

RESEARCH ARTICLE

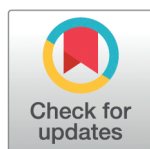
A meaningful prediction of functional decline in amyotrophic lateral sclerosis based on multi-event survival analysis

Christian Marius Lillelund^{1,2*}, Sanjay Kalra^{1,3,4}, Russell Greiner^{1,5}, The Pooled Resource Open-Access ALS Clinical Trials Consortium (PRO-ACT)[¶]

1 Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada, **2** Department of Electrical and Computer Engineering, Aarhus University, Aarhus, Denmark, **3** Neuroscience and Mental Health Institute, University of Alberta, Edmonton, Alberta, Canada, **4** Division of Neurology, Department of Medicine, University of Alberta, Edmonton, Alberta, Canada, **5** Alberta Machine Intelligence Institute, University of Alberta, Edmonton, Alberta, Canada

[¶] A complete list of members of the Pooled Resource Open-Access ALS Clinical Trials Consortium (PRO-ACT) can be found in the Acknowledgments.

* clillelund@ualberta.ca



OPEN ACCESS

Citation: Lillelund CM, Kalra S, Greiner R (2025) A meaningful prediction of functional decline in amyotrophic lateral sclerosis based on multi-event survival analysis. PLoS One 20(11): e0336476. <https://doi.org/10.1371/journal.pone.0336476>

Editor: Guanghui Liu, State University of New York at Oswego, UNITED STATES OF AMERICA

Received: December 28, 2024

Accepted: October 27, 2025

Published: November 18, 2025

Copyright: © 2025 Lillelund et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The PRO-ACT dataset can be downloaded from the Neurological Clinical Research Institute's website (<https://ncr1.partners.org/ProACT>). Users

Abstract

Amyotrophic lateral sclerosis (ALS) is a degenerative disorder of the motor neurons that causes progressive paralysis in patients. Current treatment options aim to prolong survival and improve quality of life. However, due to the heterogeneity of the disease, it is often difficult to determine the optimal time for potential therapies or medical interventions. In this study, we propose a novel method to predict the time until a patient with ALS experiences significant functional impairment (ALSFRS-R ≤ 2) for each of five common functions: speaking, swallowing, handwriting, walking, and breathing. We formulate this task as a multi-event survival problem and validate our approach in the PRO-ACT dataset ($N = 3,220$) by training five covariate-based survival models to estimate the probability of each event over the 500 days following the baseline visit. We then predict five event-specific individual survival distributions (ISDs) for a patient, each providing an interpretable estimate of when that event is likely to occur. The results show that covariate-based models are superior to the Kaplan-Meier estimator at predicting time-to-event outcomes in the PRO-ACT dataset. Additionally, our method enables practitioners to make individual counterfactual predictions—where certain covariates can be changed—to estimate their effect on the predicted outcome. In this regard, we find that Riluzole has little or no impact on predicted functional decline. However, for patients with bulbar-onset ALS, our model predicts significantly shorter time-to-event estimates for loss of speech and swallowing function compared to patients with limb-onset ALS (log-rank $p < 0.001$, Bonferroni-adjusted $\alpha = 0.01$). The proposed method can be applied to current clinical examination data to assess the risk of functional decline and thus allow more personalized treatment planning.

must create an account to access the data, but account creation is free of charge.

Funding: Funding received from the Natural Sciences and Engineering Research Council of Canada (NSERC), the Alberta Machine Intelligence Institute (AMII), and the Canadian Institute for Advanced Research (CIFAR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare no competing interests.

Introduction

Amyotrophic lateral sclerosis (ALS) is a neurological disease that causes a gradual loss of upper and lower motor neurons in the central nervous system. Unfortunately, most patients will die within 2–5 years after onset [1], primarily due to complications from respiratory failure [2]. The underlying causes of the disease are not yet understood, and as there are no known cures, current treatments aim only to prolong survival and improve quality of life. Moreover, due to its heterogeneous and unpredictable nature, and the high variability in its progression rate and clinical phenotype [5], it is often difficult to determine the appropriate treatment. This also complicates decisions about the optimal timing of medical interventions, e.g., non-invasive ventilation, or if a specific treatment can effectively slow the progression of the disease. There is therefore a significant need for tools that can predict disease progression to facilitate personalized treatment plans and improve patient care.

The ALS Functional Rating Scale-Revised (ALSFRS-R) [6] is the most widely used questionnaire to evaluate the progression of ALS. It contains 12 questions to assess functional capacity in patients, with a focus on bulbar, motor, and respiratory functions—e.g., swallowing, walking, and breathing, respectively. Each response is scored between 0 and 4, where 0 represents complete inability with regard to that function, and 4 represents normal function. As this questionnaire is administered to patients at multiple time points over the course of the disease, it provides a longitudinal view of functional decline across several areas. Based on this questionnaire, an interesting task is to predict the degree of functional decline a patient experiences during the course of the disease, rather than simply predicting when a patient is likely going to die [7,8]. Many previous projects have therefore attempted to predict the future ALSFRS-R score based on patient covariates [3,9–12]. Ong et al. [12] proposed a binary model to predict the functional decline class (i.e., whether ALS progresses slowly or rapidly in a specific patient) after baseline (trial entry), based on several patient characteristics. Their model was able to predict slow or rapid decline, but did not predict when functional decline would occur over time. Similarly, Jabbar et al. [11] proposed a method to predict disease progression as fast or slow, but as in [12], their model could only give a yes/no answer to whether a patient would experience functional decline over time. Amaral et al. [10] applied several clustering algorithms to clinical patient records, which revealed four separate groups regarding disease progression. Their method could assign newly-diagnosed patients to a progression group, but did not predict when a patient would experience functional decline based on their covariates. Finally, Vieira et al. [9] trained two machine learning models to predict a patient's future ALSFRS-R score: a voice model based on speech recordings and a movement model based on accelerometer measurements. These models showed strong predictive performance, but similar to Amaral et al. [10] and others, they predicted future ALSFRS-R scores rather than estimating the time until functional decline would occur. Moreover, disease progression in ALS is neither constant nor linear, as those studies implicitly assumed, but can vary substantially over time and between individuals [4].

Based on these shortcomings, we propose a novel method that can predict when an ALS patient will experience each of five common types of functional decline, based on the ALSFRS-R protocol (see Table 1): significant difficulty (≤ 2) in each of *Speaking*, *Swallowing*, *Handwriting*, *Walking*, and *Dyspnea*. These five events are distinct but not mutually exclusive and likely share some information in their covariates; it may be that a patient (e.g., Mr. Smith), with limb-onset ALS, will experience difficulty walking and writing earlier than another patient (e.g., Mr. Johnson), who has bulbar-onset ALS. On the other hand, Mr. Johnson may have difficulty speaking and swallowing solid foods before Mr. Smith. As a newly-diagnosed patient with ALS, our approach will help answer important questions, such as:

- How much longer will I be able to speak clearly enough to be understood?
- How much longer will I be able to eat most solid foods without difficulty?
- How much longer will I be able to write a shopping list that my wife can read?
- How much longer will I be able to walk at a normal pace and keep my balance?
- How much longer will I be able to breathe normally during daily activities?

In practice, we formulate the problem as a multi-event survival problem: given an instance (e.g., a description of Mr. Smith at a "baseline" point in time), our method predicts the time until said Mr. Smith will experience each of the events listed above. We present each prediction as an individual survival distribution (ISD) [13], which provide the probability of each event at each future time point, after the baseline visit. We describe each patient by the values of a set of covariates such as age, site of onset, forced vital capacity (FVC), and others, which are recorded in clinical records obtained at or before the baseline visit. To our knowledge, no existing method can explicitly predict the timing of distinct individual events in ALS. Fig 1 shows an outline of the proposed method. The source code is publicly available at: <https://github.com/thecml/FunctionalALS>.

Materials and methods

Clinical data and event annotation

For our empirical analyses, we use the publicly-available Pooled Resources Open-Access Clinical Trials (PRO-ACT) (Data used in the preparation of this article were obtained from the Pooled Resource Open-Access ALS Clinical Trials

Table 1. The definition of the five included functional assessments from the ALS Functional Rating Scale-Revised (ALSFRS-R) [6] and their scores. Each score ranges from 0 to 4, where 0 represents complete inability with regard to the function, and 4 represents normal function. For each patient, our method predicts the time until that patient will experience significant difficulty (ALSFRS-R ≤ 2) for that specific assessment.

