohort and maintained high specificity, PPV, and NPV in the validation cohort.

Other studies have developed prognostic models of ALS disease progression, using methods such as longitudinal support vector regression, random forest algorithms, and machine learning. These studies have found baseline ALSFRS score, ALSFRS slope,

symptom onset site, executive dysfunction and diagnosis delay time to be significantly associated with overall survival.[10-14] However, these approaches are computationally intensive, require variables not available in typical clinical practice, necessitate repeat assessments over time, or focus on overall disease progression rather than respiratory events.

A recent study created a prognostic model for time from symptom onset to a composite end point of tracheostomy, dependence on non-invasive ventilation (>23 hours per day), or death in ALS.[15] However, our study differed in several important ways. Our model included variables which 1) are routinely clinically available on all patients, 2) are assessed at baseline, and 3) prognose the short-term risk of respiratory failure onset, which is an important clinical event for patients and could be used to create an “enriched” study population for clinical trials. Sensitivity, specificity, PPV, and NPV were not presented in the Westeneng study, making it more difficult to apply the findings at the bedside for an individual patient. Perhaps most importantly, our clinical rule focuses on early stages of respiratory insufficiency, thus facilitating referral to respiratory physicians for timely interventions.

Prior literature on prognostic factors for ALS have found a significant association between age, bulbar onset disease, and diagnosis delay with worsened survival, consistent our study.[16-19] Kimura and colleagues found that change in ALSFRS-R score and symptom duration at diagnosis (analogous to diagnosis delay in the current study) identified two groups with distinct survival.[12] Crowdsourcing initiatives have used advanced machine-learning algorithms to identify 16 predictors (including time from symptom onset, FVC, age, site of symptom onset, and ALSFRS-R total score) for

distinguishing between relatively “fast” versus “slow” disease progressors by change in ALSFRS score.[13] It was estimated that with such information clinical trial enrollment sample size could be reduced by 20%.

Strengths and Limitations

Our study has several strengths. Most notably, we leveraged two large databases of prospectively collected data from ALS patients. To our knowledge, this is one of the largest studies to date for prognosing respiratory outcomes in ALS. In addition, our validation of the model in a separate, multicenter, international patient population (with different inclusion from the derivation cohort) attests to its generalizability.

We recognize several limitations to our study. The c-statistic, sensitivity, and NPV of the Penn cohort model decreased in the PRO-ACT cohort. While the prognostic rule was quite discriminating and well-calibrated (at least in the higher-risk group) in the PRO-ACT validation cohort, the cutoff selected from the Penn cohort was less sensitive. There are several possible reasons for this. First, the Penn cohort included most patients evaluated at our ALS center over 10 years, potentially making the prognostic rule generalizable to other centers, whereas the PRO-ACT cohort is composed of selected patients from clinical trials with multiple inclusion and exclusion criteria. Therefore, the two cohorts may have important differences, even if many of the demographics appeared similar.

Heterogeneity in PRO-ACT study design and variability in end point assessment could lead to bias, but this would likely be independent of our predictors. Thus, any bias

introduced by study heterogeneity would likely bias towards the null. Our strong model performance despite a very heterogenous dataset attests to its generalizability.

While six-month risk of respiratory insufficiency was essentially identical between the two cohorts (~35-39%), the incidence of the individual outcomes was somewhat different (e.g., NIV initiation was 22% in Penn cohort vs 5% in PRO-ACT). Given that PRO-ACT is a multicenter clinical trial database from cohorts around the world, our results may be affected by substantial practice variation regarding interventions for respiratory insufficiency. In the United States, guidelines and most medical insurance carriers recognize an FVC < 50% predicted normal as a threshold for initiation of NIV. However, some centers in our study may initiate NIV at a higher FVC (70-80%) while others prefer to wait for alternate physiologic changes (e.g., impaired gas exchange) even though the FVC is below 50% predicted normal. In addition, the threshold for tracheostomy placement likely differs between centers and regions.

Measurement error could have affected FVC. However, the Penn cohort used a spirometer in a neurology clinic after clinical trial training of a nurse practitioner and nurse. In addition, PRO-ACT is comprised of randomized controlled trial (RCT) data, which obtained FVC values using rigorous clinical research-grade methodologies. Of course, measurement error would likely bias to the null, unless related to both the actual value of the FVC itself and the risk of the outcome.

FVC is often used as a criterion for NIV initiation, and so our model may be influenced by using FVC as a predictor. However, we found that our model performs significantly better over using FVC alone (data not shown, $p<0.001$).

Performing a logistic regression rather than a time-to-event analysis has potential for introducing selection bias. There are several reasons why we chose the former over the latter. First, time-to-event analyses assume non-informative censoring, which likely would have been violated by including those with short follow-up times. A very large external validation cohort of 7,083 individuals likely mitigates any selection bias in our study. Second, we felt that a risk associated with a fixed time point (as in a logistic regression) is more practical for clinicians and patients rather than an arbitrary hazard ratio associated with no time point (as in a time-to-event analysis). Lastly, we felt that the six month time window was appropriate for guiding immediate interventions and discriminating “high-risk“ versus “low-risk” individuals.

Patients may have entered the Penn cohort after receiving a prior diagnosis of ALS; however, we accounted for this by using initial clinic presentation as time zero in our model. This approach facilitated validation in the PRO-ACT cohort, which began collecting data at the start of clinical trial enrollment, rather than at diagnosis date. Unmeasured confounding is possible; however, this would not impact the ability to predict events using this prognostic rule. Both the Hosmer-Lemeshow and standardized Pearson χ^2 tests indicated good calibration in the Penn cohort but less so in the PRO-ACT cohort. However, these tests are limited since they are non-specific for model fit, non-significant p-values do not indicate direction of calibration, and large sample sizes make it difficult to find a parsimonious model with a p-value above 0.05.[20] We feel that the large sample size enabled high precision estimates and thus low p-values.

Conclusions

A prognostic model for respiratory insufficiency or death in ALS may allow future studies to: (1) examine the impact of the “high-risk” phenotypes on important outcomes, (2) study novel mechanisms of disease, (3) develop an early intervention on a “high-risk” phenotype for respiratory insufficiency, and (4) identify characteristics associated with different trajectories of respiratory function, thus allowing for personalized medicine. Further study would be necessary to validate this as a tool prospectively to identify a high-risk subgroup suitable for clinical trial enrollment. In clinical practice, application of the prognostic model may help inform the optimal timing for referral of ALS clinical patients for respiratory care with the goal of delaying (or at least preparing for) the onset of respiratory insufficiency.

References Cited

1. Brown RH, Al-Chalabi A. Amyotrophic Lateral Sclerosis. N Engl J Med 2017; 377: 162–172.

2. Tsai C-P, Chang B-H, Lee CT-C. Underlying cause and place of death among patients with amyotrophic lateral sclerosis in Taiwan: a population-based study, 2003-2008. J Epidemiol 2013; 23: 424–428.

3. Paulukonis ST, Roberts EM, Valle JP, Collins NN, English PB, Kaye WE. Survival and Cause of Death among a Cohort of Confirmed Amyotrophic Lateral Sclerosis Cases. PLoS ONE 2015; 10: e0131965.

4. Yeo L, Lynch C, Hardiman O. Validating population-based registers for ALS: how accurate is death certification? Journal of Neurology 2010; 257: 1235–1239.

5. Spataro R, Re Lo M, Piccoli T, Piccoli F, La Bella V. Causes and place of death in Italian patients with amyotrophic lateral sclerosis. Acta Neurol. Scand. 2010; 122: 217–223.

6. Bourke SC, Tomlinson M, Williams TL, Bullock RE, Shaw PJ, Gibson GJ. Effects of non-invasive ventilation on survival and quality of life in patients with amyotrophic lateral sclerosis: a randomised controlled trial. The Lancet Neurology 2006; 5: 140–147.

7. Miller RG, Jackson CE, Kasarskis EJ, England JD, Forshew D, Johnston W, Kalra S, Katz JS, Mitsumoto H, Rosenfeld J, Shoesmith C, Strong MJ, Woolley SC, Quality Standards Subcommittee of the American Academy of Neurology. Practice parameter update: the care of the patient with amyotrophic lateral sclerosis: drug, nutritional, and respiratory therapies (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology; 2009; 73: 1218–1226.

8. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann. Intern. Med. American College of Physicians; 2015; 162: 55–63.

9. Brooks BR, Miller RG, Swash M, Munsat TL, World Federation of Neurology Research Group on Motor Neuron Diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Other Motor Neuron Disord. 2000; 1: 293–299.

