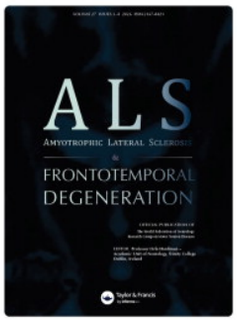
Methodological considerations in the analysis of survival data in amyotrophic lateral sclerosis

Yazdani, S.; Seitz, C.; Andersson, A.; Ingre, C.; Fang, F.; Lovik, A.



Taylor & Francis
Taylor & Francis Group

Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration

ISSN: 2167-8421 (Print) 2167-9223 (Online) Journal homepage: www.tandfonline.com/journals/iafd20

Methodological considerations in the analysis of survival data in amyotrophic lateral sclerosis

Solmaz Yazdani, Christina Seitz, John Andersson, Caroline Ingre, Fang Fang & Anikó Lovik

To cite this article: Solmaz Yazdani, Christina Seitz, John Andersson, Caroline Ingre, Fang Fang & Anikó Lovik (2026) Methodological considerations in the analysis of survival data in amyotrophic lateral sclerosis, Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 27:3-4, 445-456, DOI: 10.1080/21678421.2026.2615111

To link to this article: https://doi.org/10.1080/21678421.2026.2615111

Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 2026; 27: 445–456

REVIEW ARTICLE

Methodological considerations in the analysis of survival data in amyotrophic lateral sclerosis

SOLMAZ YAZDANI[1], CHRISTINA SEITZ[2], JOHN ANDERSSON[2], CAROLINE INGRE[1,3], FANG FANG[2] & ANIKÓ LOVIK[2,4]

[1]*Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden,* [2]*Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden,* [3]*Neurology Clinic, Karolinska University Hospital, Stockholm, Sweden, and* [4]*Institute of Psychology, Leiden University, Leiden, The Netherlands*

Abstract

Survival outcomes are commonly analyzed in studies with data from patients with progressive, neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS). Given the fast progression of ALS, survival analyses are, however, often difficult to perform and interpret. In this methodological article we demonstrate on real-world data how the choices we make in the study design, data collection, and analysis could influence the results. The factors we consider in this study are length of follow-up, sample size, timing of sample collection, and choice of covariables adjusted for in the models. We further discuss the importance of each of these contributing factors and about how to avoid mistakes in interpreting and reporting survival data in ALS and other rare, progressive diseases.

Keywords: *Cox regression, methodology, statistical analysis, survival analysis, tutorial*

Introduction

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease that affects the upper and lower motor neurons, causing a gradual deterioration of the voluntary muscle function, and usually leads to death within 2–4 years from symptom onset (i.e. as few as 10% of the patients survive longer than 10 years). ALS is a relatively rare disease with an incidence of 0.3–4.6 per 100,000 person-years that can be variable depending on the region and ethnicity (1–3). ALS can present with a wide range of symptoms and constitutes therefore a heterogeneous disease (2,4). The rapidly progressing course of the disease, the rarity, and the heterogeneity pose difficulties in the research of ALS (5).

In many clinical studies, the outcome of interest is the time until a specific event such as death, a particular biomarker reaching a certain threshold or a patient requiring mechanical ventilation – commonly referred to as ventilation-free survival in ALS trials. This type of data, known as survival data, is commonly analyzed using the Cox proportional hazards model. This model is frequently employed for several reasons. First, many researchers are familiar with the basics of regression models and the interpretation of Cox regression (as we will show in the next section) is straightforward with a basic understanding of regression. Second, Cox models are implemented in many if not all popular statistical software including SAS, Stata, Statistica, SPSS, R and Python, and require no technical or advanced statistical programming skills. Last, Cox models are also popular because they allow the inclusion of multiple predictors alongside the primary exposure or treatment group. For instance, in observational studies, the exposure might be the level of a biomarker, while in clinical trials, it could involve comparing a new treatment method to the standard of care (6). However, the outcome of interest, the decisions made in the study design and performing data collection and analysis could play important roles in the results and the interpretation of

Correspondence: Anikó Lovik, Institute of Environmental Medicine, Karolinska Institutet, Nobels väg 13, 171 77 Stockholm, Sweden. E-mail: aniko.lovik@ki.se.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/21678421.2026.2615111.

