




Development of a prognostic model of respiratory insufficiency or death in amyotrophic lateral sclerosis

Jason Ackrivo¹, John Hansen-Flaschen ¹, E. Paul Wileyto²,
Richard J. Schwab¹, Lauren Elman^{3,4} and Steven M. Kawut^{1,2,4}

Affiliations: ¹Dept of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. ²Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. ³Dept of Neurology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. ⁴These authors contributed equally.

Correspondence: Jason Ackrivo, Pulmonary Allergy and Critical Care Division, 839 West Gates Building, 3400 Spruce Street, Philadelphia, PA 19104, USA. E-mail: jason.ackrivo@uphs.upenn.edu



@ERSpublications

ALS has a heterogeneous progression to respiratory failure. A clinician can use the clinical prognostic rule to estimate a 6-month risk of respiratory failure onset, thus facilitating referrals and respiratory interventions <http://ow.ly/LLzP30nqm2D>

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; 53: 1802237 [<https://doi.org/10.1183/13993003.02237-2018>].

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 6 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-centre cohort of 765 ALS patients who presented between 2006 and 2015. External validation was performed using the multicentre Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) database with 7083 ALS patients. Predictors included baseline characteristics at first outpatient visit. The primary outcome was respiratory insufficiency within 6 months, defined by initiation of noninvasive ventilation, forced vital capacity (FVC) <50% predicted, tracheostomy, or death.

Of 765 patients in our centre, 300 (39%) had respiratory insufficiency or death within 6 months. Six baseline characteristics (diagnosis age, delay between symptom onset and diagnosis, FVC, symptom onset site, amyotrophic lateral sclerosis functional rating scale-revised (ALSFERS-R) total score and ALSFERS-R dyspnoea score) were used to prognose the risk of the primary outcome. The derivation cohort c-statistic was 0.86 (95% CI 0.84–0.89) and internal cross-validation produced a c-statistic of 0.86 (95% CI 0.85–0.87). External validation of the model using the PRO-ACT 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.

This article has supplementary material available from erj.ersjournals.com

Received: Nov 24 2018 | Accepted after revision: Jan 17 2019

Copyright ©ERS 2019

Introduction

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease with high morbidity and universal mortality. ALS is usually sporadic with an incidence of approximately 1–2 per 100 000 people in the US per year and a prevalence of approximately 3–5 per 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 noninvasive 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 the 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 for preparing patients for shared decision making, for improving the timeliness of referrals for respiratory interventions and for 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 6 months of presentation to an ALS centre.

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, PA, USA; the Penn cohort) with a first visit between January 01, 2006 and December 31, 2015. The source population for the validation cohort was the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) database and details on this dataset (the PRO-ACT cohort) can be found in the supplementary material.

This study was approved by the University of Pennsylvania institutional review board and adheres to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (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 01, 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 forced vital capacity (FVC) <50% predicted. We also excluded anyone with less than 6 months of follow-up time who were censored as “alive”. Further details regarding data collection for the Penn cohort dataset can be found in the supplementary material.

The PRO-ACT database includes de-identified data from 23 Phase II/Phase III clinical trials (supplementary material). PRO-ACT inclusion criteria commonly included being 18 years of age or older, an ability to provide informed consent, clinical diagnosis of ALS, FVC \geq 50% predicted, serum creatinine $<1.5 \text{ mg}\cdot\text{dL}^{-1}$ ($133 \text{ }\mu\text{mol}\cdot\text{L}^{-1}$) and disease duration of \leq 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. Further information regarding the amyotrophic lateral sclerosis functional rating scale-revised (ALSFERS-R) score can be found in the supplementary material.

Outcomes

We sought to establish a discriminating clinical prognostic rule for respiratory insufficiency within 6 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 supplementary material.

Statistical analysis

We created a prognostic model using multivariable logistic regression to identify associations between baseline characteristics and respiratory insufficiency within 6 months. After deriving the model on the Penn cohort and internally cross-validating, we externally validated it with the PRO-ACT cohort. We represented our model performance graphically using receiver operating characteristic (ROC) curves, Kaplan–Meier curves and calibration plots. Further details can be found in the supplementary material.