10. Du W, Cheung H, Goldberg I, Thambisetty M, Becker K, Johnson CA. A Longitudinal Support Vector Regression for Prediction of ALS Score. IEEE Int Conf Bioinform Biomed Workshops IEEE; 2015; 2015: 1586–1590.

11. Hothorn T, Jung HH. RandomForest4Life: a Random Forest for predicting ALS disease progression. Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration 2014; 15: 444–452.

12. Kimura F, Fujimura C, Ishida S, Nakajima H, Furutama D, Uehara H, Shinoda K, Sugino M, Hanafusa T. Progression rate of ALSFRS-R at time of diagnosis predicts survival

time in ALS. Neurology Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology; 2006; 66: 265–267.

13. Küffner R, Zach N, Norel R, Hawe J, Schoenfeld D, Wang L, Li G, Fang L, Mackey L, Hardiman O, Cudkowicz M, Sherman A, Ertaylan G, Grosse-Wentrup M, Hothorn T, van Ligtenberg J, Macke JH, Meyer T, Schölkopf B, Tran L, Vaughan R, Stolovitzky G, Leitner ML. Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. Nat. Biotechnol. Nature Publishing Group; 2015; 33: 51–57.

14. Elamin M, Bede P, Montuschi A, Pender N, Chiò A, Hardiman O. Predicting prognosis in amyotrophic lateral sclerosis: a simple algorithm. Journal of Neurology Springer Berlin Heidelberg; 2015; 262: 1447–1454.

15. Westeneng H-J, Debray TPA, Visser AE, van Eijk RPA, Rooney JPK, Calvo A, Martin S, McDermott CJ, Thompson AG, Pinto S, Kobeleva X, Rosenbohm A, Stubendorff B, Sommer H, Middelkoop BM, Dekker AM, van Vugt JJFA, van Rheenen W, Vajda A, Heverin M, Kazoka M, Hollinger H, Gromicho M, Körner S, Ringer TM, Rödiger A, Gunkel A, Shaw CE, Bredenoord AL, van Es MA, et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. The Lancet Neurology 2018; 17: 423–433.

16. Czaplinski A, Yen AA, Appel SH. Amyotrophic lateral sclerosis: early predictors of prolonged survival. Journal of Neurology 2006; 253: 1428–1436.

17. Chiò A, Logroscino G, Hardiman O, Swingler R, Mitchell D, Beghi E, Traynor BG, On Behalf of the Eurals Consortium. Prognostic factors in ALS: A critical review. Amyotrophic Lateral Sclerosis 2009; 10: 310–323.

18. del Aguila MA, Longstreth WT, McGuire V, Koepsell TD, van Belle G. Prognosis in amyotrophic lateral sclerosis: a population-based study. Neurology 2003; 60: 813–819.

19. Chio A, Mora G, Leone M, Mazzini L, Cocito D, Giordana MT, Bottacchi E, Mutani R, Piemonte and Valle d'Aosta Register for ALS (PARALS). Early symptom progression rate is related to ALS outcome: a prospective population-based study. Neurology 2002; 59: 99–103.

20. Allison P. Measures of fit for logistic regression. https://statisticalhorizons.com/wp-content/uploads/GOFForLogisticRegression-Paper.pdf. Date last updated: 2014. Date last accessed: September 1 2018.

21. Pinto S, de Carvalho M. The R of ALSFRS-R: does it really mirror functional respiratory involvement in amyotrophic lateral sclerosis? Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration 2015; 16: 120–123.

22. Gordon PH, Miller RG, Moore DH. ALSFRS-R. Amyotroph. Lateral Scler. Other Motor Neuron Disord. 2004; 5 Suppl 1: 90–93.

23. Gordon PH, Cheung YK. Progression rate of ALSFRS-R at time of diagnosis predicts survival time in ALS. Neurology 2006; 67: 1314–5.

24. Rooney J, Burke T, Vajda A, Heverin M, Hardiman O. What does the ALSFRS-R really measure? A longitudinal and survival analysis of functional dimension subscores in amyotrophic lateral sclerosis. J. Neurol. Neurosurg. Psychiatr. 2017; 88: 381–385.

25. Bakker LA, Schröder CD, van Es MA, Westers P, Visser-Meily JMA, van den Berg LH. Assessment of the factorial validity and reliability of the ALSFRS-R: a revision of its measurement model. Journal of Neurology 2017; 264: 1413–1420.

26. Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, Nakanishi A. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). Journal of the Neurological Sciences 1999; 169: 13–21.

27. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. Source Code Biol Med 2008; 3: 17.

28. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Statist. Surv. 2010; 4: 40–79.

29. N Dudley PhD W, Wickham PhD RN AOCN R, Coombs MS N. An Introduction to Survival Statistics: Kaplan-Meier Analysis. JADPRO 2016; 7.

30. Demler OV, Paynter NP, Cook NR. Tests of calibration and goodness-of-fit in the survival setting. Stat Med 2015; 34: 1659–1680.

31. McCullagh P. The Conditional Distribution of Goodness-of-Fit Statistics for Discrete Data. Journal of the American Statistical Association 1986; 81: 104.

32. Enders CK. Multiple imputation as a flexible tool for missing data handling in clinical research. Behav Res Ther 2017; 98: 4–18.

33. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int. J. Methods Psychiatr. Res. 2011; 20: 40–49.

Figure Legends

Figure 1A. Receiver operating characteristic curve for prognosis of respiratory insufficiency (including death) within six months in the Penn cohort.

Figure 1B. Kaplan-Meier curves with 95% confidence intervals stratified by prognostic probability of respiratory insufficiency in the Penn cohort, truncated at 180 days.

Figure 2A. Receiver operating characteristic curves for derivation and external validation of respiratory insufficiency within six months. Test of equality, $p<0.001$.

Figure 2B. Kaplan-Meier curves with 95% confidence intervals stratified by prognostic probability of respiratory insufficiency in the PRO-ACT cohort, truncated at 180 days.

Figure 1A.

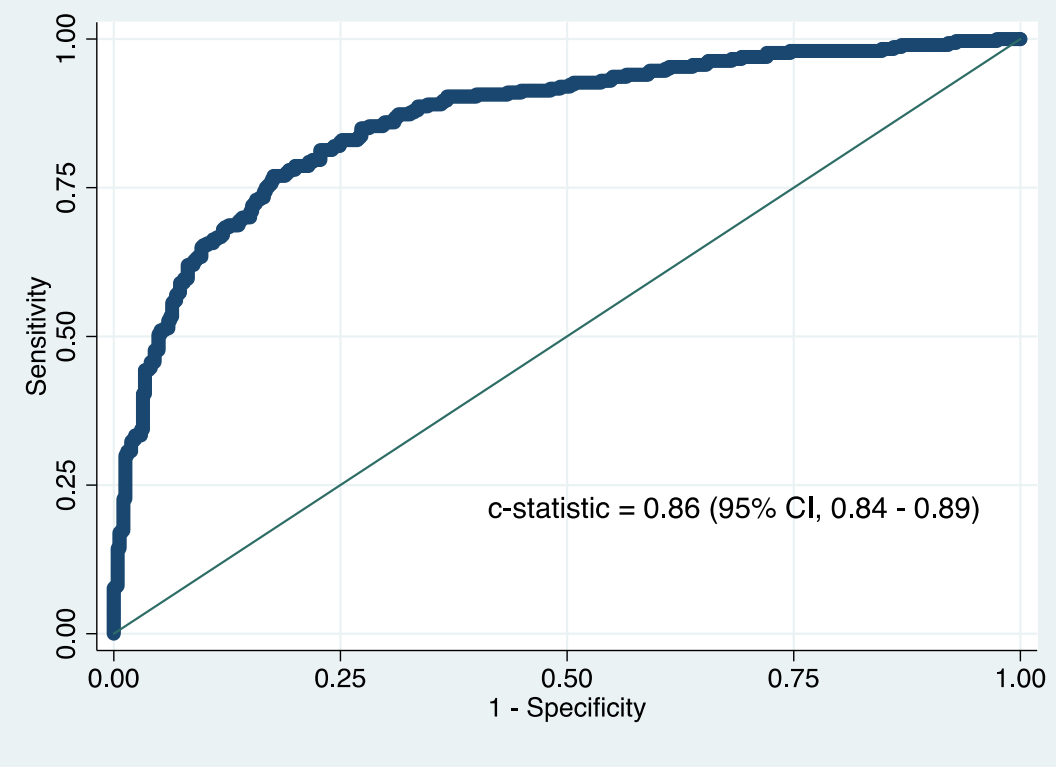

Sensitivity
0.00
0.25
0.50
0.75
1.00
1 - Specificity
0.00
0.25
0.50
0.75
1.00
c-statistic = 0.86 (95% CI, 0.84 - 0.89)

Figure 1B.