(Received 28 May 2025; revised 18 December 2025; accepted 7 January 2026)

ISSN 2167-8421 print/ISSN 2167-9223 online
DOI: 10.1080/21678421.2026.2615111

the results. Yet, in the field of ALS, there has been little consideration given to some of these factors and many are concerned with clinical trial design such as defining outcome measures (7) or outcomes (8). As an example from the non-trial literature, studies on the serum level of neurofilament light chain (NfL) as a prognostic marker in ALS could help highlight the influence of methodological concerns on survival analysis of ALS. Four studies (9–12), using a sample size of 64 to 383 patients, reported different HRs for survival in relation to a serum NfL levels above the median, compared to below median, ranging from 2.21 (95%CI: 1.51–3.24) in one study to 6.05 (1.68–21.87) in another. This variation could be attributed to factors discussed in the present study, such as differing sample sizes (64, 124, 198, 383, event numbers were only reported for one study (78) in the main text), follow-up durations (15, 60, 120 and 150 months), definition of survival time (only reported for one study), inclusion/exclusion criteria, variable selection or different median (cut-off) values for the biomarkers, among other things. This emphasizes the importance of transparent reporting in survival analysis of ALS data.

In this discussion article, we will explore some of the contributing factors that can impact the results of survival analysis such as the follow-up time, the sample size, the time of sampling for biospecimens and the covariables adjusted for in the model. In our analysis we will focus on issues and decisions made when analyzing data employing classical survival analysis (i.e. Cox models), that are specific to rare, progressive diseases such as ALS. Our aim is to illustrate the potential impact of these decisions using a real data example and to draw the attention of ALS researchers to these issues.

A short introduction to the Cox proportional hazards model

The Cox proportional hazards model (or Cox regression) was introduced by Sir David Cox in 1972 (13) to analyze survival data and test the effect of several predictor variables on the time to an event (in ALS research often death or the start of invasive ventilation). This effect is commonly given in the form of hazard ratios (HR), which can be interpreted similarly to odds ratios obtained from a logistic regression. For instance, comparing treatment to control group (denominator), a hazard ratio of 1 means the same probability of an event occurring in the two groups at any given time, while a HR of 2 would indicate an event at twice the rate in the treatment compared to the control group. Note that the HR is a relative measure of risk and not an absolute one. A more detailed explanation of the Cox model is given in Figure 1.

Because the Cox model estimates a single, time-independent hazard ratio (HR), the model assumes that the hazards are proportional for the entire study period. This is called the proportional hazards (PH) assumption. Several methods exist to assess this assumption, including graphical approaches and statistical tests based on Schoenfeld residuals. However, the PH assumption is often violated in practice, which can lead to biased or misleading estimates if not appropriately addressed (14–16).