Domain	Assessment	Description	Score
Bulbar	Speech	Ability to speak clearly and be understood.	4 = Speaks clearly; 2 = Intelligible with repeating; 0 = Loss of useful speech
Bulbar	Swallowing	Ability to swallow without choking.	4 = Normal eating habits; 2 = Difficulty with some foods; 0 = Needs feeding tube
Fine motor	Handwriting	Ability to write or type.	4 = Normal writing; 2 = Not all words legible; 0 = Unable to grip pen
Gross motor	Walking	Ability to walk without assistance.	4 = Walks normally; 2 = Walks with assistance; 0 = Cannot walk
Respiratory	Dyspnea	Shortness of breath.	4 = None; 2 = Moderate shortness of breath; 0 = Severe shortness of breath

<https://doi.org/10.1371/journal.pone.0336476.t001>

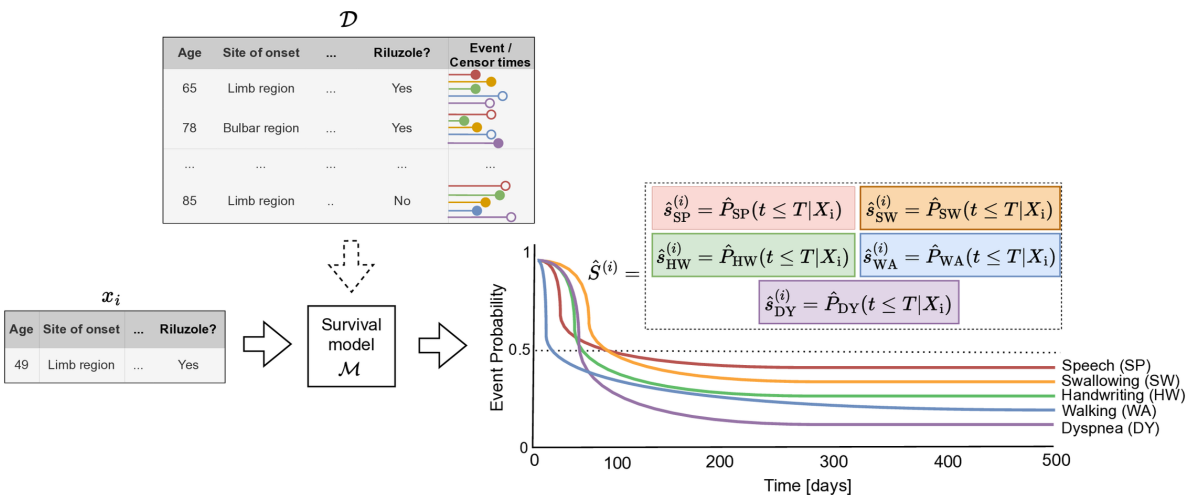


Fig 1. Outline of the proposed method. Historical data about ALS patients form a patient dataset $\mathcal{D}$ with N training instances and d covariates, and the time to event (filled dot) or censoring (hollow dot) from a patient's first visit—the "baseline time". Censoring indicates that the event of interest was not observed for a patient within the study period, so their exact event time remains unknown. We consider five separate but related events, i.e., *Speech*, *Swallowing*, *Handwriting*, *Walking*, and *Dyspnea*. We use the recorded covariates (taken at the baseline time) and event information to train a survival model $\mathcal{M}$ that can accurately estimate the individual survival distribution (ISD) of each of these five events, for a novel patient $\mathbf{x}_i$, denoted as $\hat{S}^{(i)}$. These ISDs give the probability of each of these five events occurring after t days after the baseline visit, for all $t > 0$. They can also be used to estimate the time to event for this $\mathbf{x}_i$ patient, for example, when the survival curve intersects the dashed horizontal line at 50%, which is called the median survival time, for each of the five events.

<https://doi.org/10.1371/journal.pone.0336476.g001>

(PRO-ACT) Database. The data available in the PRO-ACT Database have been volunteered by PRO-ACT Consortium members. The dataset was obtained on October 1st, 2024.) [14] dataset, which is the largest ALS dataset in the world. It includes patient demographics, lab and medical records, as well as family histories, of over 11,600 ALS patients from 23 clinical trials (see Table 2). In each clinical trial, patients follow a schedule of visits until the end of the study or until they become too ill to participate. Covariates recorded at each visit include the total ALSFRS-R score, age, sex, region of onset (e.g., limb, bulbar), the mean FVC score, time since diagnosis, disease progression rate (Defined as $(48 - \text{the total ALSFRS-R score})/(\text{days since diagnosis}/30)$), and whether or not the patient is taking Riluzole. In this study, we only use covariates recorded at the baseline visit for each individual patient.

To define our events of interest, each patient's disease state is assessed by a clinician at every visit using the ALSFRS-R scale. We focus on five specific events: *Speech*, *Swallowing*, *Handwriting*, *Walking*, and *Dyspnea*. To

Table 2. Patient demographics in the PRO-ACT [14] dataset.

Demographic/Data	PRO-ACT (N = 3,220) Pct. or mean ($\pm$ SD)
Age (years)	55.8 (11.6)
Height (cm)	172 (9.5)
Weight (kg)	76.3 (14.6)
Body mass index (BMI)	25.7 (4)
% Female	32.2
% Caucasian	94.6
Site of Onset; % Limb, % Bulbar	43.3, 13.4
Baseline ALSFRS-R Score	39.4 (5)
Time in study (days)	271.4 (122.7)
% Riluzole	75.7

<https://doi.org/10.1371/journal.pone.0336476.t002>

Table 3. Overview of the dataset. “SP”, “SW”, “HA”, “WA” and “DY”, are the *Speech*, *Swallowing*, *Handwriting*, *Walking* and *Dyspnea* events, respectively. Event distribution indicates the percentage of uncensored instances of each type in the dataset. *N*: sample size, *d*: number of covariates, *K*: number of event types, *T*: observed time (of the uncensored instances) for any event in days.

Dataset	N	d	K	T_{min}	T_{max}	T_{mean}	Event distribution (%):
PRO-ACT	3,220	8	5	1	498	130	SP: 37.8, SW: 31.9 HA: 50.3, WA: 60.6 DY: 27.0

<https://doi.org/10.1371/journal.pone.0336476.t003>

annotate our dataset, we define an event as having occurred the first time a patient scores 2 or fewer in that specific assessment during any follow-up visit following their baseline visit. These five events can occur in any order and are not mutually exclusive. There are several types of censoring, including death, being unable to perform the assessment or leaving the study prematurely. Censoring thus represents incomplete event information, where the true event time is only known to exceed the observed follow-up time. We use a maximum follow-up time of 500 days to focus on the early stages of the disease and because data become sparse with high censoring afterward. We consider patients who score 2 or fewer on any of the five assessments at baseline as having already had the event, indicating that the loss of function occurred before the start of the trial, and exclude these patients from the study. We also exclude patients with no recorded ALSFRS-R history. After applying these criteria and accounting for censoring, the final dataset comprises 3,220 patients with observed or right-censored event times. The code for loading and annotating the data is available in the source code repository. Table 3 provides an overview of the dataset. Fig 2 shows the distribution of event and censoring times. A list of covariates is available in the Supplement (S1 Table).

Survival analysis and notation

Survival analysis models the time until some event of interest occurs. This event can either be observed or censored—the latter typically happens if the patient is still alive at the end of the study or decides to drop out [15, Ch. 11]. Survival analysis has become an important tool for understanding diseases and predicting outcomes, which helps medical researchers evaluate the significance of prognostic factors in applications like comatose after cardiac arrest [16], liver transplantation [17] or fall risk [18]. In ALS, survival analysis has been used mainly to predict risk scores with respect to death [4,7,19].