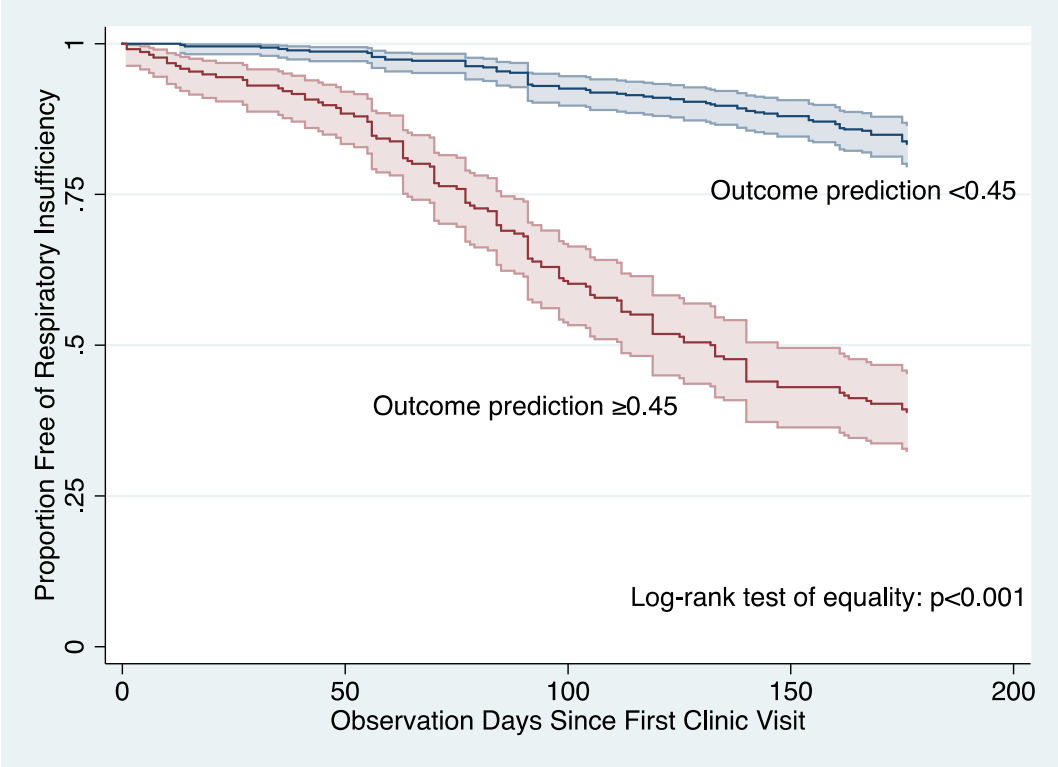

Proportion Free of Respiratory Insufficiency
0
.25
.5
.75
1
Outcome prediction <0.45
Outcome prediction ≥0.45
Log-rank test of equality: p<0.001
0
50
100
150
200
Observation Days Since First Clinic Visit

Figure 2A and 2B.

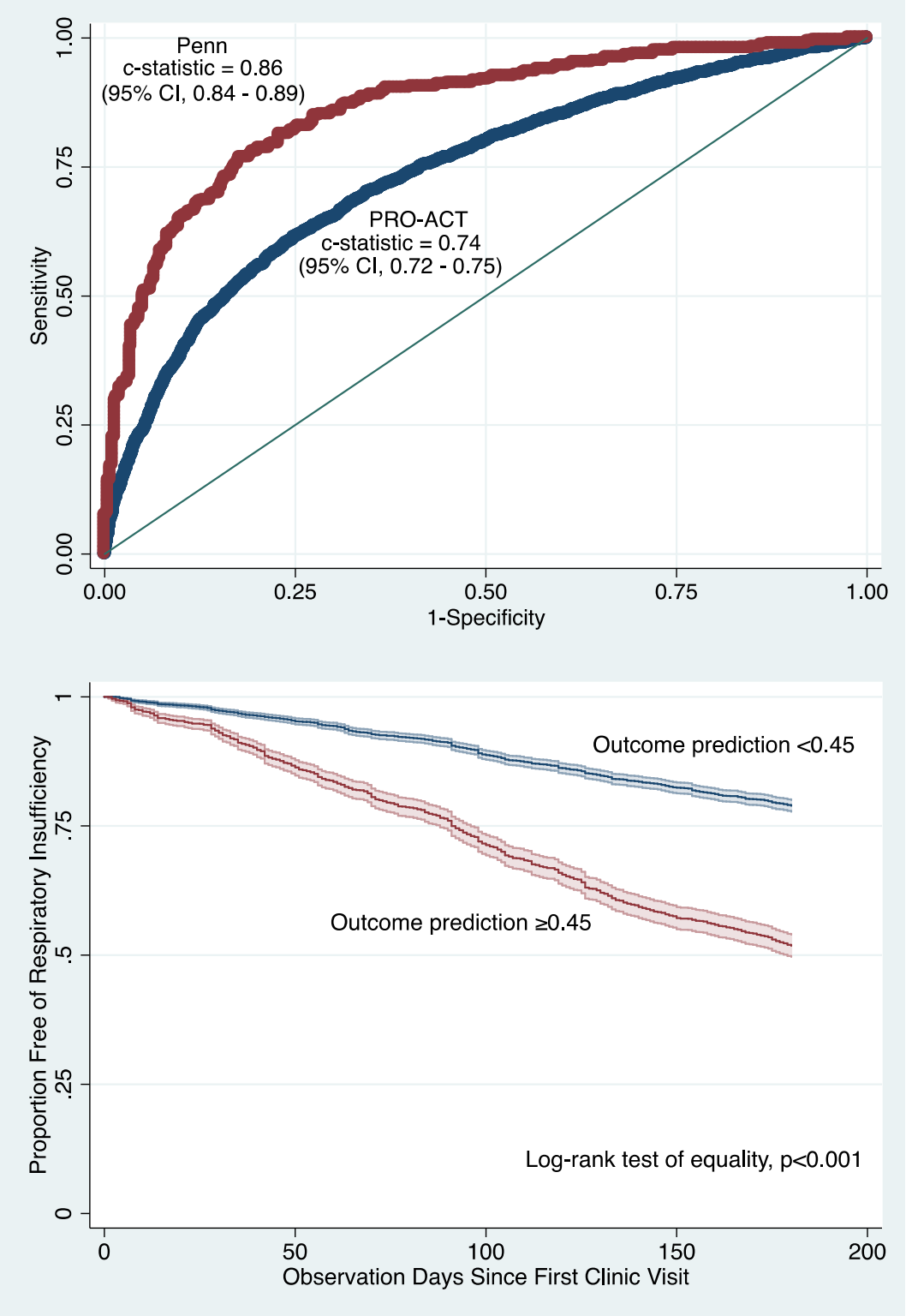


Table 1. Baseline characteristics of Penn cohort (n=765)

Variable	
Age at diagnosis, years	63 ± 12
Male sex, n (%)	440 (58)
Race, n (%)	
Caucasian	662 (87)
African-American	53 (7)
Other	50 (6)
BMI class, n (%)	
<18.5 kg/m^2	30 (4)
18.5 - 24.9 kg/m^2	318 (42)
25 - 29.9 kg/m^2	270 (35)
>30 kg/m^2	147 (19)
Diagnosis delay, years	1.0 (0.6, 1.7)
El Escorial criteria, n (%)	
Definite ALS	152 (20)
Possible ALS	197 (26)
Probable ALS	239 (31)
Suspected ALS	177 (23)
Symptom onset site, n (%)	
Limb	596 (78)
Bulbar	169 (22)
FVC % predicted	81 ± 18
ALSFRS-R total score	37 ± 6
ALSFRS-R dyspnea, n (%)	
4	454 (59)
3	205 (27)
2	74 (10)
1	31 (4)
0	1 (<1)
ALSFRS-R orthopnea, n (%)	
4	657 (86)
3	43 (5)
2	42 (5)
1	2 (<1)
0	21 (3)
Smoking history, n (%)	
Current	74 (10)
Previous	285 (37)
Never	406 (53)
Coronary artery disease, n (%)	70 (9)
Diabetes mellitus, n (%)	81 (11)
Hypertension, n (%)	297 (39)

Definition of abbreviations: BMI = body mass index; ALS = amyotrophic lateral sclerosis, FVC = forced vital capacity; ALSFRS-R = ALS functional rating scale – revised

Data are mean ± SD or median (25th percentile, 75th percentile).

