Predicting respiratory failure in amyotrophic lateral sclerosis: still a long way to go

To the Editor:

Amyotrophic lateral sclerosis (ALS) is a relentlessly progressive neurodegenerative disease that affects the brain and spinal cord motor neurons leading to progressive muscle weakness, irreversible disability, respiratory failure and death within a time range of a few months to more than 10 years [1]. Although use of noninvasive ventilation (NIV) is standard of care in ALS and improves health-related quality of life and survival [2, 3], no accurate model to predict respiratory failure in ALS was available until the recently published work by ACKRIVO *et al.* [4]. The authors should therefore be congratulated for this much awaited contribution to this field.

The authors used a single-centre cohort of ALS patients (the Penn cohort, n=765) to develop a prognostic model. Their model included six variables gathered at first outpatient visit to predict respiratory failure within 6 months. An external validation was performed using the PRO-ACT cohort, a large ALS dataset with individual data from 23 ALS clinical trials (n=7083). Discrimination, *i.e.* the ability to sort patients who will reach the endpoint from those who will not, was lower in the validation cohort (area under the curve (AUC) c-statistic 0.74, 95% CI 0.72–0.75) than in the derivation cohort (AUC c-statistic 0.86, 95% CI 0.84–0.89). External calibration using the PRO-ACT cohort also demonstrated a gross under-estimation of the risk in the lowest risk group. Indeed, as many as 15% of patients in this group developed respiratory failure within 6 months of their first outpatient visit (refer to table E3 of the original publication).

To be clinically useful, prognostic models should perform well in terms of discrimination and calibration and remain valid across different populations in order to provide an accurate risk prediction in single individuals.

We tested the performance of the model proposed by ACKRIVO *et al.* [4] in a single-centre cohort of 68 ALS patients, followed in a multidisciplinary programme in Geneva, Switzerland between 2012 and 2016 (registered at clinicaltrials.gov as NCT03536962) [5]. For validation of the prognostic score in our population, we assessed model discrimination using AUC c-statistic. Calibration was also assessed by visual inspection of a calibration plot [6].

Among 68 patients enrolled in our cohort, 16 patients were excluded from the analysis because they were already treated by NIV at baseline visit. Two more patients were excluded from the analysis because forced vital capacity was missing at baseline. Clinical characteristics of the 50 patients included in our analysis were slightly different than the Penn state cohort in terms of age at diagnosis (mean±SD 68±13 years), gender (44% male), body mass index (median 23 kg·m^{-2}, interquartile range 22–26 kg·m^{-2}) and risk factors included in the proposed model (time to diagnosis (0.8 years, interquartile range 0.6–1.3 years), bulbar disease (20%), forced vital capacity (93±20% pred) and functional impairment as measured by ALSFRS-R scores (39±6)). In our centre, NIV is most often started according to European Federation of Neurological Societies recommendations [3].

Among 50 patients included in the analysis, 10 patients (20%) developed respiratory insufficiency (or died) within 6 months of baseline visit. The AUC c-statistic was 0.74 (95% CI 0.58–0.91). Overall, the prognostic score overestimated the risk of respiratory failure, with an average predicted risk of 24%. Inspection of the
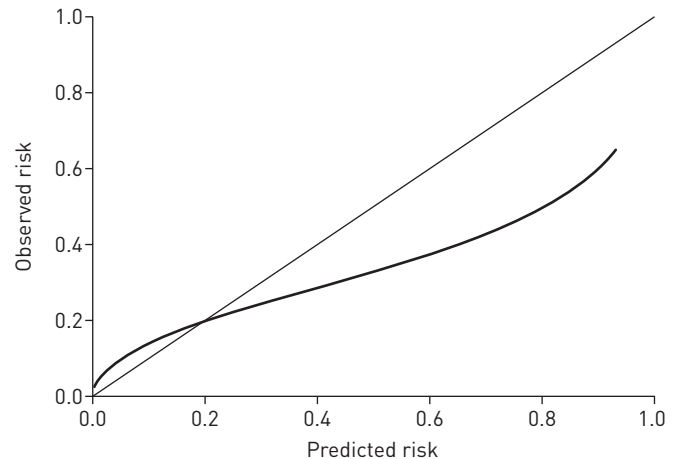
FIGURE 1 Calibration curve for external validation of prognostic model in the Geneva amyotrophic lateral sclerosis cohort (n=50). The prognostic model under-evaluated the risk of the composite outcome (noninvasive ventilation, death, tracheostomy or forced vital capacity <50% pred) in patients at low risk and over-evaluated the risk in patients at higher risk.