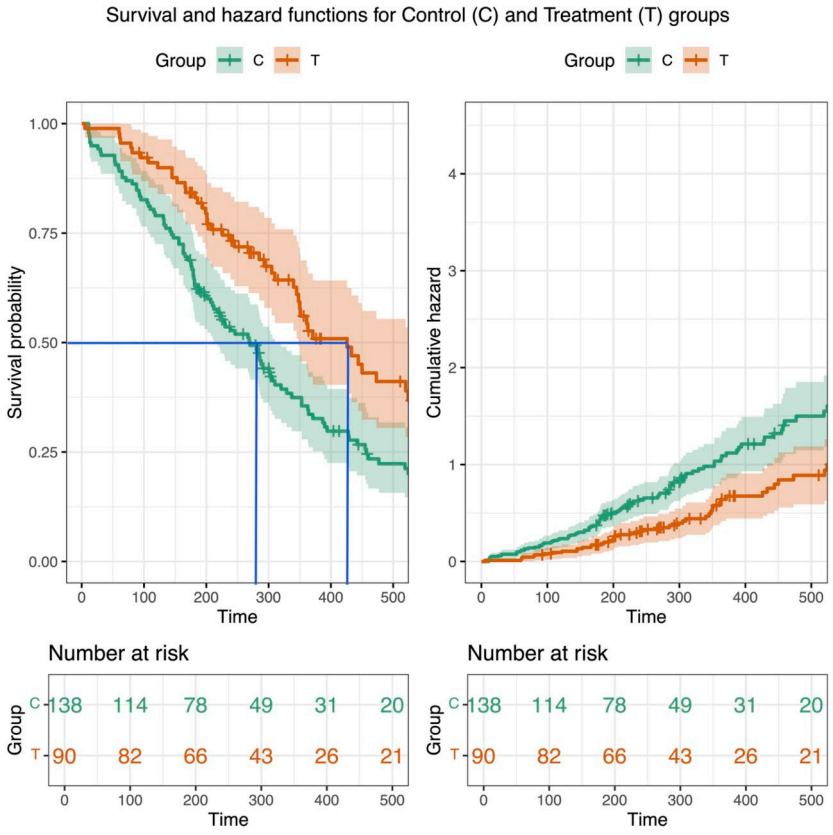
Even when the proportional hazards assumption is tested and holds, extrapolating the results to longer timeframes may be problematic, as we assume that the PH assumption will hold beyond the time period studied. This issue may be exacerbated in progressive diseases like ALS where dropouts are often linked to outcome, exposure, or covariates. This, on one hand, may render the results invalid, but also hinders the comparability and reproducibility of findings between studies. Consequently, inconsistent findings in analysis of ALS survival data may partially stem from differing follow-up durations and untestable violations of the PH assumption (17). Next, we will focus on the small decisions made during a study.

Challenges in survival analysis of ALS patients

Sample size. A critical challenge in ALS research is limited sample size, which is a common issue for all rare diseases (17–19). Due to the heterogeneity of ALS patients, however, the in- or exclusion of specific patients, or a slightly smaller or larger sampling fraction from the source patient population may have a larger impact in ALS than in diseases with more homogeneous patient populations. Moreover, minimum sample size requirements are often not met (20), resulting in wide confidence intervals and non-reproducible study results.

A commonly cited rule of thumb for the sample size recommends at least 10 events per predictor variable to ensure model stability and to reduce the risk of overfitting (21). More recent guidelines, however, advocate for a context-specific approach that accounts for model complexity, outcome incidence, and intended use (20).

Statistical power in survival analysis is driven primarily by the number of events (a combination of sample size, follow-up duration and the probability of an event), not by the total sample size alone. Small numbers of events can lead to imprecise and potentially biased effect estimates. In addition, lower power increases the risk of Type II error (i.e. false negatives). Importantly, the Type I error rate (i.e. false positives) remains controlled under correct model specification, even in small samples. Meeting these requirements remains a substantial challenge in ALS research due to small sample sizes and the large number of clinically relevant variables.



On the left-hand side, we see the survival functions for control (C) and treatment (T) groups. The vertical axis represents the survival probability, the horizontal axis time in days (since start of the study). The blue lines indicate the median survival times for both groups. We can see from the blue lines that the median survival is less than 300 days for the control and more than 400 days for the treatment group.

On the right-hand side, we see the matching cumulative hazard functions.

The tables under both figures depict the patients at risk (patients still in the study, so uncensored and still alive at the moment).