Table 2. Baseline Penn cohort characteristics by composite outcome including death.

Variable	No composite outcome (n=465)	Composite outcome achieved (n=300)	P Value
Age at diagnosis, years	61 ± 12	65 ± 11	<0.001
Male sex, n (%)	272 (59)	168 (56)	0.50
Race, n (%)			
Caucasian	410 (88)	252 (84)	0.25
African-American	28 (6)	25 (8)	
Other	27 (6)	23 (8)	
BMI class, n (%)			
<18.5 kg/m^2	9 (2)	21 (7)	<0.001
18.5 - 24.9 kg/m^2	186 (40)	132 (44)	
25 - 29.9 kg/m^2	177 (38)	93 (31)	
>30 kg/m^2	93 (20)	54 (18)	
Diagnosis delay, years	1.0 (0.6, 2.0)	0.8 (0.5, 1.3)	<0.001
El Escorial criteria, n (%)			
Definite ALS	62 (13)	90 (30)	<0.001
Possible ALS	125 (27)	72 (24)	
Probable ALS	154 (33)	85 (28)	
Suspected ALS	124 (27)	53 (18)	
Symptom onset site, n (%)			
Limb	380 (82)	216 (72)	0.002
Bulbar	85 (18)	84 (28)	
FVC % predicted	89 ± 15	69 ± 14	<0.001
ALSFRS-R total score	38 ± 5	34 ± 6	<0.001
ALSFRS-R dyspnea, n (%)			
4	316 (68)	138 (46)	<0.001
3	118 (25)	87 (29)	
2	26 (6)	48 (16)	
1	4 (1)	27 (9)	
0	1 (<1)	0 (0)	
ALSFRS-R orthopnea, n (%)			
4	432 (93)	225 (75)	<0.001
3	20 (4)	23 (7)	
2	10 (2)	32 (11)	
1	0 (0)	2 (1)	
0	3 (1)	18 (6)	
Smoking history, n (%)			
Never	256 (55)	150 (50)	0.11
Previous	160 (34)	125 (42)	
Current	49 (11)	25 (8)	
Coronary artery disease, n (%)	34 (7)	36 (12)	0.028
Diabetes mellitus, n (%)	31 (7)	50 (17)	<0.001
Hypertension, n (%)	161 (35)	136 (45)	0.003

Definition of abbreviations: BMI = body mass index; ALS = amyotrophic lateral sclerosis; FVC = forced vital capacity; ALSFRS-R = ALS functional rating scale – revised.
Data are mean ± SD.
Data compared using t-test, chi-squared test, or Wilcoxon-Mann-Whitney test.

Table 3. Results of logistic regression analysis for respiratory insufficiency (n=765)

	Univariate Analysis			Multivariate Analysis		
Variable	**OR**	**95% CI**	**P Value**	**OR**	**95% CI**	**P Value**
Age at diagnosis, per decade	1.49	1.31 – 1.70	<0.001	1.13	0.96 – 1.32	0.14
Male sex	0.90	0.67 – 1.21	0.50			
Race						
Caucasian	--	--	--			
African-American	1.45	0.83 – 2.55	0.19			
Other	1.39	0.78 – 2.47	0.27			
BMI class (kg/m^2)						
<18.5	3.29	1.46 – 7.41	0.004			
18.5 - 24.9	--	--	--			
25 - 29.9	0.74	0.53 – 1.04	0.08			
>30	0.82	0.55 – 1.22	0.33			
Diagnosis delay, per year	0.93	0.84 – 1.01	0.09	0.77	0.67 – 0.88	<0.001
El Escorial criteria						
Definite ALS	--	--	--			
Possible ALS	0.40	0.26 – 0.61	<0.001			
Probable ALS	0.38	0.25 – 0.58	<0.001			
Suspected ALS	0.29	0.19 – 0.46	<0.001			
Symptom onset site						
Limb	--	--	--	--	--	--
Bulbar	1.62	1.15 – 2.30	0.006	1.70	1.08 – 2.67	0.02
FVC, per 10% decrease	2.65	2.30 – 3.06	<0.001	2.36	2.04 – 2.74	<0.001
ALSFRS-R total, per 6 decrease	2.12	1.80 – 2.50	<0.001	1.59	1.29 – 1.95	<0.001
ALSFRS-R dyspnea						
>2	--	--	--	--	--	--
≤2	4.67	2.98 – 7.31	<0.001	1.82	1.02 – 3.26	0.04
ALSFRS-R orthopnea						
>2	--	--	--			
≤2	7.29	3.89 – 13.65	<0.001			
Smoking history						
Never	--	--	--			
Previous	1.33	0.98 – 1.82	0.07			
Current	0.87	0.52 – 1.47	0.603			
Coronary artery disease	1.73	1.06 – 2.83	0.03			
Diabetes mellitus	2.8	1.74 – 4.50	<0.001			
Hypertension	1.57	1.16 – 2.11	0.003			

Definition of abbreviations: OR = odds ratio; CI = confidence interval; ALS = amyotrophic lateral sclerosis; FVC = forced vital capacity.

Table 4. Derivation and external validation of the prognostic rule for respiratory insufficiency and death.

		c-statistic	95% CI	Sensitivity* (95% CI)	Specificity* (95% CI)	PPV* (95% CI)	NPV* (95% CI)
Respiratory insufficiency	Derivation	0.86	0.84 – 0.89	83% (71% – 91%)	81% (72% – 89%)	77% (65% – 86%)	86% (77% – 93%)
	External validation	0.74	0.72 – 0.75	53% (51% – 55%)	82% (81% – 83%)	62% (60% – 64%)	76% (75% – 77%)
Death	Derivation	0.84	0.80 – 0.89	83% (63% – 95%)	81% (73% – 87%)	46% (30% – 61%)	96% (91% – 99%)
	External validation	0.72	0.71 – 0.74	47% (44% – 50%)	82% (81% – 83%)	36% (34% – 38%)	88% (87% – 89%)

Definition of abbreviations: CI = confidence interval; PPV = positive predictive value; NPV = negative predictive value.
*Sensitivity, specificity, PPV, and NPV based on the following cut-points: respiratory insufficiency, ≥0.45; death, ≥0.11.

Table 5. Baseline characteristics of PRO-ACT cohort (n=7,083)

Variable	
Age at diagnosis, years	56 ± 12
Male sex, n (%)	4371 (62)
Race, n (%)	
Caucasian	6763 (96)
African-American	106 (1)
Other	214 (3)
BMI class, n (%)	
<18.5 kg/m^2	464 (7)
18.5 - 24.9 kg/m^2	3060 (43)
25 - 29.9 kg/m^2	2376 (33)
>30 kg/m^2	1183 (17)
Diagnosis delay, years	0.8 (0.4, 1.3)
Symptom Onset Site, n (%)	
Limb	5517 (78)
Bulbar	1566 (22)
FVC % predicted	88 ± 20
ALSFRS-R total score	37 ± 6
ALSFRS-R dyspnea	
4	4887 (69)
3	1133 (16)
2	779 (11)
1	213 (3)
0	71 (1)
ALSFRS-R orthopnea	
4	5902 (84)
3	708 (10)
2	334 (5)
1	71 (1)
0	48 (<1)

Definition of abbreviations: BMI = body mass index; FVC = forced vital capacity; ALSFRS-R = amyotrophic lateral sclerosis functional rating scale – revised.

Data are mean ± SD or median (interquartile range 25th percentile, 75th percentile).

Online Data Supplement.

Methods

Study Samples and Data Collection

The Penn cohort data were prospectively entered into a secure online data portal known as the Penn Integrated Neurodegenerative Disease Database (INDD). Starting in 2006, patients diagnosed with ALS by an attending neurologist using the World Federation of Neurology El Escorial Criteria were approached for consent.[9] The El Escorial criteria categorize the presence of upper and lower motor neuron signs and classify patients as clinically “Definite”, “Probable”, “Possible”, and “Suspected” ALS, and these are standard inclusion criteria for patients in ALS clinical trials. After each clinic visit, an attending neurologist completed clinical data entry. Death was noted from hospital record, caregiver notification, or public record. Subjects were followed via outpatient neurology clinic visits at approximately three-month intervals. Follow-up was conducted until September 1, 2016.

PRO-ACT Cohort Additional Information

PRO-ACT is a longitudinal dataset follows patients from placebo and interventional arms from 23 phase II/III clinical trials. Data were provided by organizations such as the Northeast ALS Consortium, Novartis, Regeneron Pharmaceuticals Inc., Sanofi, and

Teva Pharmaceuticals Industries, Ltd. Prize4Life is a not-for-profit organization which created the PRO-ACT database in partnership with the Northeast ALS Consortium and the Neurological Clinical Research Institute at Massachusetts General Hospital with funding from the ALS Therapy Alliance.