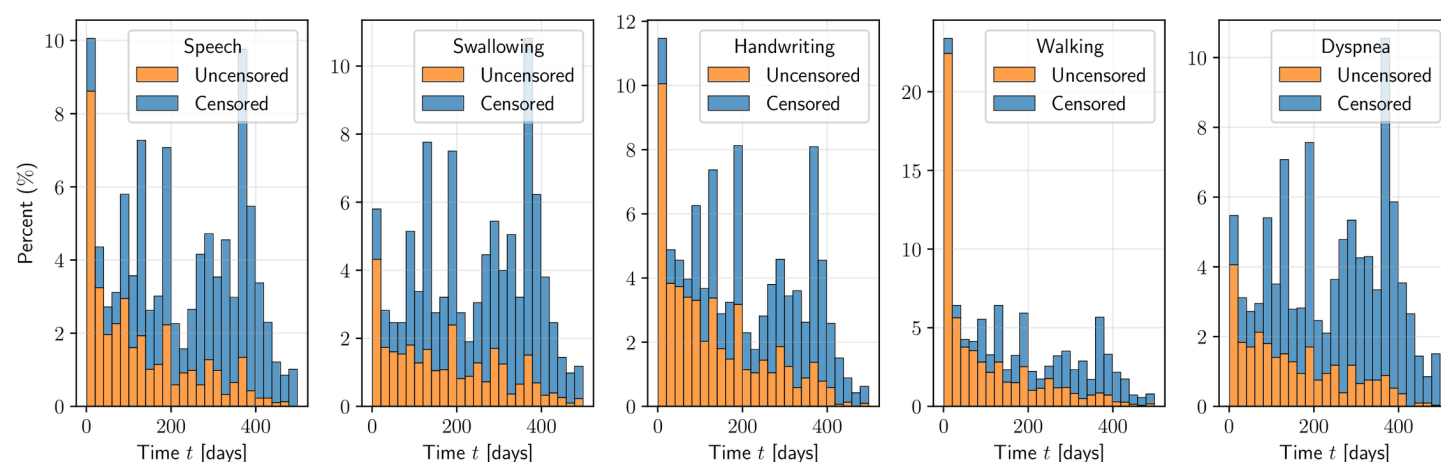


Fig 2. Distribution of uncensored and censored times in the PRO-ACT dataset for the five events.

<https://doi.org/10.1371/journal.pone.0336476.g002>

We briefly introduce the notation: Let $\mathcal{D} = \{(\mathbf{x}^{(i)}, \mathbf{t}^{(i)}, \delta^{(i)})\}_{i=1}^N$ be the dataset, where for each i , $\mathbf{x}^{(i)} \in \mathbb{R}^d$ is a d -dimensional vector of covariates. We consider discrete times, where $\mathbf{t}^{(i)} \in \{1, 2, \dots, T_{\max}\}^K$ is a vector of observed times for the i -th patient for K different events, each at one of these $T_{\max}$ possible times, and $\delta^{(i)} \in \{0, 1\}^K$ is a vector of event indicators for each event. Moreover, let $e_k^{(i)} \in \{1, 2, \dots, T_{\max}\}$ denote the event time and let $c_k^{(i)} \in \{1, 2, \dots, T_{\max}\}$ denote the censoring time for the i -th instance for event k over the event horizon, thus we have $t_k^{(i)} = e_k^{(i)}$ if $\delta_k^{(i)} = 1$, otherwise, $t_k^{(i)} = c_k^{(i)}$ if $\delta_k^{(i)} = 0$. A survival model predicts the probability that some event k occurs at time T later than t , i.e., the so-called survival or event probability, which is denoted $S(t) = \Pr(T > t) = 1 - \Pr(t \leq T)$. In multi-event survival analysis, we want to predict the probability of different events occurring over time. For each patient i , this results in a matrix of survival functions, $\mathbf{S}^{(i)} = \{\mathbf{s}_1^{(i)}, \mathbf{s}_2^{(i)}, \dots, \mathbf{s}_K^{(i)}\}$. Here, $\mathbf{s}_k^{(i)}$ represents the probabilities that patient i does not experience event k up to each time point t , i.e., $\mathbf{s}_k^{(i)} = (P(e_k^{(i)} > 1), P(e_k^{(i)} > 2), \dots, P(e_k^{(i)} > T_{\max}))$. See Fig 1 for a visual overview.

Experiments and results

Setup and preprocessing

We follow Sechidis et al. [20] and first split the dataset into a training, validation and test set by 70%, 10%, and 20% using a stratified procedure. This ensures that the event times and censoring rates are consistent across the three sets and the K events. The stratification procedure uses a random seed (0-9). After splitting the data, we impute missing values using the sample mean for real-valued covariates or the mode for categorical covariates based on the training set only. A list of missing values is available in the Supplement (S2 Table). After imputing missing values, we encode categorical covariates using a one-hot encoding strategy and apply a z-score data normalization to speed up the training process and improve stability. No outlier detection was performed.

Our evaluation procedure consists of training five survival models on the training set and evaluating them on the holdout test set. For evaluation fairness, we do not perform hyperparameter optimization for specific models, but instead configure all models with sensible default parameters. A list of hyperparameters is available in the Supplement (S3 Table). For deep learning models, we use a single hidden layer with 32 nodes. For methods that support it, we use early stopping during training to terminate the process if the validation loss does not improve for 10 consecutive epochs. All data preprocessing and experiments were implemented in Python 3.9. A list of software packages used is available in the source code repository.

Survival models

To evaluate both single-event and multi-event survival modeling approaches, we distinguish between models that predict the time to a single event and those that jointly model multiple event types. As single-event models, we implement the traditional Cox Proportional Hazards (CoxPH) model using a linear estimator [21], Random Survival Forests (RSF) [22], DeepSurv [23] and MTLR [24]. As a multi-event model, we implement the Multi-Event Network for Survival Analysis (MENSA) model [25]. To explain these models: **CoxPH** is a popular semiparametric method to fit a regression model to survival data. By adopting a multiplicative form for the contribution of several covariates to each individual's event time, the CoxPH model is a powerful yet simple tool for assessing the simultaneous effect that different covariates have on the event times. CoxPH assumes proportional hazards, which means that the effect of covariates on the hazard function is constant over time. **RSF** extends decision trees for survival analysis and recursively splits training data based on some survival-specific criterion, e.g., log-rank splitting, with the goal of maximizing separation between event times between nodes. RSF can model nonlinear relationships between the covariates and the event, and does not assume proportional hazards. **DeepSurv** is a multilayer perceptron (MLP) based on the CoxPH model, where the risk score is a nonlinear function of the covariates. DeepSurv assumes proportional hazards. **MTLR** is a discrete survival model that estimates the individual survival function as a multinomial logistic regression model and does not assume proportional hazards. Lastly,

MENSA is an MLP that learns the individual survival distribution for K events as a mixture of Weibull distributions with a shared covariate layer to model dependencies between events. MENSA does not assume proportional hazards.

Evaluation metrics

To comprehensively assess the survival models, we select a range of evaluation metrics that capture three key aspects of model performance: discrimination, absolute/squared error, and calibration. Discrimination, usually measured by the concordance index (CI), is relevant when we want to identify which patients are likely to experience functional decline earlier versus later at the group level. Squared or absolute error metrics, like the mean absolute error (MAE), are used to assess how accurately the model predicts the time of functional impairment for individual patients. Finally, calibration measures how well the predicted probabilities match the observed frequencies of events at specific time points. For example, if a model predicts that a patient has an 80% chance of not experiencing significant functional decline before 100 days, then among 100 similar patients, about 80 should not experience significant functional decline by that time. We use the following metrics to assess discrimination, absolute/squared error, and calibration:

CI: Harrell's CI [26] measures discrimination performance by calculating the proportion of concordant pairs among all comparable pairs given predicted risk scores. A pair is considered comparable if the event order can be determined. Specifically, for a comparable pair $\{i, j\}$ with event times $e_i < e_j$ and $\delta_i = 1$ (indicating that i experienced the event), we calculate risk scores $\hat{\eta}_i$ and $\hat{\eta}_j$ as the negative median survival times. The median survival time is the time point at which the estimated survival probability drops to 0.5. If the predicted risk score for j is greater than that for i at the time i experienced the event (while j remains event-free), the pair is concordant. A CI of 0.5 indicates chance-level ranking. We also report Uno's CI [27], which is a CI that uses a technique to correct for the bias introduced by censoring.

BS/IBS: The Brier Score (BS) calculates the squared error between the predicted probability of the event and the Heaviside step function of the observed event. The integrated Brier score [28] (IBS) aggregates Brier scores over multiple time points to provide a single measure of the squared error.

MAE: The MAE calculates the average absolute difference between the true event time and the predicted event time. Given an ISD, $S(t | \mathbf{x}_i) = \Pr(T > t | \mathbf{x}_i)$, we calculate the predicted event time $\hat{t}_i$ as the median survival time [29]. Since our dataset contains censoring, we calculate the margin MAE (mMAE) as proposed by [13], which assigns a "best-guess" estimate to each censored patients using the Kaplan-Meier (KM) [30] estimator.