calibration curve showed slight under-estimation of the risk in patients at low risk and possibly an important overestimation of the risk in patients at higher risk (figure 1). Using a cut-off of 0.45 for the odds of presenting a respiratory insufficiency, as in the original publication by ACKRIVO et al. [4], model sensitivity was 50% (95% CI 19–81%) and specificity was 77% (95% CI 62–89%).

Although the model developed by ACKRIVO et al. [4] emphasises the prognostic importance of specific risk factors for respiratory failure within 6 months of a first visit to an outpatient ALS clinic, the clinical benefit of their model is yet to be demonstrated and clinicians should not rely solely on this model for starting NIV or planning advanced care. Indeed, our findings are in close agreement with the external validation results reported by ACKRIVO et al. [4]: a modest c-index of 0.74 and a calibration curve suggesting under-estimation of the risk in low-risk groups and an overestimation of the risk in higher-risk groups. These findings are suggestive of "over-optimism" in the development of the initial model. As illustrated here and in the PRO-ACT external validation, performance of the score is insufficient for prediction in individuals and may also be suboptimal in different populations. This might be due to real, and perhaps unmeasured, differences between populations (i.e. medical care, social or genetic background, or practices regarding tracheostomy). Other possibilities include referral bias and variable interpretation of tests [7, 8]. With this in mind, developing a new, improved tool to select patients at high risk of impending respiratory failure is still much-needed. Such a tool could also be used in clinical research to enrich a population at risk in order to improve the efficiency of interventional randomised control trials.

Dan Adler[1,2], Antoine Poncet[2,3], Ruxandra Iancu Ferfoglia[2,4], André Truffert[2,4] and Jean-Paul Janssens[1,2]
[1]Division of Pulmonary Diseases, Dept of Medicine, Geneva University Hospitals, Geneva, Switzerland. [2]Faculty of Medicine, University of Geneva, Geneva, Switzerland. [3]Center for Clinical Research and Division of Clinical Epidemiology, Dept of Health and Community Medicine, University of Geneva and University Hospitals of Geneva, Geneva, Switzerland. [4]Division of Neurology, Dept of Clinical Neurosciences, Geneva University Hospitals, Geneva, Switzerland.

Correspondence: Dan Adler, Division of Pulmonary Diseases, Geneva University Hospitals, Rue Gabrielle-Perret-Gentil 4, 1205, Geneva, Switzerland. E-mail: dan.adler@hcuge.ch

Conflict of interest: None declared.

References

1 Brown RH, Al-Chalabi A. Amyotrophic lateral sclerosis. N Engl J Med 2017; 377: 162–172.
2 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.
3 Andersen PM, Abrahams S, Borasio GD, et al. EFNS guidelines on the clinical management of amyotrophic lateral sclerosis (MALS) – revised report of an EFNS task force. Eur J Neurol 2012; 19: 360–375.
4 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.
5 Janssens JP, Adler D, Iancu Ferfoglia R, et al. Assessing inspiratory muscle strength for early detection of respiratory failure in motor neuron disease: should we use MIP, SNIP, or both? Respiration 2019; in press [https://doi.org/10.1159/000498972].
6 Finazzi S, Poole D, Luciani D, et al. Calibration belt for quality-of-care assessment based on dichotomous outcomes. PLoS One 2011; 6: e16110.
7 Barras AC H, Adler D, Iancu Ferfoglia R, et al. Is tracheostomy still an option in amyotrophic lateral sclerosis? Reflections of a multidisciplinary work group. Swiss Med Wkly 2013; 143: w13830.