The PRO-ACT database included de-identified data from standard protocol approvals. Registration and patient consents were obtained by the participating medical centers. In the rare cases where data were not already anonymized, they were further anonymized following the Health Insurance Portability and Accountability Act de-identification conventions for personal health information. All observation time in PRO-ACT was measured in days since trial enrollment as a point of reference, and thus trial enrollment day was used as the baseline start time in PRO-ACT.

PRO-ACT inclusion criteria commonly included being 18 years of age or older; ability to provide informed consent; clinical diagnosis of ALS; FVC ≥ 50% predicted; serum creatinine < 1.5 mg/dl (133 μmol/L) and disease duration of ≤ 5 years from symptom onset. Exclusion criteria for trials from PRO-ACT included known sensitivity or intolerability to the study drug under investigation in each trial; recent exposure to the study drug; exposure to other investigational agents within the last 30 days; malnourishment; substance abuse within the last year; active significant medical (cardiac, pulmonary, renal, hepatic, endocrine, hematologic, active malignancy or infectious disease) or psychiatric disease (psychosis or untreated major depression within 90 days of screening visit); human immunodeficiency virus; acquired immune deficiency syndrome (AIDS) or AIDS-related complex; pregnancy or breastfeeding; or significant cardiac conduction abnormality identified on screening electrocardiogram.

ALSFRS-R Score

The ALS Functional Rating Scale-Revised (ALSFRS-R) score is a standardized, widely-used method for staging functional status of ALS patients in clinical care as well as clinical trials.[21-24] The ALSFRS-R has been shown to correlate with progression of disease and survival, as well as have validity and reliability.[23, 25] There are 12 questions covering four domains, including gross motor tasks, fine motor tasks, bulbar function, and respiratory function. Each question rates an individual's function for that domain on a scale of 0 (minimal function) to 4 (maximal function). The range of the total ALSFRS-R score is from 0 (most severe symptoms) to 48 (minimal to no symptoms). The ALSFRS-R serves as a granular version of the ALSFRS for assessing respiratory function by splitting the ALSFRS respiratory score into three separate questions: dyspnea, orthopnea, and respiratory insufficiency.[26]

Statistical Analysis

Data were summarized using mean ± standard deviation or median (interquartile range) for continuous variables, and number of subjects (%) for categorical variables. Baseline characteristics of those with and without respiratory insufficiency were compared using Student *t* tests, Wilcoxon rank-sum tests, Kruskal-Wallis test, χ^2 tests or Fisher exact tests, as appropriate.

Logistic Regression Univariable Analysis

Using the Penn cohort, a logistic regression model incorporated predictors of respiratory insufficiency at six months from presentation. A univariable logistic regression analysis was performed with baseline characteristics such as age, self-reported race, gender, diagnosis age, symptom-onset date, diagnosis delay (time between symptom onset and diagnosis date), symptom-onset site, El Escorial criteria at first evaluation, ALSFRS-R scores (total, dyspnea, orthopnea, and respiratory insufficiency categories), body mass index (BMI), comorbidities, FVC, any history of smoking, and smoking pack-years.

Results were expressed as odds ratios (ORs) and their corresponding 95% confidence intervals (95% CI). We used a p-value <0.2 for selecting variables from the univariate analysis for inclusion into the multivariate model. We used purposeful backward selection[27] to retain all covariates with a p-value of <0.2 in the final multivariable model.

Derivation and Validation of Prognostic Model

Derivation of the multivariable logistic regression model was performed using the entire Penn cohort. To assess model discrimination capabilities, we used receiver operating characteristic (ROC) curves and c-statistics, also known as the area under the curve (AUC). Internal validation of the multivariable logistic regression model was performed using k-fold cross validation within the Penn cohort.[28] Using random assignment, out-of-sample data were created in approximately 4-to-1 ratio of estimation to prediction sample. This procedure was repeated 500 times. Using in-sample data

from the Penn cohort, we determined the probability of respiratory insufficiency. Using the Penn cohort out-of-sample data, a probability cut-off for a “positive test” was determined that prioritized sensitivity over specificity in order to capture individuals at high risk for the outcome. After constructing the prognostic rule and cutoff from the Penn cohort, we pursued external validation by applying the model to the PRO-ACT dataset. We applied the above approach to develop separate probability cut-offs for the composite outcome of respiratory insufficiency (including death) and death alone.

Assessment of Model Fit

We depicted time to respiratory insufficiency stratified by the prognosis probability cut-off using Kaplan-Meier curves and compared them using the log-rank test.[29] We assessed the calibration of the prognostic rule using both the Hosmer-Lemeshow goodness-of-fit test and the standardized Pearson χ^2 test.[30, 31] We graphically represented calibration for external validation in the PRO-ACT cohort. A lowess smoother line depicted expected to observed events among individuals with similar event probabilities across 12 equal portions of the cohort.

Multiple Imputation

Missing data from both the Penn and PRO-ACT cohorts were addressed using multiple imputation by chained equations with creation of 20 imputed datasets.[32, 33] We registered variables with complete data as predictors for imputation, including diagnosis age, age at symptom onset, gender, race, smoking history, and visit date.

Statistical significance was determined by p-values < 0.05. All analyses were performed using Stata version 15.0 (StataCorp LP, College Station, TX).

Multiple Imputation for ALSFRS-R Scores

Within the PRO-ACT cohort, 32% (n=3,412) of subjects had their ALS staged by the ALSFRS and 69% (7,311) by the ALSFRS-R. The difference between the ALSFRS and the ALSFRS-R lies in the assessment of the respiratory domain. The ALSFRS has 10 questions, with one of them focusing on the respiratory domain. The first nine questions of the ALSFRS-R are identical to the ALSFRS; however, questions 10-12 of the ALSFRS-R assess the respiratory domain in a more granular fashion by separately scoring the degree of dyspnea, orthopnea, and respiratory insufficiency. To maximize power in PRO-ACT, we converted all subjects with an ALSFRS score to an ALSFRS-R score by imputing the missing respiratory scores (dyspnea, orthopnea, and respiratory insufficiency). We used existing data of individuals with ALSFRS-R scores to perform multiple imputation by chained equations to complete the three missing respiratory questions for all individuals with an ALSFRS score. Finally, we summed the answers of the 9 original ALSFRS questions to the three new, imputed respiratory questions to obtain an ALSFRS-R total score for all individuals.

Sensitivity Analyses

Given that in the United States guidelines for ALS respiratory care and the Centers for Medicare & Medicaid Services both recognize FVC < 50% as a threshold for starting NIV, we hypothesized that FVC may be tightly linked to NIV initiation timing. Therefore,

we performed separate sensitivity analyses after removing NIV initiation and FVC from the outcome. In the first sensitivity analysis, we assessed our model performance with a composite outcome including FVC < 50% of predicted, tracheostomy placement, or death. In the second sensitivity analysis, we assessed model performance with a composite outcome including initiation of NIV, tracheostomy placement, or death. Lastly, we used the original composite outcome to compare performance of our full multivariate model against using FVC alone as a predictor.

Results

Sensitivity Analyses

When removing NIV from the composite outcome, the Penn cohort derivation c-statistic was 0.86 (95% CI, 0.83 – 0.89) and the internal cross-validation c-statistic was 0.85 (95% CI, 0.84 – 0.86). The PRO-ACT external validation c-statistic was 0.73 (95% CI, 0.72 – 0.75) (**Figure E7**). The cut-point of 0.45 corresponded to a sensitivity of 51% (95% CI, 49 – 53%), specificity of 84% (95% CI, 83 – 85%), PPV of 62% (60 – 64%), and NPV of 77% (95% CI, 77 – 79%).

When removing FVC from the composite outcome, the Penn cohort derivation c-statistic was 0.84 (95% CI, 0.81 – 0.87) and the internal cross-validation c-statistic was 0.84 (95% CI, 0.83 – 0.84). Using the cut-point of 0.45, the PRO-ACT external validation c-statistic was 0.76 (95% CI, 0.74 – 0.77) (**Figure E8**). The cut-point of 0.45 corresponded to a sensitivity of 43% (95% CI, 41 – 46%), specificity of 88% (95% CI, 88 – 89%), PPV of 49% (95% CI, 46 – 52%), and NPV of 86% (95% CI, 85 – 87%).