D-calibration: Distribution calibration [13], also known as D-Cal, measures how well the predicted survival function, $\hat{S}(t)$, is calibrated for each event. D-calibration assesses this using a Pearson's chi-squared test with $\alpha = 0.05$. For any probability interval $[a, b] \in [0, 1]$, we define $D_m(a, b)$ as the group of patients in the dataset D whose predicted probability of an event is in the interval $[a, b]$ [31]. A model is D-calibrated if the proportion of patients $|D_m(a, b)|/|D|$ is statistically similar to the amount $b - a$. Intuitively, if a model is well-calibrated, the predicted probabilities match observed event frequencies—for instance, about 20% of patients predicted to have a 0.2 event probability should actually experience the event. In our experiments, we report the number of times each model is D-calibrated (at $p > 0.05$) for each event across 10 experiments.

Model results

We present our empirical results (Table 4 and Fig 3) in terms of the aforementioned aspects with respect to model performance: discrimination, absolute/squared error, and calibration. Below, we discuss the results and reflect on their implications.

Discrimination: Good discrimination performance is important if we need to rank patients by their disease severity to prioritize access to limited clinical resources, for example, portable ventilators. Across all models, we see the highest Harrell's and Uno's C-indices for the *Speech* (approx. 74–75% and 70–71%, respectively) and *Swallowing* (approx. 75–76% and 70–71%, respectively) events. On the other hand, functional decline with respect to limb functions is harder

Table 4. Mean ($\pm$ SD.) prediction performance, averaged over 10 independent experiments. D-calibration counts the number of times the respective model was D-calibrated. Harrell's CI, Uno's CI and IBS results are multiplied by 100.

Model	Event	Harrell's CI $\uparrow$	Uno's CI $\uparrow$	IBS $\downarrow$	mMAE $\downarrow$	D-Cal $\uparrow$
CoxPH [21]	Speech	74.52 $\pm$ 1.39	70.44 $\pm$ 3.34	14.41 $\pm$ 0.43	206.85 $\pm$ 9.88	(10/10)
	Swallowing	75.75 $\pm$ 1.53	70.96 $\pm$ 4.58	12.58 $\pm$ 0.72	146.15 $\pm$ 25.50	(10/10)
	Handwriting	67.57 $\pm$ 0.92	63.75 $\pm$ 2.29	16.73 $\pm$ 0.70	141.85 $\pm$ 7.40	(10/10)
	Walking	68.67 $\pm$ 1.20	67.15 $\pm$ 1.40	16.60 $\pm$ 0.50	145.04 $\pm$ 4.50	(7/10)
	Dyspnea	68.54 $\pm$ 1.71	66.99 $\pm$ 6.30	13.79 $\pm$ 0.41	235.46 $\pm$ 27.57	(10/10)
RSF [22]	Speech	75.05 $\pm$ 1.13	70.08 $\pm$ 2.75	14.69 $\pm$ 0.44	164.51 $\pm$ 19.12	(10/10)
	Swallowing	75.53 $\pm$ 1.46	69.65 $\pm$ 2.98	13.02 $\pm$ 0.74	111.71 $\pm$ 11.46	(10/10)
	Handwriting	67.38 $\pm$ 1.39	63.70 $\pm$ 2.05	16.87 $\pm$ 0.57	146.97 $\pm$ 6.88	(10/10)
	Walking	68.44 $\pm$ 1.47	66.68 $\pm$ 1.54	16.78 $\pm$ 0.56	145.98 $\pm$ 4.63	(10/10)
	Dyspnea	78.23 $\pm$ 1.07	78.95 $\pm$ 4.01	11.89 $\pm$ 0.33	850.45 $\pm$ 471.39	(10/10)
DeepSurv [23]	Speech	74.18 $\pm$ 3.05	70.77 $\pm$ 3.57	14.36 $\pm$ 0.89	232.15 $\pm$ 28.48	(10/10)
	Swallowing	75.61 $\pm$ 1.37	70.67 $\pm$ 5.52	12.48 $\pm$ 0.65	178.77 $\pm$ 39.68	(10/10)
	Handwriting	67.12 $\pm$ 1.08	63.72 $\pm$ 2.15	16.81 $\pm$ 0.84	151.83 $\pm$ 18.59	(9/10)
	Walking	69.53 $\pm$ 0.93	68.44 $\pm$ 1.11	16.23 $\pm$ 0.51	148.17 $\pm$ 6.60	(8/10)
	Dyspnea	71.72 $\pm$ 3.99	70.61 $\pm$ 7.60	13.11 $\pm$ 0.61	384.96 $\pm$ 284.17	(10/10)
MTLR [24]	Speech	74.22 $\pm$ 1.67	70.17 $\pm$ 3.23	13.95 $\pm$ 0.44	179.35 $\pm$ 5.34	(10/10)
	Swallowing	75.13 $\pm$ 1.43	70.63 $\pm$ 5.41	12.34 $\pm$ 0.59	122.93 $\pm$ 19.52	(10/10)
	Handwriting	67.32 $\pm$ 1.11	63.71 $\pm$ 2.42	16.72 $\pm$ 0.66	140.94 $\pm$ 6.29	(9/10)
	Walking	68.97 $\pm$ 1.16	67.71 $\pm$ 1.06	16.30 $\pm$ 0.51	149.93 $\pm$ 10.99	(9/10)
	Dyspnea	66.54 $\pm$ 2.22	65.90 $\pm$ 6.85	13.50 $\pm$ 0.50	203.36 $\pm$ 24.57	(10/10)
MENSA [25]	Speech	74.21 $\pm$ 1.11	70.56 $\pm$ 3.47	14.53 $\pm$ 0.41	181.81 $\pm$ 7.92	(9/10)
	Swallowing	74.94 $\pm$ 1.36	70.85 $\pm$ 5.20	12.61 $\pm$ 0.56	119.16 $\pm$ 7.12	(10/10)
	Handwriting	66.46 $\pm$ 0.90	63.80 $\pm$ 2.65	17.10 $\pm$ 0.66	141.97 $\pm$ 5.70	(9/10)
	Walking	68.63 $\pm$ 1.29	66.94 $\pm$ 1.50	17.04 $\pm$ 0.61	149.50 $\pm$ 6.49	(6/10)
	Dyspnea	66.56 $\pm$ 1.77	64.40 $\pm$ 7.07	13.78 $\pm$ 0.30	189.43 $\pm$ 24.28	(10/10)

<https://doi.org/10.1371/journal.pone.0336476.t004>

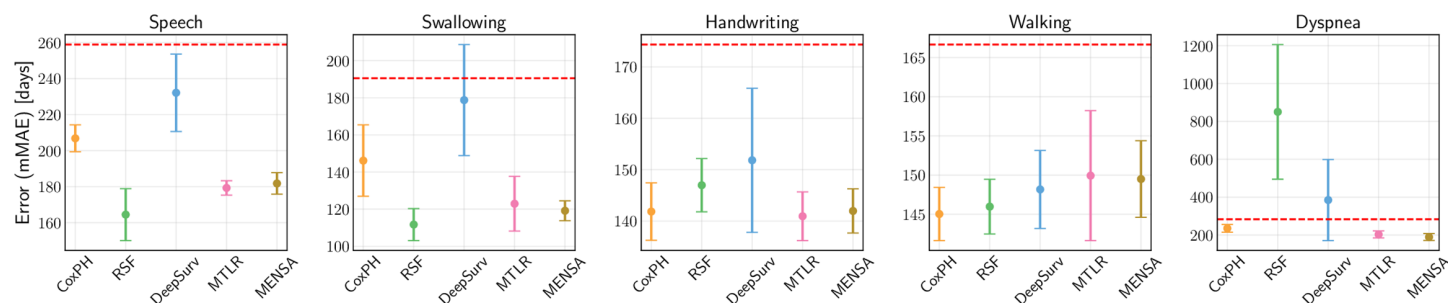


Fig 3. The mMAE (in days) as a function of covariate-based models (x-axis) in the PRO-ACT test set, with error bars representing empirical 95% confidence intervals. The horizontal dashed line is the KM estimator. Lower is better.

<https://doi.org/10.1371/journal.pone.0336476.g003>

to discriminate, with lower C-indices (approx. 67–69%) for the *Handwriting* and *Walking* events. Although the *Dyspnea* event typically has moderate CI values (approx. 66–68%), RSF achieves a Harrell's CI of 78.23% and a Uno's CI of 78.95% for this event, making it the best model in this case. DeepSurv and MTLR also demonstrate consistently good discrimination performance across events, and MENSA matches or slightly exceeds these for the *Speech* and *Swallowing* events, though its performance dips slightly for the *Walking* and *Dyspnea* events.