A bit more technical explanation of the Cox regression

The Cox proportional hazards model takes the form for each patient *i* for a single covariate:

$$h(t|X_i) = h_0(t) \cdot \exp(\beta_1 X_{i1})$$

Where $h(t|X_i)$ or $h(t)$ is a **hazard function** ($h_0(t)$ being the **baseline hazard function**, identical for all patients) and $\exp(\beta_1 X_{i1})$ is the **scaling factor** for patient *i*. In general terms, this means that the hazard given observed covariates X_i is the baseline hazard multiplied by the (exponentiated) **linear predictor** (which takes the form $\exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip})$ for covariates $X_1, X_2, \ldots, X_p$. The coefficient for covariate *j* denoted as $\exp(\beta_j)$ is called a **hazard ratio** (HR).

The **proportional hazards assumption** of Cox model postulates that the ratio of the hazards of an event should be constant over time (a.k.a. proportional hazards (PH) assumption). To demonstrate this, let's consider two patients, *i* and *j*. The ratio of the hazards:

$$\frac{h(t|X_i)}{h(t|X_j)} = \frac{h_0(t) \cdot \exp(\beta \cdot X_i)}{h_0(t) \cdot \exp(\beta \cdot X_j)} = \frac{\exp(\beta \cdot X_i)}{\exp(\beta \cdot X_j)} = \exp\left(\beta \cdot (X_i - X_j)\right)$$

Because the baseline hazard (which is the same for all patients) is cancelled out, the right-hand side is not dependent on time, hence the ratio of the hazards is constant, i.e., the hazards are proportional. Provided that the PH assumption holds, we can estimate the parameter estimate β without considering the entire hazard function, which is the main advantage of the Cox model.

Figure 1. Survival and hazard functions comparing two groups and a brief introduction to the Cox regression and its assumption with formulae. Note for the figure on top: fictional example representing survival and hazard functions for a discrete predictor comparing control and treatment groups (Note that continuous variables such as biomarker levels cannot be represented.

Time of biomarker sampling in a progressive disease. In ALS, the timing of biospecimen collection is a critical methodological consideration. The disease follows a progressive course, and there is often a substantial delay between disease initiation, symptom onset, and clinical diagnosis which can span from several months to over a year. Moreover, patients exhibit diverse disease trajectories after diagnosis as well. Biomarker measurements can therefore be obtained at very different points along this disease timeline. For many exploratory (i.e. not yet established and/or not ALS-specific) biomarkers, we do not yet understand their trajectory over time and how this

trajectory is related to disease progression. Therefore, differences in measured levels are not necessarily a direct reflection of disease progression but could be attributed to individual differences, different disease subtype, etc. As a result, samples taken at different disease stages may not be directly comparable.

To reduce variability related to disease stage, many studies adopt narrow sampling windows, such as collecting samples within the first three months after diagnosis to capture patients at a more comparable stage. In contrast, sampling later in the disease course can introduce selection bias or at least unwanted differences and heterogeneity in the study population, as these cohorts are more likely to include patients with slower progression and longer survival. Differences in sampling strategy make comparisons across studies challenging.

From a statistical perspective, Cox models can account for timing, but this requires careful specification of the time scale and appropriate reporting. In ALS biomarker research, potential starting points include the date of biomarker measurement, time since diagnosis, or time since symptom onset. The latter two approaches implicitly assume that biomarker levels remain constant over time – a strong assumption that may not hold, especially in case of a progressive disease. If violated, this can lead to immortal time bias, whereby survival time is artificially inflated because patients must survive until the biomarker measurement is taken. Therefore, correct model specification and transparent reporting of time origin are essential for valid inference.

Multivariable adjustment is a common approach used to address heterogeneity and enhance the statistical efficiency of data analysis (13,22). However, larger sample sizes are often required to accommodate adjustments for an increased number of covariables. Covariables selection is typically informed by prior literature and expert knowledge; but over-adjustment and overfitting are frequent challenges. Another issue, though not addressed in the present study, is multicollinearity among covariables, which can sometimes yield results that are difficult to interpret. For these reasons, careful selection of covariables is essential, as we will demonstrate.