When using FVC only as a predictor of the composite outcome, the c-statistic was 0.83 (95% CI, 0.80-0.86) in the Penn cohort, while the full model c-statistic was significantly higher at 0.86 (95% CI, 0.84-0.89) ($p<0.001$ for the comparison). The c-statistic for FVC only in the PRO-ACT external validation was 0.70 (95% CI, 0.69-0.72) while the full model c-statistic was also significantly higher at 0.74 (95% CI, 0.72-0.75) ($p<0.001$ for the comparison).

Figure E1. Receiver operating characteristic curve for internal cross-validation of Penn cohort prognosis of respiratory insufficiency (including death) within six months.

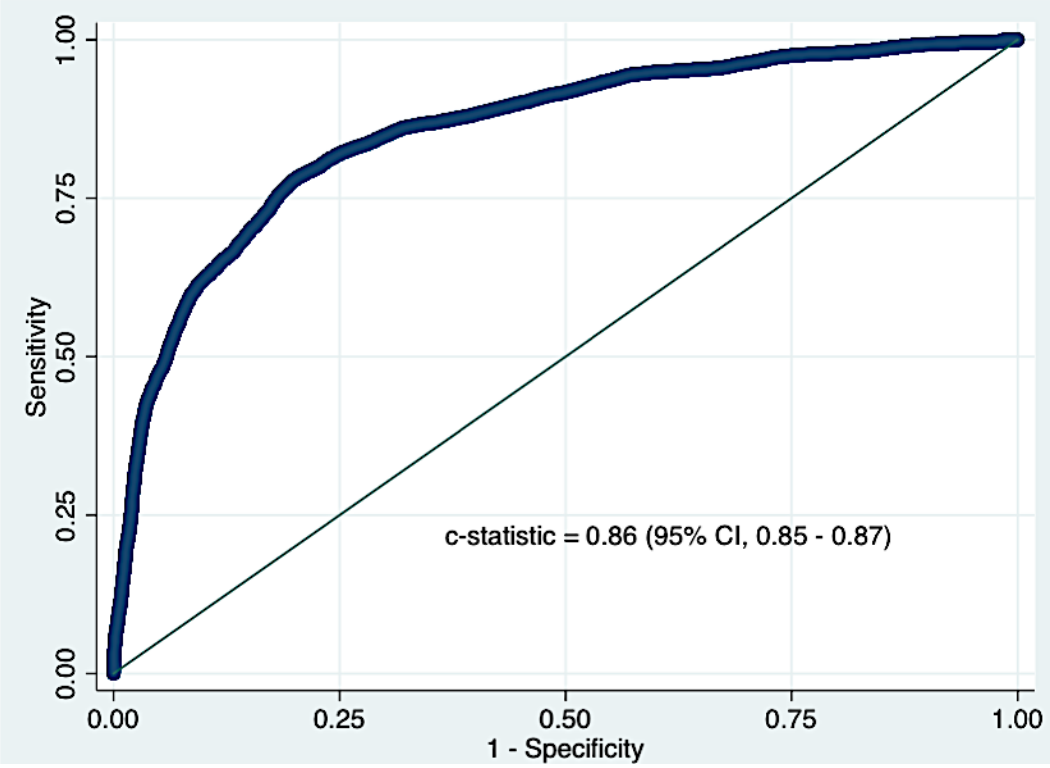


Figure E2. Kaplan-Meier curves with 95% confidence intervals stratified by prognostic probability of respiratory insufficiency in the Penn cohort, truncated at 365 days.

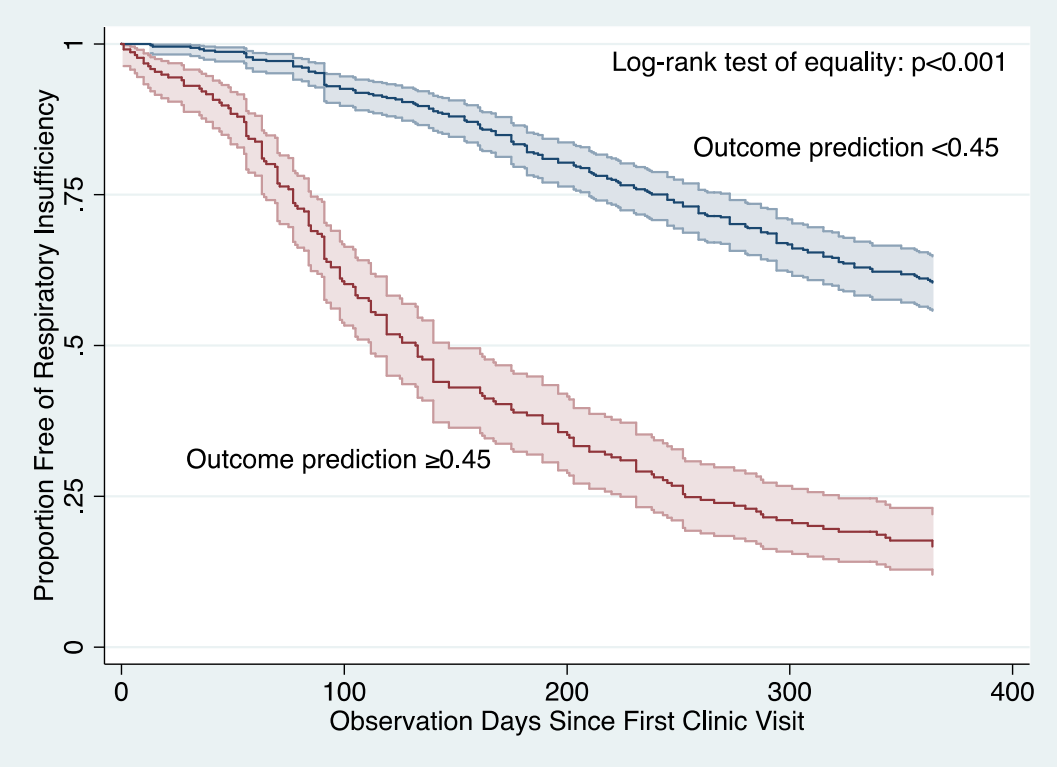


Figure E3. Receiver operating characteristic curve for overall Penn cohort prognosis of death within six months.

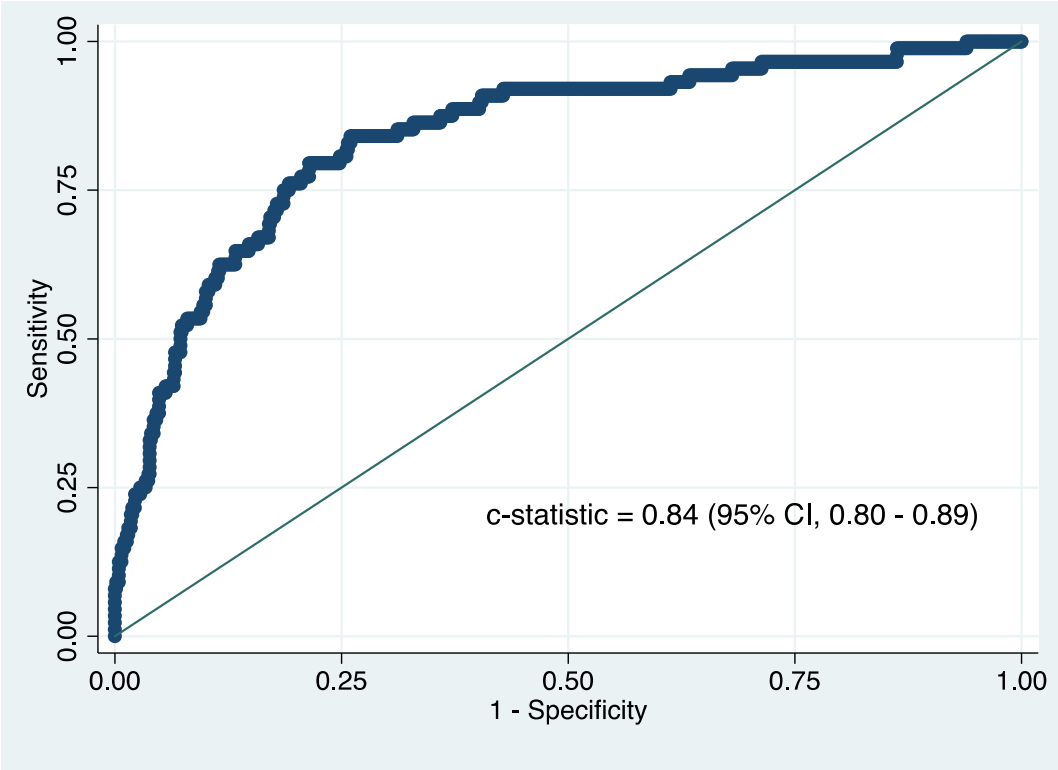


Figure E4. Receiver operating characteristic curve for internal cross-validation of death within six months.