Absolute/squared error: Accurate time-to-event estimates or survival probabilities are important when a decision or intervention depends on the precise timing of the event. Across all models, the *Swallowing* event consistently exhibits some of the lowest IBS (approx. 12–13) and mMAE (approx. 110–150) numbers, meaning that this event is the easiest

to predict the timing of. For the *Dyspnea* event, the IBS is also moderate (11.9–13.8), but the RSF model is notable with a low IBS but a relatively high mMAE (approx. 850±471 days). In contrast, other models produce much more moderate mMAE numbers for this event (approx. 200–400 days), as seen with the MENSA model. This highlights an important point: simply calculating the squared error of the predicted survival probabilities does not fully capture a model's ability to predict the timing of an event. MENSA achieves reasonable mMAE numbers (approx. 120–190 days) in all cases, except for the *Speech* event (181±8 days), however, and tends to perform particularly well for the *Swallowing* event, similar to MTLR.

Fig 3 shows the prediction error in days between the population-level KM estimator [30] and the covariate-based models. The KM estimate is calculated per event on observed and censored event times, without considering individual differences (i.e., covariates). In all five events, the covariate-based models generally give superior time-to-event estimates compared to the KM method, with some variation between models. This suggests that incorporating patient-specific information can lead to more accurate time-to-event predictions when predicting functional decline, rather than simply relying on a population-level estimate.

Calibration: Good calibration performance is important when we have to use the actual predicted probabilities to guide our decision-making. For example, underestimating the risk of respiratory decline or bulbar involvement may result in mistimed interventions or suboptimal treatment choices. Across all models, the *Walking* event consistently poses a calibration challenge, as predicted event probabilities generally deviate from the observed event frequencies here (i.e., the models are sometimes not calibrated). However, RSF achieves perfect D-calibration (10/10) across all events and experiments, demonstrating robust and consistent calibration performance. DeepSurv and MTLR also perform well, with most experiments achieving D-calibration. MENSA achieves perfect calibration for the *Swallowing* and *Dyspnea* (10/10) events, but shows slightly poorer calibration for the *Speech* (9/10) and *Handwriting* (9/10) events, and especially the *Walking* event (only 6/10).

Individual disease predictions

Our method can predict a patient's disease trajectory as K individual survival distributions, where K is the number of events (here, $K = 5$). Fig 4 shows such ISDs, from baseline ($t = 0$) to 500 days later, for the i -th patient. As an example, let us return to our patient, Mr. Smith. He is a 72-year-old male, who had been diagnosed with limb-onset ALS prior to study entry and had experienced difficulty speaking and weakness in his legs, but all five of his scores were above 2, so that he could participate in the study. Mr. Smith's covariates are available in the Supplement (S4 Table). The neurologist predicted that talking and walking difficulties would be among Mr. Smith's earliest functional impairments, and our method

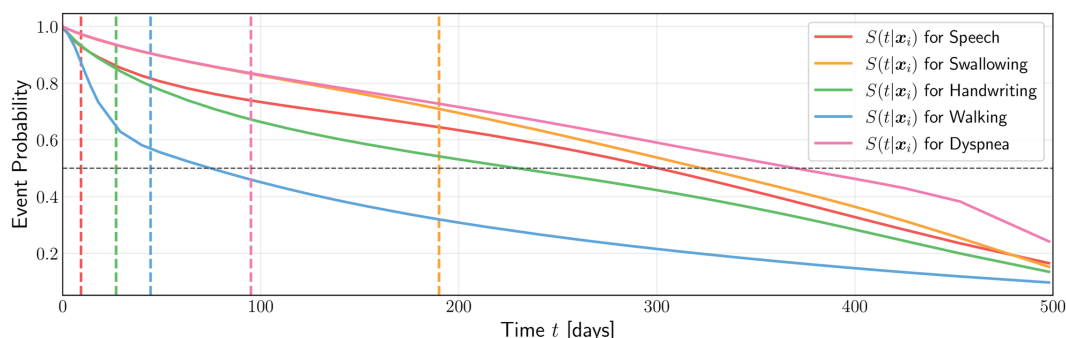


Fig 4. Predicted ISDs for Mr. Smith (i). The point where the survival curve intersects the dashed horizontal line at 50% indicates the predicted time of event—so this predicts, for example, *Walking* at approximately 80 days, *Handwriting* at 240 days, and *Speech* at 300 days. The dashed vertical lines are Mr. Smith's actual time of event for the respective events—so, for example, *Speech* at approximately 7 days, *Handwriting* at 35 days, and *Walking* at 45 days.

<https://doi.org/10.1371/journal.pone.0336476.g004>

confirmed her suspicion. From the outset, Mr. Smith was provided with a walking aid and a personalized long-term care plan that included therapy, support systems, and home modifications. Shortly after his hospital visit, he had to repeat many of his words to be understood and relied on assistance from a cane or his wife to walk. Meanwhile, handwriting also became challenging for him, but he did not experience shortness of breath or difficulty swallowing whole foods until much later.

Individual counterfactual predictions

Our method can also make individual counterfactual predictions [32], that is, alternative scenarios in which certain patient covariates are changed to see their (predicted) effect on the probability of the event. Such predictions can give novel insights into disease mechanisms and assess treatment options on an individual level. In the following figures, we have used MENSA [25] to make individual counterfactual predictions based on our PRO-ACT dataset. These provide insights into the relationship between common disease predictors and functional decline, and ask how changing certain covariates affects disease outcomes. Again, we use our patient, Mr. Smith, as the instance i . Additionally, we separate all patients in the test set into groups by the same covariate we used for the individual patient and employ the log-rank test [33] to measure the statistical difference between the groups' event times. We use a Bonferroni correction to correct for multiple comparisons by dividing the significance level ($\alpha = 0.05$) by the number of comparisons ($n = 5$), i.e., the significance level is then $\alpha_{adjusted} = 0.05/5 = 0.01$. For the remainder of this section, we will discuss our counterfactual predictions.

As a starting point, many newly diagnosed patients with ALS need to decide whether to take certain pharmaceutical drugs. Currently, there are three approved drugs by the US Food and Drug Administration (FDA) to treat ALS, the most common being Riluzole. Riluzole works by blocking the release of glutamate, which delays the onset of ventilator dependence or tracheostomy in some people and may increase overall survival by two to three months [34]. However, the drug can cause significant side effects, including nausea, anorexia, and diarrhea [4]. Therefore, it is important to determine whether the drug will be effective—i.e., significantly delay functional decline in this regard—for each individual, in order to decide whether that person should receive the treatment. Upon diagnosis, Mr. Smith decided not to take Riluzole, but what if he instead had decided to take it. Fig 5 shows the predicted ISDs for each of the five events using Mr. Smith's covariates ($\mathbf{x}_i$), with the Riluzole covariate being yes (resp., no) at the baseline visit. The two survival curves have nearly the same slopes, with Riluzole only being marginally above non-Riluzole, and nearly identical estimated event times. There are no statistical differences in event times between the groups either ($p > 0.01$). As an example, for the *Speech* event, this model predicts that taking Riluzole would delay the onset of speaking difficulties by approximately 20 days.

In ALS, the site of onset refers to the specific part of the body where the initial symptoms began. ALS can affect different regions, and the site of onset can influence the progression of the disease. The two most common sites of onset are bulbar-onset and limb-onset, the former being associated with faster disease progression in general [11]. Mr. Smith

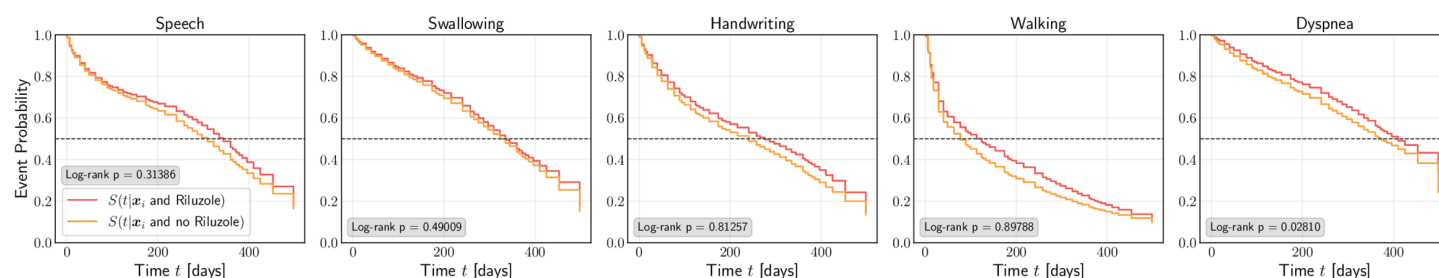


Fig 5. Predicted ISDs for Mr. Smith based on his use of Riluzole.