Results

The University of Pennsylvania Comprehensive ALS Center cohort (the Penn Cohort)

Patients with ALS (n=1061) were evaluated at the University of Pennsylvania Comprehensive ALS Center during the study period (the Penn cohort). After excluding those with NIV or tracheostomy at baseline (n=33), those with baseline FVC <50% predicted (n=168), those with unusable data or non-physiologic values (n=64) and those with less than 6 months follow-up (n=31), there were 765 individuals left in the Penn cohort. Mean age at diagnosis was 63 years, 58% of subjects were male and 87% were Caucasian (table 1). Average baseline FVC was 81%, mean baseline ALSFRS-R total score was 37 and 53% of the cohort classified themselves as “never smokers”. Median follow-up time was 2.3 years (interquartile range (IQR) 1.5–4.0 years) and 39% of the cohort (n=300) had respiratory insufficiency (or were deceased) by 6 months of observation.

We compared baseline characteristics of patients who had respiratory insufficiency by 6 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” under the El Escorial criteria, higher

TABLE 1 Baseline characteristics of the Penn cohort (n=765)

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

Data are presented as n (%), mean±SD or median (interquartile range). BMI: body mass index; ALS: amyotrophic lateral sclerosis; FVC: forced vital capacity; ALSFRS-R: amyotrophic lateral sclerosis functional rating scale-revised.

TABLE 2 Baseline Penn cohort characteristics by composite outcome (including death)

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

Data are presented as n (%), mean±SD, or median (interquartile range). Comparisons between data are made using the t-test, the Chi-squared test, or the Wilcoxon–Mann–Whitney test. BMI: body mass index; ALS: amyotrophic lateral sclerosis; FVC: forced vital capacity; ALSFRS-R: amyotrophic lateral sclerosis functional rating scale-revised.

proportion of bulbar symptom onset, lower FVC, lower ALSFRS-R total score, lower ALSFRS-R dyspnoea and ALSFRS-R orthopnoea 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 ethnicity.

In the Penn cohort, 167 subjects (22%) initiated NIV, 246 (32%) had FVC <50% predicted, two (0.3%) underwent tracheostomy and 88 (12%) died by 6 months.

Prognosis of respiratory insufficiency or death in the Penn cohort

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 dyspnoea score. The model was well calibrated by both the Hosmer–Lemeshow test (p=0.45) (supplementary table E1) and the standardised Pearson

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

Characteristic	Univariate analysis		Multivariate analysis	
	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		
Ethnicity				
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	0.4 (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
ALSFRRS-R total score per 6 decrease	2.12 (1.80–2.50)	<0.001	1.59 (1.29–1.95)	<0.001
ALSFRRS-R dyspnoea score				
>2	–	–	–	–
≤2	4.67 (2.98–7.31)	<0.001	1.82 (1.02–3.26)	0.04
ALSFRRS-R orthopnoea score				
>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		

BMI: body mass index; ALS: amyotrophic lateral sclerosis; FVC: forced vital capacity; ALSFRS-R: amyotrophic lateral sclerosis functional rating scale-revised.

Chi-squared 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 four-fold (four-to-one) cross-validation of the multivariate logistic model in the Penn cohort for the composite outcome. The ROC curve 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) (supplementary figure E1). Using the internally cross-validated ROC curve, we selected a cut-off point of 0.45 for the odds of reaching respiratory insufficiency (including death) at 6 months, corresponding to a sensitivity of 83% and a 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 event 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 as being at high risk of respiratory insufficiency within 6 months.

Patients with a “positive” prediction for the odds of respiratory insufficiency had a significantly higher risk of the primary outcome within 6 months (figure 1b) and at 1 year (supplementary figure E2) (both $p<0.001$). The median time to respiratory insufficiency was 123 days (IQR 77–239 days) for the group with a “positive” prediction (odds of an event ≥ 0.45) and 469 days (IQR 259–889 days) for those with a “negative” prediction. A calculator for prognosticating the risk of respiratory insufficiency at 6 months is included in the supplementary material.