8   Westeneng HJ, 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.

From the authors:

We appreciate the interest of D. Adler and colleagues in our recent manuscript describing a prognostic model for early respiratory insufficiency in amyotrophic lateral sclerosis (ALS) [1]. After applying our prediction model in a cohort of 50 ALS patients at their centre in Geneva, they obtained a c-statistic, sensitivity and specificity all virtually identical to our external validation in the Pooled Resource Open-Access Clinical Trials (PRO-ACT) ALS database. In addition, the Geneva cohort calibration curve resembles figure E5 from PRO-ACT in our supplementary material [1].

In this setting, perhaps the primary conclusion is that D. Adler and co-workers' single-centre results are very similar to those from our large multicentre validation cohort, despite differences in the study samples. First, 24% of patients in the Geneva group were using noninvasive ventilation (NIV) at baseline and were excluded, compared to 1% in the Penn cohort (and 4% in the PRO-ACT). The European Federation of Neurological Societies guidelines propose a NIV initiation threshold at forced vital capacity (FVC) <80% [2], which is much higher than that of the American Academy of Neurology guidelines (FVC <50%), which may explain these differences [3]. Only 20% of the 50 patients included in the Geneva cohort developed respiratory insufficiency or died within 6 months of observation, compared to 39% and 35% in Penn and PRO-ACT, differences likely due the exclusion of the sicker patients already using NIV from the study sample (selection bias).

Due to the differences in the underlying risk of respiratory failure in the Geneva sample, we calculated a lower positive predictive value (36%, 95% CI 13–65%) and higher negative predictive value (86%, 95% CI 71–95%) than PRO-ACT (62% and 76%, respectively). Of course, the small sample size of the Geneva cohort caused extremely wide 95% confidence intervals for the discrimination estimates and likely for the calibration curve (although they are not shown). Despite these differences and wide confidence intervals, the findings by D. Adler and co-workers closely resemble our external validation findings, testifying to the robustness and generalisability of our model.

The properties of discrimination and calibration of prediction rules support different uses. Both are key measurements for assessing the validity of prediction models. Calibration refers to the agreement between predicted and observed outcomes in a population. Discrimination refers to the model's ability to distinguish patients *versus* without an outcome. A model that predicts all individuals to have a risk equal to the actual incidence of an outcome would be a model with excellent calibration but poor discrimination. Highlighting that an average predicted risk of 24% is higher than the actual incidence of 20% does not fully characterise the model's discrimination or calibration abilities. However, the Geneva cohort had a similar c-statistic, sensitivity and specificity to PRO-ACT (thus similar discrimination) and a similar calibration curve which provided reasonably accurate estimates, realising that identical and consistent calibration at all levels of risk of the outcome may not be a realistic goal [4, 5].

We agree with the need for a useful, discriminating and calibrated prediction model for respiratory events in ALS to expedite timeliness of care, shape patient expectations and enrich clinical trial design. We also agree that further research is necessary before widespread clinical use of the prediction rule. For example, next steps may include applying the prediction rule to identify high-risk patients for inclusion in randomised clinical trials; stratification of randomised patients by the predicted risk of respiratory failure; or assessing how randomising patients/clinicians to receiving prediction results affects quality of life and respiratory outcomes. We agree with D. Adler and colleagues that more work needs to be done in early identification and treatment of ALS patients at high risk of respiratory failure.

Jason Ackrivo[1], Lauren Elman[2] and Steven M. Kawut[1,3]
[1]Dept of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. [2]Dept of Neurology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. [3]Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA.

Correspondence: Jason Ackrivo, University of Pennsylvania, Perelman School of Medicine, 3400 Spruce Street, 5048 West Gates Building, Philadelphia, PA 19104, USA. E-mail: jason.ackrivo@uphs.upenn.edu

References

1   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.
2   Andersen PM, Abrahams S, Borasio GD, *et al.* EFNS guidelines on the clinical management of amyotrophic lateral sclerosis (MALS) – revised report of an EFNS task force. *Eur J Neurol* 2012; 19: 360–375.
3   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.
4   Van Calster B, Nieboer D, Vergouwe Y, *et al.* A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 2016; 74: 167–176.
5   Alba AC, Agoritsas T, Walsh M, *et al.* Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA* 2017; 318: 1377–1384.