<https://doi.org/10.1371/journal.pone.0336476.g005>

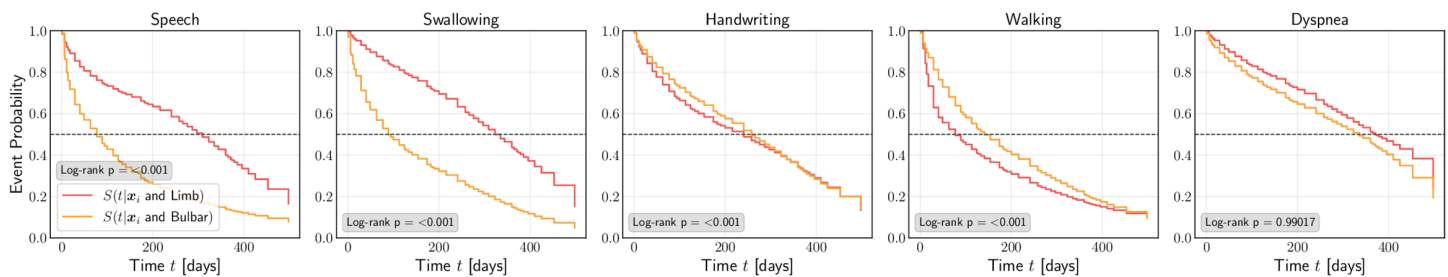


Fig 6. Predicted ISDs for Mr. Smith based on whether he has limb-onset or bulbar-onset.

<https://doi.org/10.1371/journal.pone.0336476.g006>

was diagnosed with limb-onset ALS, but what if instead he had been diagnosed with bulbar-onset ALS. Fig 6 shows the predicted ISDs for Mr. Smith based on whether he has limb-onset or bulbar-onset. We see significant differences in the predictions for the *Speech* and *Swallowing* events, indicating a worse prognosis for Mr. Smith if he has bulbar-onset, but more agreement between the curves for the *Handwriting*, *Walking* and *Dyspnea* events.

Next, forced Vital Capacity (FVC) is a measure of lung function and represents the total amount of air a person can forcibly exhale after taking a deep breath. It is commonly used to assess lung respiratory function and is measured in liters. In ALS, FVC is a key clinical assessment for evaluating respiratory muscle strength, which progressively weakens as the disease progresses. Mr. Smith's total FVC score at baseline was 2.69. Again, we ask the counterfactual question of what if instead he had a higher FVC or a lower FVC. Fig 7 shows the predicted ISDs for Mr. Smith based on whether he has a high or low FVC. As expected, the *Handwriting* event shows only a weak relationship with FVC, as evidenced by the high p -value and the alignment of the two survival curves. In contrast, the *Swallowing* event exhibits a significant relationship with FVC, and the model provides substantially different time-to-event predictions based solely on this covariate.

Lastly, we consider the total ALSFRS-R score at baseline and its effect on predicted functional decline. This score provides a snapshot of Mr. Smith's initial physical capabilities when he is relatively healthy before the disease has advanced notably. Mr. Smith's total ALSFRS-R score at baseline was 37. Again, we make a counterfactual prediction if instead Mr. Smith had a significantly higher or lower ALSFRS-R score. Fig 8 shows the predicted ISDs for Mr. Smith based on whether he has a high or low ALSFRS-R score. Consistent with the literature [4], if Mr. Smith had a high ALSFRS-R score to begin with, he would (presumably) experience functional decline much later than if he had a low ALSFRS-R score. This is indicated by the discrepancy between the survival curves and the low p -values across all the events. Obviously, initial physical ability has a strong relationship with loss of function after a certain time, and patients may be at different stages of the disease and have different levels of strength when they enter the trial. In addition, research in ALS has suggested the presence of various disease phenotypes with different progression rates [35].

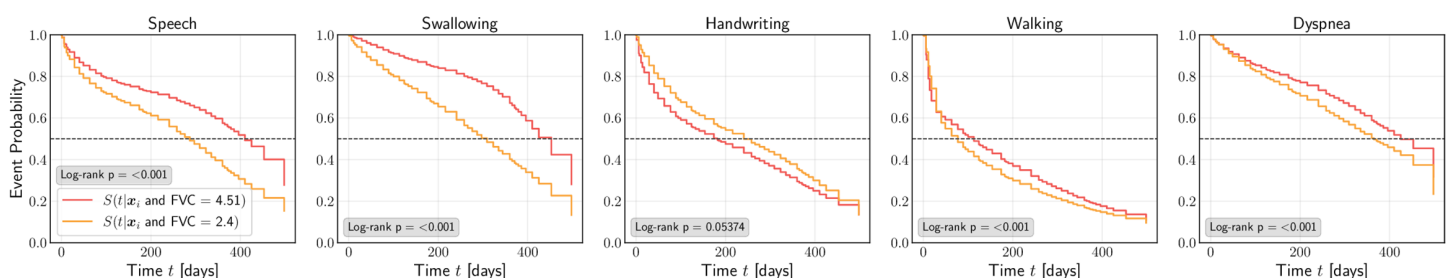


Fig 7. Predicted ISDs for Mr. Smith based on his FVC score being high or low.

<https://doi.org/10.1371/journal.pone.0336476.g007>

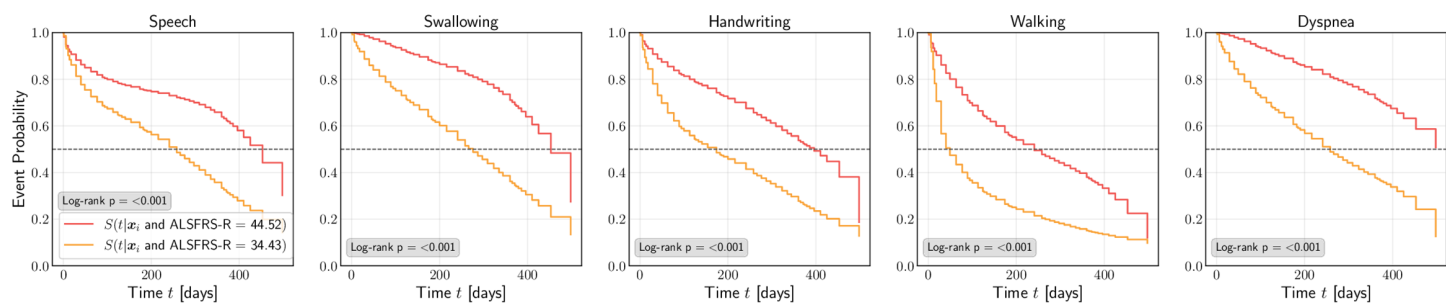


Fig 8. Predicted ISDs for Mr. Smith based on whether his initial ALSFRS-R score is high or low.

<https://doi.org/10.1371/journal.pone.0336476.g008>

Discussion

Contribution

This research has three goals: (1) determine whether covariate-based survival models can predict the time to individual functional decline ($\text{ALSFRS-R} \leq 2$) in ALS patients more accurately than population estimators, (2) precisely estimate risk scores and event times, and (3) explore how counterfactual scenarios affect the probability of functional decline over the course of the disease.

For the first goal, individualized predictions are crucial for tailoring healthcare and therapeutic decisions to the individual patient. Models that estimate an individual's risk rather than a group average enable us to answer important questions from the patient's point of view such as: how much longer will I be able to speak clearly enough to be understood? We make a novel contribution in this regard, as we empirically show that most covariate-based models, even simple linear models, outperform the KM estimator (which does not use covariates) in predicting the time to functional decline. Given that ALS affects multiple functional systems (bulbar, motor, respiratory) at varying rates, our analysis also shows clear advantages of adopting non-proportional hazards models, such as Random Survival Forests and MENSA. Unlike other methods for predicting functional decline in ALS [3,9–12], our method uniquely provides patients and clinicians with an actual time (in days) when functional decline is likely to occur.

For the second goal, we rigorously evaluate five survival models across three aspects of model performance: discrimination, absolute/squared error, and calibration. The main takeaway is that some types of functional decline in ALS patients are easier to predict than others. For example, the *Speech* event has an average Harrell's CI of approximately 74% across the five models—the highest among all events—but also an average mMAE (a variant of the MAE that supports censoring) of 192.4 days, the second-highest. This means that even if a model is good at ranking patients by their relative risk (independent of time) for some event, it cannot necessarily predict accurately when the event will occur. In contrast, the *Walking* event has a lower average CI of approximately 67%, but an average mMAE of only 147.7 days. This suggests that, although it is harder to rank patients by risk for this event compared to *Speech*, it is easier, by contrast, to predict when the event will occur. This leads to the realization that strong ranking performance does not necessarily translate into accurate time-to-event predictions.