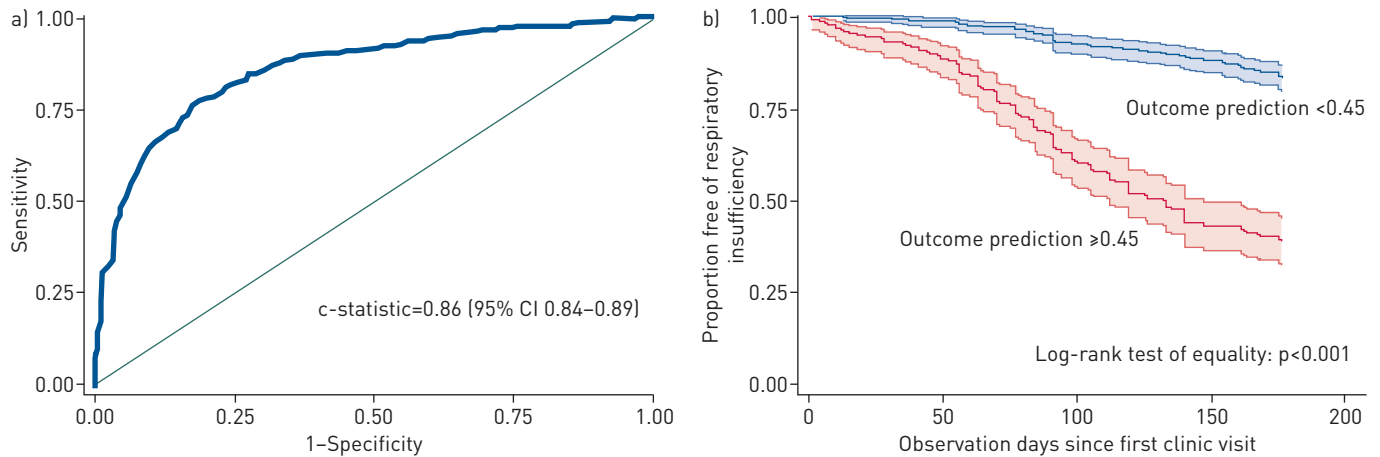


FIGURE 1 a) Receiver operating characteristic curve for prognosis of respiratory insufficiency (including death) within 6 months in the Penn cohort. b) Kaplan-Meier curves with 95% confidence intervals stratified by prognostic probability of respiratory insufficiency in the Penn cohort, truncated at 180 days.

Mortality in the Penn cohort

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

The Pooled Resource Open-Access ALS Clinical Trials cohort (the PRO-ACT cohort)

The PRO-ACT cohort contained 10723 subjects. We excluded 2473 patients with less than 6 months of follow-up time and those with prior tracheostomy or prior NIV (n=108), as well as anyone with FVC <50% predicted at the start of observation (n=1059). The final PRO-ACT cohort thus contained 7083 individuals with a mean age of 56 years. Of these subjects, 62% were male and 96% were Caucasian (table 5). Average baseline FVC was 88% and mean baseline ALSFRS-R total score was 37. The median follow-up time was 0.98 years (IQR 0.65–1.32 years) and 35% of subjects (n=2453) had respiratory insufficiency (including death) by 6 months of observation.

We compared baseline characteristics of individuals who did and did not meet our composite outcome within 6 months of diagnosis (supplementary 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 ethnicity.

At 6 months, 2453 individuals (35%) had respiratory insufficiency, 360 (5%) were initiated on NIV, 1398 (20%) had FVC <50% predicted and 1168 (16%) had died. No one had received a tracheostomy.

Prognosis of respiratory insufficiency or death in the PRO-ACT cohort

We applied the prognostic model and cut-off from the Penn cohort to the PRO-ACT cohort. The model yielded a c-statistic of 0.74 (95% CI 0.72–0.75) (figure 2a) and model performance using the outcome

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

Variable	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)

PPV: positive predictive value; NPV: negative predictive value. Sensitivity, specificity, PPV and NPV were based on the following cut-off points: respiratory insufficiency ≥ 0.45 and death ≥ 0.11 .

TABLE 5 Baseline characteristics of the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) cohort (n=7083)

Characteristic	Result
Age at diagnosis years	56±12
Male sex	4371 (62)
Ethnicity	
Caucasian	6763 (96)
African-american	106 (1)
Other	214 (3)
BMI class kg·m⁻²	
<18.5	464 (7)
18.5–24.9	3060 (43)
25–29.9	2376 (33)
≥30	1183 (17)
Diagnosis delay years	0.8 [0.4–1.3]
Symptom onset site	
Limb	5517 (78)
Bulbar	1566 (22)
FVC % predicted	88±20
ALSFRS-R total score	37±6
ALSFRS-R dyspnoea score	
4	4887 (69)
3	1133 (16)
2	779 (11)
1	213 (3)
0	71 (1)
ALSFRS-R orthopnoea score	
4	5902 (84)
3	708 (10)
2	334 (5)
1	71 (1)
0	48 (<1)