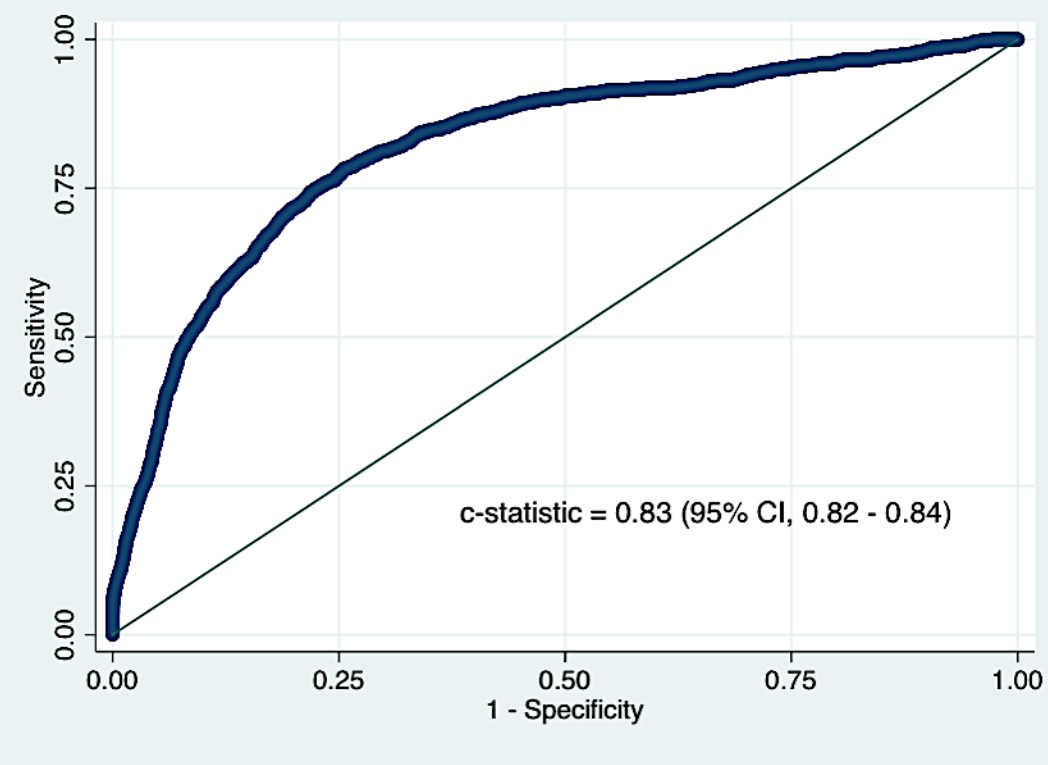
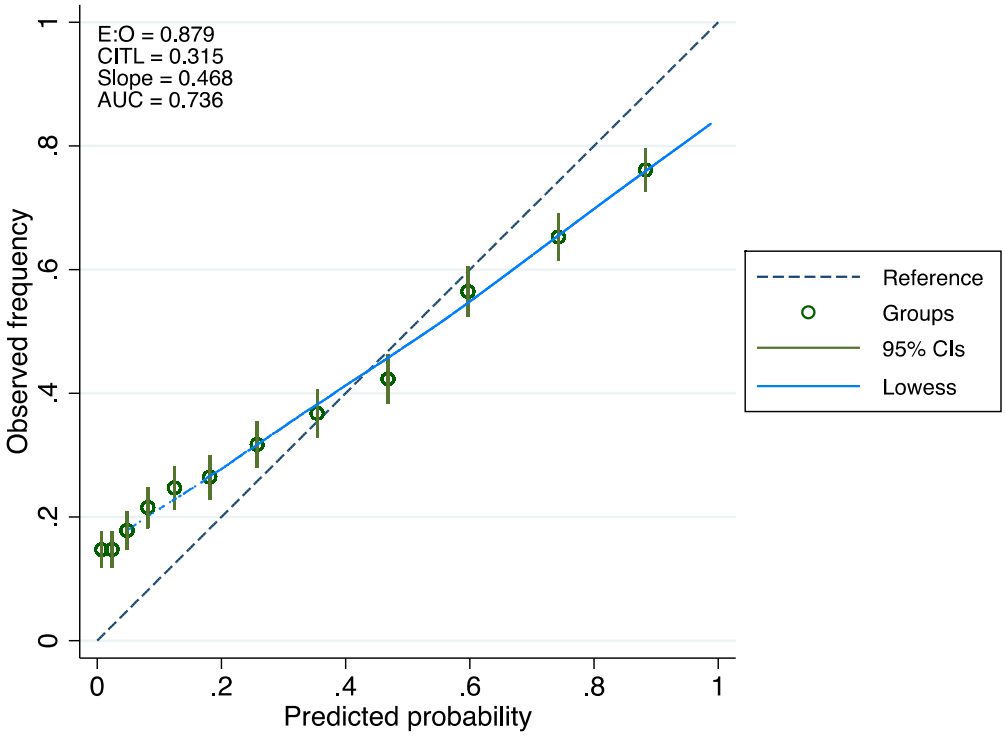


Figure E5. Calibration curve for external validation of prognostic model.



Definition of abbreviations: E:O = ratio of expected to observed events; CITL = calibration in the large; AUC = area under the curve.
Groups represent 12 equal portions of the cohort for comparing expected:observed events across different ranges of probability estimates.

Figure E6. Receiver operating characteristic curve for derivation and external validation of death within six months. Test of equality, p<0.001.

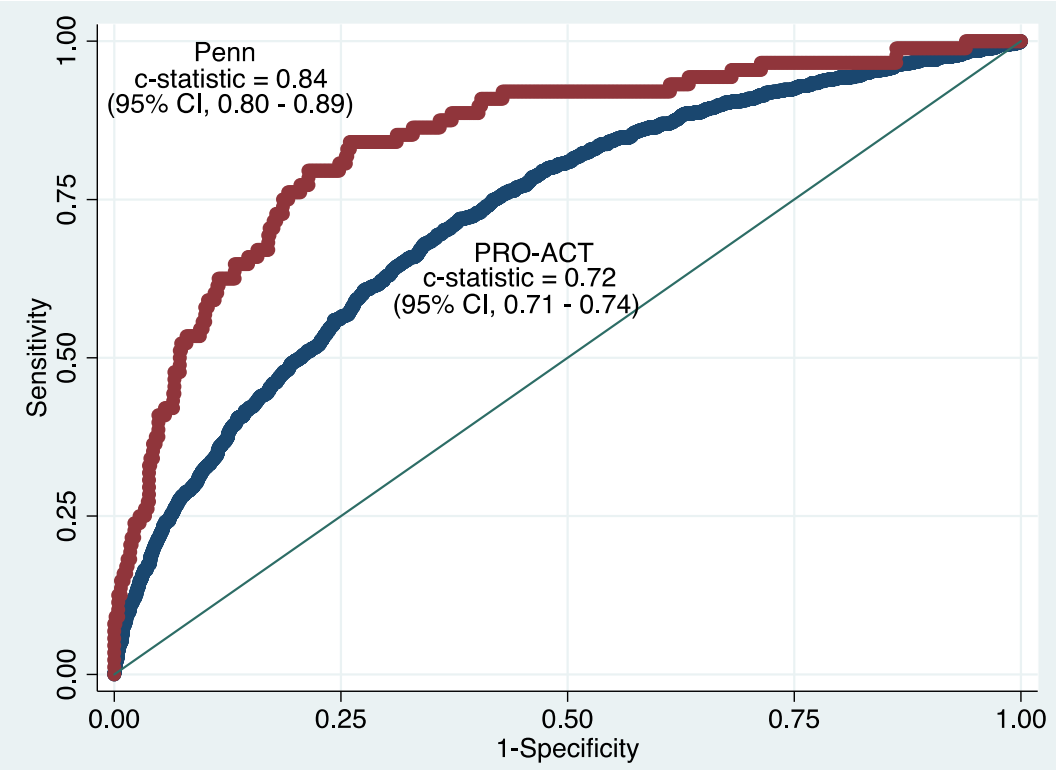


Figure E7. Receiver operating characteristic curves for derivation and external validation of the composite outcome of FVC < 50% predicted, tracheostomy placement, or death within six months. Test of equality, p<0.001.