Finally, for the third goal, we use our method to make counterfactual predictions, investigating how modifying specific covariates affects predicted disease outcomes for individual patients. To our knowledge, no prior work has explored this use case in ALS. However, such predictions are highly relevant for medical professionals who must make important treatment decisions after first meeting a patient. For example, consider our patient, Mr. Smith: based on his covariates, there were no significant differences in the predicted survival probabilities (i.e., the risk of functional decline) should he decide to take Riluzole, a drug used to treat ALS. In other words, whether he decided to take Riluzole or not, this decision would not significantly change his predicted risk of functional decline over the next 500 days from the first time he saw the

neurologist. We believe that these individualized predictions can be very beneficial for treatment planning and clinical decision-making.

Application

Our method has several important applications. First, it can help medical doctors make informed decisions by tailoring treatments and interventions to the individual's predicted disease trajectory. This is crucial in ALS, where timely interventions, such as assistive devices and nutritional or respiratory support, can significantly impact a patient's quality of life. Second, patients and their families are often faced with difficult decisions regarding treatment options, lifestyle adjustments, and end-of-life planning. By providing a clear, individualized prognosis, clinicians can help patients make more informed decisions about their care, improve their quality of life, and give them a sense of control over their treatment. As part of the evaluation, we show how our method can make counterfactual predictions, answering the what-if questions: Mr. Smith was diagnosed with limb-onset ALS and the model predicts a specific path of functional decline for him, but what would have happened if he had been diagnosed with bulbar-onset ALS instead. Third, personalized predictions can also help design and evaluate clinical trials. Understanding the likely timing of functional decline in patients with ALS can help researchers identify suitable candidates for clinical studies, ensuring that the trials are conducted with participants at the appropriate stage of disease. This can lead to more effective testing of potential therapies and interventions that could slow or stop disease progression.

Concretely, if practitioners wish to implement our method in everyday clinical practice, we recommend the following steps:

1. Integrate model training with existing electronic health record (EHR) systems or clinical decision support tools to automatically retrieve patient data (covariates) and past ALSFRS-R scores (event information). After obtaining the relevant data, they should train the model, evaluate it, deploy it, and expose an application programming interface (API) that allows clinicians to make predictions on new patients.
2. Develop a user-friendly software interface or web-based application that allows clinicians to easily input or access patient information and obtain individual predictions of functional decline over time. When a new patient is diagnosed with ALS and has their first visit at the hospital, clinicians can use the application to enter the patient's covariates and obtain a prediction over the next 500 days.
3. Update, retrain, and deploy the model regularly with new patient data to maintain accuracy and incorporate new biomarkers or additional clinical variables, if such become available. This can be done using Machine Learning Operations (MLOps), which is a set of practices that automate and simplify machine learning workflows and deployments.

Limitations

All survival models evaluated in this work assume conditional independent censoring, i.e., that the event time is independent of the censoring time, given the patient's covariates. For example, patients who leave the study early do so for reasons unrelated to the event of interest or because of poor lung function captured by the forced vital capacity covariate. However, ALS is a heterogeneous disease that can quickly deteriorate and no study can capture all intrinsic disease mechanisms. Generally, clinical trials in ALS experience high dropout rates [36], which is also evident in our study through the high censoring rates shown in Table 3 and Fig 2. It is likely that this phenomenon is related to disease progression and therefore functional decline, but since we do not observe functional decline in patients who leave the study, this kind of censoring may violate the assumption of conditional independent censoring and bias the results. According to Atassi et al. [36], the three most common reasons for attrition (i.e., dropout) in ALS trials were withdrawal of consent for no specific reason (37%), death (28%), and withdrawal secondary to an adverse event (17%).

Future work

Our method has several avenues for future research. First, addressing the aforementioned dependency between the event and censoring times is relevant, given the relatively few covariates provided in the PRO-ACT dataset and because all evaluated survival models assume that censoring is conditionally independent. To address this issue, when estimating model parameters, we can learn a so-called copula function [37]. This is a link function that describes the dependency structure between the event and censoring distributions. Copulas have recently gained attention in survival analysis due to their flexibility in capturing complex dependency structures [38]. Second, it is highly relevant to include covariates captured using magnetic resonance imaging (MRI) into the predictive models, since such features can explain early signs of neurodegeneration, for example, cerebral atrophy in the motor cortex and corticospinal tract [39]. At present, such features have only been used to predict mortality or disease aggressiveness in patients with ALS [19,39,40], but not to predict functional decline. Thus, this presents an exciting avenue for future research.

Conclusion

This article proposes a novel and interpretable method for estimating the time to functional decline in ALS patients. We formulate this task as a multi-event survival problem and validate our method in the PRO-ACT dataset by training five covariate-based survival models to predict significant functional impairment with respect to five common functions. The proposed method has the following advantages over state-of-the-art approaches: (1) It explicitly predicts the time until functional decline is likely to occur for an individual patient. (2) It gives superior time-to-event estimates over commonly used population estimators, such as the KM estimator. (3) It allows for individual counterfactual predictions, where certain covariates can be changed to see their effect on the predicted outcome. In conclusion, we recommend using the proposed method on patient records in clinical settings, subject to additional rigorous validation, to assess the risk of imminent functional decline and enable more personalized treatment planning.

Supporting information

S1 Table. List of covariates.

(DOCX)

S2 Table. List of missing data rows.

(DOCX)

S3 Table. List of model hyperparameters.

(DOCX)

S4 Table. Mr. Smith's covariates.

(DOCX)

Acknowledgments

Data used in the preparation of this article were obtained from the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) Database. The following organizations and individuals within the PRO-ACT Consortium contributed to the design and implementation of the PRO-ACT Database and/or provided data, but did not participate in the analysis of the data or the writing of this report: ALS Therapy Alliance, Cytokinetics, Inc., Amylyx Pharmaceuticals, Inc., Knopp Biosciences, Neuraltus Pharmaceuticals, Inc., Neurological Clinical Research Institute (MGH), Northeast ALS Consortium, Novartis, Orion Corporation, Prize4Life Israel, Regeneron Pharmaceuticals, Inc., Sanofi, Teva Pharmaceutical Industries, Ltd., and The ALS Association.

Author contributions

Conceptualization: Christian Marius Lillelund, Sanjay Kalra, Russell Greiner.

Data curation: Christian Marius Lillelund.

Formal analysis: Christian Marius Lillelund.

Funding acquisition: Russell Greiner.

Investigation: Christian Marius Lillelund, Sanjay Kalra, Russell Greiner.

Methodology: Christian Marius Lillelund, Sanjay Kalra, Russell Greiner.

Project administration: Russell Greiner.

Resources: Christian Marius Lillelund, Sanjay Kalra.

Software: Christian Marius Lillelund.

Supervision: Sanjay Kalra, Russell Greiner.

Validation: Christian Marius Lillelund, Sanjay Kalra, Russell Greiner.

Visualization: Christian Marius Lillelund.

Writing – original draft: Christian Marius Lillelund.

Writing – review & editing: Christian Marius Lillelund, Sanjay Kalra, Russell Greiner.