Data are presented as n (%), mean±sd or median [interquartile range]. BMI: body mass index; FVC: forced vital capacity; ALSFRS-R: amyotrophic lateral sclerosis functional rating scale-revised.

probability cut-off of ≥ 0.45 , which produced a sensitivity of 53% and a specificity of 82%, is shown in Table 4. The Hosmer–Lemeshow goodness-of-fit test gave $p < 0.001$ (supplementary table E3) and the standardised Pearson Chi-squared test gave $p = 0.001$. The calibration plot (supplementary 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 6 months ($p < 0.001$) (figure 2b). For the group with a “positive” prediction (odds of an event ≥ 0.45), median time to respiratory insufficiency was 182 days (IQR 91–344 days), while for those with a “negative” prediction it was 381 days (IQR 204–581 days).

Mortality in the PRO-ACT cohort

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

Sensitivity analyses

Details of sensitivity analysis can be found in the supplementary material.

Discussion

We found that younger age, less diagnostic delay, lower FVC, bulbar symptom onset site, lower ALSFRS-R total score and ALSFRS-R dyspnoea score of two or less at baseline were significantly associated with a

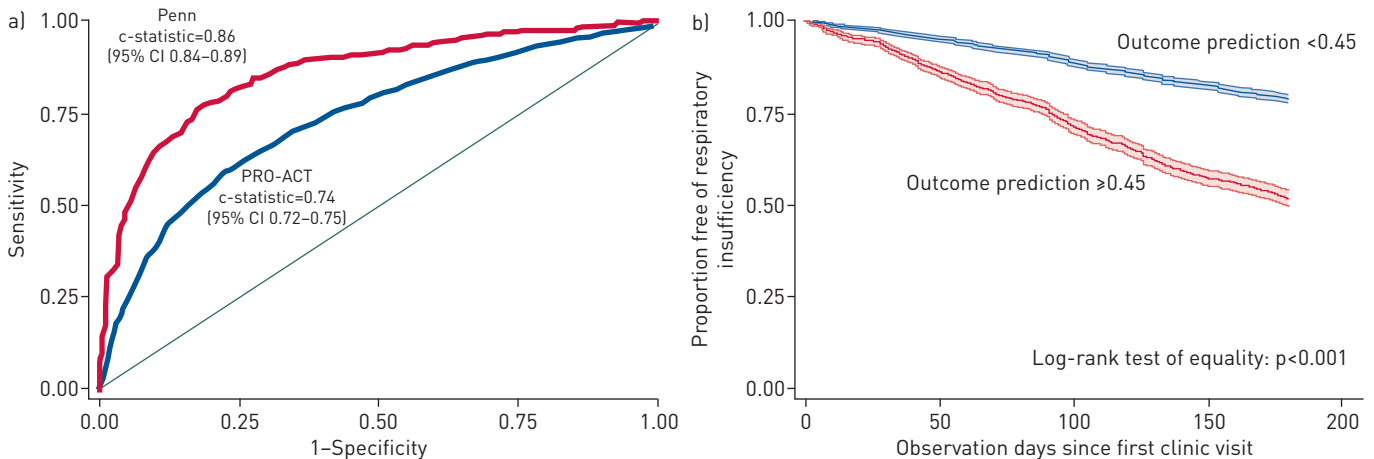


FIGURE 2 a) Receiver operating characteristic curves for derivation and external validation of respiratory insufficiency within 6 months (test of equality, $p < 0.001$). b) Kaplan-Meier curves with 95% confidence intervals stratified by prognostic probability of respiratory insufficiency in the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) cohort, truncated at 180 days.

higher risk of respiratory insufficiency or death at 6 months in a large single-centre 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 (and ALSFRS-R) score, ALSFRS (and ALSFRS-R) 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 in ALS of tracheostomy, dependence on NIV ($>23 \text{ h}\cdot\text{day}^{-1}$), or death [15]; however, our study differed in several important ways. Our model included variables which 1) are routinely clinically available for all patients, 2) are assessed at baseline, and 3) prognose the short-term risk of respiratory failure onset (an important clinical event for patients which could be used to create an “enriched” study population for clinical trials). Sensitivity, specificity, PPV and NPV were not presented in the study by WESTENENG *et al.* [15], 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 significant associations between age, bulbar onset disease and diagnosis delay with worsened survival that are consistent with our study [16–19]. KIMURA *et al.* [12] 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 results. Furthermore, crowdsourcing initiatives have used advanced machine-learning algorithms to identify 15 predictors (including time from symptom onset, FVC, age, site of symptom onset and ALSFRS total score) for distinguishing between “fast” versus “slow” relative disease progressors by change in ALSFRS score [13]. It was estimated that with such information sample size for clinical trial enrolment 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, multicentre, international patient population (with different inclusion criteria from the derivation cohort) attests to its generalisability.