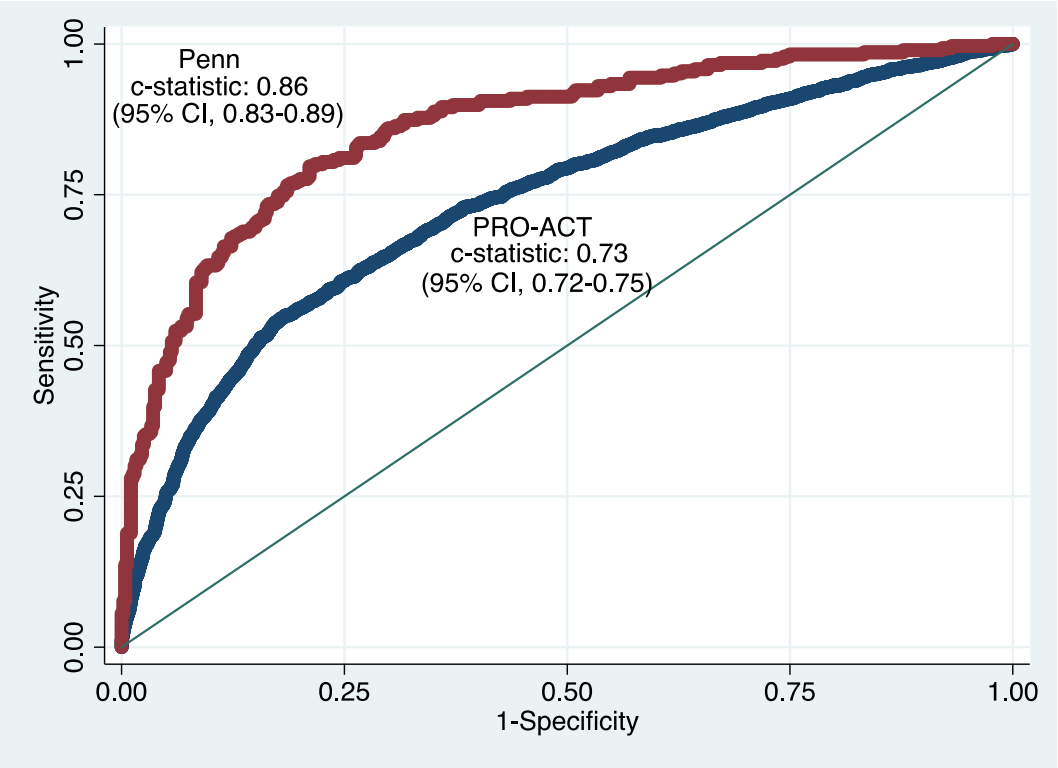


Figure E8. Receiver operating characteristic curves for derivation and external validation of the composite outcome of NIV initiation, tracheostomy placement, or death within six months. Test of equality, p<0.001.

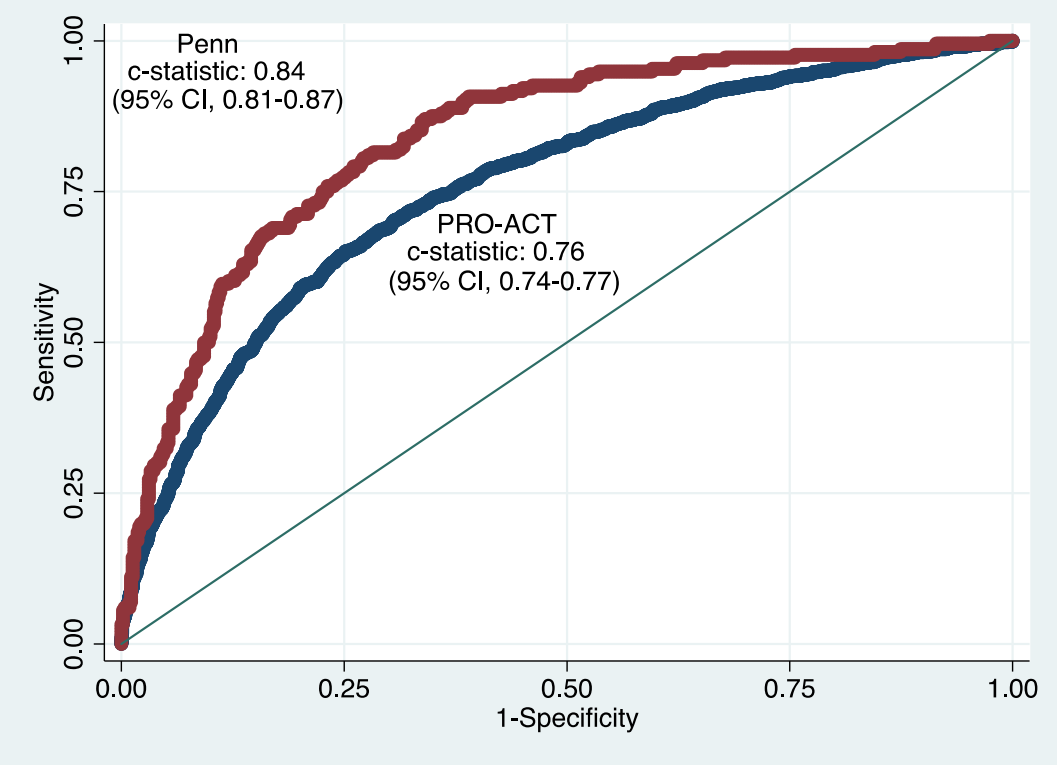


Table E1. Hosmer-Lemeshow goodness-of-fit test for respiratory insufficiency model in Penn derivation cohort (p=0.45).

Group	Probability	Observed outcomes	Expected outcomes	Total individuals
1	0.03	3	1	64
2	0.08	4	3	64
3	0.11	7	6	64
4	0.16	10	9	63
5	0.23	5	12	64
6	0.33	16	18	64
7	0.43	24	24	64
8	0.56	31	31	63
9	0.66	39	39	64
10	0.77	49	46	64
11	0.89	54	53	64
12	0.98	58	59	63

Table E2. Baseline PRO-ACT cohort characteristics by composite outcome including death.

Variable	No composite outcome (n=4630)	Composite outcome achieved (n=2453)	P value
Age at diagnosis, years	55 ± 12	58 ± 12	<0.001
Male sex, n (%)	2928 (63)	1443 (59)	<0.001
Race, n (%)			
Caucasian	4429 (96)	2334 (95)	0.57
African-American	65 (1)	41 (2)	
Other	136 (3)	78 (3)	
BMI class, n (%)			
<18.5 kg/m^2	217 (5)	247 (10)	
18.5 - 24.9 kg/m^2	1864 (40)	1196 (49)	<0.001
25 - 29.9 kg/m^2	1668 (36)	708 (29)	
>30 kg/m^2	881 (19)	302 (12)	
Diagnosis delay, years	0.8 (0.5, 1.3)	0.8 (0.4, 1.2)	0.002
Symptom onset site, n (%)			
Limb	3849 (83)	1668 (68)	<0.001
Bulbar	781 (17)	785 (32)	
FVC % predicted	94 ± 19	78 ± 19	<0.001
ALSFRS-R total score	38 ± 6	35 ± 7	<0.001
ALSFRS-R dyspnea, n (%)			
4	3566 (78)	1641 (65)	
3	577 (13)	467 (19)	<0.001
2	316 (7)	322 (13)	
1	84 (2)	89 (3)	
0	11 (<1)	10 (<1)	
ALSFRS-R orthopnea, n (%)			
4	4128 (91)	2074 (82)	
3	288 (6)	282 (11)	
2	112 (3)	140 (6)	<0.001
1	22 (<1)	23 (1)	
0	4 (<1)	10 (<1)	

Definition of abbreviations: BMI = body mass index; ALS = amyotrophic lateral sclerosis; FVC = forced vital capacity; ALSFRS-R = ALS functional rating scale – revised.
Data are mean ± SD or median (25th percentile, 75th percentile).
Data compared using t-test, chi-squared test, or Wilcoxon-Mann-Whitney test.

Table E3. Hosmer-Lemeshow goodness-of-fit test for respiratory insufficiency model in PRO-ACT cohort ($p<0.001$).

Group	Probability	Observed outcomes	Expected outcomes	Total individuals
1	0.01	87	4	591
2	0.04	87	14	590
3	0.06	105	28	590
4	0.10	127	48	590
5	0.15	146	74	591
6	0.22	156	107	590
7	0.30	187	152	590
8	0.41	217	209	590
9	0.53	250	277	591
10	0.67	333	352	590
11	0.81	385	438	590
12	0.99	449	521	590