References

- Morris J. Amyotrophic Lateral Sclerosis (ALS) and related motor neuron diseases: an overview. *Neurodiagn J*. 2015;55(3):180–94. <https://doi.org/10.1080/21646821.2015.1075181> PMID: 26630810
- Talbot K. Motor neuron disease: the bare essentials. *Pract Neurol*. 2009;9(5):303–9. <https://doi.org/10.1136/jnnp.2009.188151> PMID: 19762894
- Turabieh H, Afshar AS, Statland J, Song X. Towards a machine learning empowered prognostic model for predicting disease progression for amyotrophic lateral sclerosis. *AMIA Annu Symp Proc*. 2024;2023:718–25.
- Kjældgaard A-L, Pilely K, Olsen KS, Jessen AH, Lauritsen AØ, Pedersen SW, et al. Prediction of survival in amyotrophic lateral sclerosis: a nationwide, Danish cohort study. *BMC Neurol*. 2021;21(1):164. <https://doi.org/10.1186/s12883-021-02187-8> PMID: 33865343
- Feldman EL, Goutman SA, Petri S, Mazzini L, Savelieff MG, Shaw PJ, et al. Amyotrophic lateral sclerosis. *Lancet*. 2022;400(10360):1363–80. [https://doi.org/10.1016/S0140-6736\(22\)01272-7](https://doi.org/10.1016/S0140-6736(22)01272-7) PMID: 36116464
- Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). *J Neurol Sci*. 1999;169(1–2):13–21. [https://doi.org/10.1016/S0022-510X\(99\)00210-5](https://doi.org/10.1016/S0022-510X(99)00210-5) PMID: 10540002
- Kuan L-H, Parnianpour P, Kushol R, Kumar N, Anand T, Kalra S, et al. Accurate personalized survival prediction for amyotrophic lateral sclerosis patients. *Sci Rep*. 2023;13(1):20713. <https://doi.org/10.1038/s41598-023-47935-7> PMID: 38001260
- van der Burgh HK, Schmidt R, Westeneng H-J, de Reus MA, van den Berg LH, van den Heuvel MP. Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis. *Neuroimage Clin*. 2016;13:361–9. <https://doi.org/10.1016/j.nicl.2016.10.008> PMID: 28070484
- Vieira FG, Venugopalan S, Premasiri AS, McNally M, Jansen A, McCloskey K, et al. A machine-learning based objective measure for ALS disease severity. *NPJ Digit Med*. 2022;5(1):45. <https://doi.org/10.1038/s41746-022-00588-8> PMID: 35396385
- M Amaral D, Soares DF, Gromicho M, de Carvalho M, Madeira SC, Tomás P, et al. Temporal stratification of amyotrophic lateral sclerosis patients using disease progression patterns. *Nat Commun*. 2024;15(1):5717. <https://doi.org/10.1038/s41467-024-49954-y> PMID: 38977678
- Din Abdul Jabbar MA, Guo L, Nag S, Guo Y, Simmons Z, Pioro EP, et al. Predicting amyotrophic lateral sclerosis (ALS) progression with machine learning. *Amyotroph Lateral Scler Frontotemporal Degener*. 2024;25(3–4):242–55. <https://doi.org/10.1080/21678421.2023.2285443> PMID: 38052485
- Ong M-L, Tan PF, Holbrook JD. Predicting functional decline and survival in amyotrophic lateral sclerosis. *PLoS One*. 2017;12(4):e0174925. <https://doi.org/10.1371/journal.pone.0174925> PMID: 28406915
- Haider H, Hoehn B, Davis S, Greiner R. Effective ways to build and evaluate individual survival distributions. *J Mach Learn Res*. 2020;21(1):3289–351.

14. Atassi N, Berry J, Shui A, Zach N, Sherman A, Sinani E, et al. The PRO-ACT database: design, initial analyses, and predictive features. *Neurology*. 2014;83(19):1719–25. <https://doi.org/10.1212/WNL.0000000000000951> PMID: 25298304
15. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: with applications in R. 2nd ed. New York: Springer; 2021.
16. Shen X, Elmer J, Chen GH. Neurological prognostication of post-cardiac-arrest coma patients using EEG data: a dynamic survival analysis framework with competing risks. In: *Proceedings of the 8th Machine Learning for Healthcare Conference*. 2023. p. 667–90.
17. Andres A, Montano-Loza A, Greiner R, Uhlich M, Jin P, Hoehn B, et al. A novel learning algorithm to predict individual survival after liver transplantation for primary sclerosing cholangitis. *PLoS One*. 2018;13(3):e0193523. <https://doi.org/10.1371/journal.pone.0193523> PMID: 29543895
18. Lillelund CM, Harbo M, Pedersen CF. Prognosis of fall risk in home care clients: a noninvasive approach using survival analysis. *J Public Health (Berl)*. 2024. <https://doi.org/10.1007/s10389-024-02317-9>
19. Lajoie I, Canadian ALS Neuroimaging Consortium (CALSNIC), Kalra S, Dadar M. Regional cerebral atrophy contributes to personalized survival prediction in amyotrophic lateral sclerosis: a multicentre, machine learning, deformation-based morphometry study. *Ann Neurol*. 2025;97(6):1144–57. <https://doi.org/10.1002/ana.27196> PMID: 39985309
20. Sechidis K, Tsoumakas G, Vlahavas I. On the stratification of multi-label data. *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer; 2011. p. 145–58. https://doi.org/10.1007/978-3-642-23808-6_10
21. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 1972;34(2):187–202. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>
22. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat*. 2008;2(3). <https://doi.org/10.1214/08-aos169>
23. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol*. 2018;18(1):24. <https://doi.org/10.1186/s12874-018-0482-1> PMID: 29482517
24. Yu CN, Greiner R, Lin HC, Baracos V. Learning patient-specific cancer survival distributions as a sequence of dependent regressors. In: *Proceedings of the 24th International Conference on Neural Information Processing Systems*, Granada, Spain, 2011. p. 1845–53.
25. Lillelund CM, Gharari Foomani AH, Sun W, Qi S, Greiner R. MENSA: a multi-event network for survival analysis with trajectory-based likelihood estimation. *arXiv preprint* 2023. <https://arxiv.org/abs/2409.06525>
26. Harrell FE. Evaluating the yield of medical tests. *JAMA*. 1982;247(18):2543. <https://doi.org/10.1001/jama.1982.03320430047030>
27. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med*. 2011;30(10):1105–17. <https://doi.org/10.1002/sim.4154> PMID: 21484848
28. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Statist Med*. 1999;18(17–18):2529–45. [https://doi.org/10.1002/\(sici\)1097-0258\(19990915/30\)18:17/18<2529::aid-sim274>3.0.co;2-5](https://doi.org/10.1002/(sici)1097-0258(19990915/30)18:17/18<2529::aid-sim274>3.0.co;2-5)
29. Qi S, Sun W, Greiner R. SurvivalEVAL: a comprehensive open-source python package for evaluating individual survival distributions. *AAAI-SS*. 2024;2(1):453–7. <https://doi.org/10.1609/aaaiss.v2i1.27713>
30. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*. 1958;53(282):457–81. <https://doi.org/10.1080/01621459.1958.10501452>
31. Qi S-A, Kumar N, Farrokh M, Sun W, Kuan L-H, Ranganath R, et al. An effective meaningful way to evaluate survival models. *Proc Mach Learn Res*. 2023;202:28244–76. PMID: 40895293
32. Prosperi M, Guo Y, Sperrin M, Koopman JS, Min JS, He X, et al. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nat Mach Intell*. 2020;2(7):369–75. <https://doi.org/10.1038/s42256-020-0197-y>
33. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*. 1966;50(3):163–70. PMID: 5910392
34. Jaiswal MK. Riluzole and edaravone: a tale of two amyotrophic lateral sclerosis drugs. *Med Res Rev*. 2019;39(2):733–48. <https://doi.org/10.1002/med.21528> PMID: 30101496
35. Swinnen B, Robberecht W. The phenotypic variability of amyotrophic lateral sclerosis. *Nat Rev Neurol*. 2014;10(11):661–70. <https://doi.org/10.1038/nrneurol.2014.184> PMID: 25311585
36. Atassi N, Yerramilli-Rao P, Szymonifka J, Yu H, Kearney M, Grasso D, et al. Analysis of start-up, retention, and adherence in ALS clinical trials. *Neurology*. 2013;81(15):1350–5. <https://doi.org/10.1212/WNL.0b013e3182a823e0> PMID: 24005339
37. Emura T, Chen Y-H. Analysis of survival data with dependent censoring. Singapore: Springer; 2018. <https://doi.org/10.1007/978-981-10-7164-5>
38. Foomani AHG, Cooper M, Greiner R, Krishnan RG. Copula-based deep survival models for dependent censoring. In: *Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence*. 2023. p. 669–80.
39. Dadar M, Manera AL, Zinman L, Korngut L, Genge A, Graham SJ, et al. Cerebral atrophy in amyotrophic lateral sclerosis parallels the pathological distribution of TDP43. *Brain Commun*. 2020;2(2):fcaa061. <https://doi.org/10.1093/braincomms/fcaa061> PMID: 33543125
40. Dieckmann N, Roediger A, Prell T, Schuster S, Herdick M, Mayer TE, et al. Cortical and subcortical grey matter atrophy in amyotrophic lateral sclerosis correlates with measures of disease accumulation independent of disease aggressiveness. *Neuroimage Clin*. 2022;36:103162. <https://doi.org/10.1016/j.nicl.2022.103162> PMID: 36067613