We recognise 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 in the PRO-ACT validation cohort (at least in the higher-risk group), the cut-off 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 centre over 10 years, potentially making the prognostic rule generalisable to other centres, 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 appear 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 heterogeneous dataset attests to its generalisability.

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 the Penn cohort *versus* 5% in the PRO-ACT cohort). Given that PRO-ACT is a multicentre clinical trial database from cohorts around the world, our results may be affected by substantial practice variation regarding interventions for respiratory insufficiency. In the USA, guidelines and most medical insurance carriers recognise an FVC <50% predicted normal as a threshold for initiation of NIV. However, some centres 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 centres 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 randomised controlled trial data, which obtained FVC values using rigorous clinical research grade methodologies. Measurement error would, of course, likely bias to the null unless related to both the actual value of 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 ($p < 0.001$, data not shown).

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 7083 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 6 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 enrolment 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 test and the standardised Pearson Chi-squared test 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 “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 personalised medicine. Further study would be necessary to validate this as a tool to prospectively identify a high-risk subgroup suitable for clinical trial enrolment. 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.

Conflict of interest: None declared.

Support statement: Funds were provided by the National Institutes of Health (grants: T32 HL-007891, K24 HL-103844, P01 AG032953-A1 and F32 HL144145. Funding information for this article has been deposited with the Crossref Funder Registry.

References

- 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, *et al.* 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? *J Neurol* 2010; 257: 1235–1239.
- 5 Spataro R, Re Lo M, Piccoli T, *et al.* 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, *et al.* Effects of non-invasive ventilation on survival and quality of life in patients with amyotrophic lateral sclerosis: a randomised controlled trial. *Lancet Neurol* 2006; 5: 140–147.
- 7 Miller RG, Jackson CE, Kasarskis EJ, *et al.* 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* 2009; 73: 1218–1226.
- 8 Collins GS, Reitsma JB, Altman DG, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015; 162: 55–63.
- 9 Brooks BR, Miller RG, Swash M, *et al.* 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, *et al.* 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. *Amyotroph Lateral Scler Frontotemporal Degener* 2014; 15: 444–452.
- 12 Kimura F, Fujimura C, Ishida S, *et al.* Progression rate of ALSFRS-R at time of diagnosis predicts survival time in ALS. *Neurology* 2006; 66: 265–267.
- 13 Küffner R, Zach N, Norel R, *et al.* Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. *Nat Biotechnol* 2015; 33: 51–57.
- 14 Elamin M, Bede P, Montuschi A, *et al.* Predicting prognosis in amyotrophic lateral sclerosis: a simple algorithm. *J Neurol* 2015; 262: 1447–1454.
- 15 Westeneng H-J, Debray TPA, Visser AE, *et al.* Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. *Lancet Neurol* 2018; 17: 423–433.
- 16 Czaplinski A, Yen AA, Appel SH. Amyotrophic lateral sclerosis: early predictors of prolonged survival. *J Neurol* 2006; 253: 1428–1436.
- 17 Chiò A, Logroscino G, Hardiman O, *et al.* Prognostic factors in ALS: a critical review. *Amyotroph Lateral Sclerosis* 2009; 10: 310–323.
- 18 del Aguila MA, Longstreth WT, McGuire V, *et al.* Prognosis in amyotrophic lateral sclerosis: a population-based study. *Neurology* 2003; 60: 813–819.
- 19 Chio A, Mora G, Leone M, *et al.* 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 PD. Measures of fit for logistic regression (paper 1485-2014). SAS Global Forum, Washington, DC, March 23–26, 2014. <https://statisticalhorizons.com/wp-content/uploads/GOFForLogisticRegression-Paper.pdf> Date last accessed: March 15, 2018.