Each of these issues could produce invalid or biased results, leading to erroneous conclusions. Furthermore, coexistence of multiple problems can, and does, occur and lead to greater validity concerns. In the present study, we show how each of these issues may impact the results of survival analysis in ALS in a relatively large, population-based sample of newly diagnosed ALS patients from Stockholm, Sweden. We propose specific aspects of the survival analysis to report and emphasize when evaluating the validity of the study results and before generalizing the findings to other settings.

In the Introduction, we have presented factors related to study design and statistical analyses and indicated how they could impact the results of survival analysis in ALS. The remainder of the article is organized as follows: instead of the usual materials and methods, we first introduce the *Motivating case study* to illustrate the effect of analytical choices in survival analysis in a real-world setting then detail the *Statistical analysis of the motivating case study*. These sections are followed by the *Results* we have obtained, and the article concludes with a *Discussion* of the findings and recommendations for ALS researchers performing survival analysis.

Motivating case study

The motivating dataset consisted of 159 individuals with ALS who were diagnosed and enrolled in the study – by donating blood and cerebrospinal fluid (CSF) samples –between March 2016 and May 2023 at the ALS Clinical Research Center of Karolinska University Hospital in Stockholm, Sweden. Of these, 115 individuals donated a single time one sample of blood and one of CSF, whereas 44 contributed multiple samples over time. All participants received a diagnosis of probable, possible, or definite ALS, according to the El Escorial criteria. To characterize immune system involvement, flow cytometry was used to analyze T cells subsets from fresh blood and CSF samples collected at the time of enrollment (a brief description of the biomarkers can be found in Supplementary Table 1). Detailed protocols for sample processing and analysis have been described elsewhere (23). During the follow-up period, 67 participants reached the primary endpoint of death or initiation of invasive ventilation. Oral and written informed consent was collected from all patients. The study was approved by the Swedish Ethical Review Authority (ALSrisc: DNRs 2014/1815-31/4 (extended 2018/1065-31, 2021-06397-02), MND Registry: 2017/1895-31/1).

Information on age, sex, date of diagnosis, date and site of symptom onset, ALSFRS-R score and body mass index (BMI) was obtained for all participants through linkage to the Swedish Motor Neuron Disease Quality Registry. Through linkage to this registry, all individuals were followed from the date of first sampling of biospecimen, until the date of death, the use of invasive ventilation, or July 29th, 2023, whichever occurred first. Progression rate at the time of sampling was calculated as 48 minus ALSFRS-R score around the time of sampling divided by the time interval between symptom

onset and the time of ALSFRS-R score measurement closest to the sampling (in months).

Statistical analysis of the motivating case study

Cox proportional hazards model was used to estimate the risk of death (alternatively use of invasive ventilation) in relation to the level of different biomarkers (specifically, T cell subsets). The model was adjusted for sex, age and BMI at the time of sampling, site of onset, diagnostic delay and ALSFRS-R score closest to the time of sampling. We chose these variables as examples as these variables are often adjusted for in multivariable survival models for ALS. A typical model could take the form:

$$h(t) = h_0(t)\cdot\exp(\beta_1\cdot Biomarker + \beta_2\cdot Sex + \beta_3\cdot Age + \beta_4\cdot BMI + \beta_5\cdot OnsetSite + \beta_6\cdot DiagDelay + \beta_7\cdot ALSFRS)$$

where *Biomarker* represents the T cell subsets measured in blood and CSF described above (our original model). To analyze the impact of follow-up length, sample size, and choice of multivariable adjustment, we selected each patient's sample that was closest to the time of diagnosis, ensuring it was within 3 months of diagnosis. To examine the effect of sampling timing, we included only patients with multiple measurements and used both the sample closest to and the sample furthest from the diagnosis date.

Follow-up time. To assess the impact of the length of the follow-up period, we analyzed baseline samples taken within 3 months following the date of diagnosis ($n = 106$). If multiple measurements were available during this period, the one closest to the diagnosis date was selected, creating the baseline dataset ($n = 105$). For follow-up analysis, we tracked these patients over two timeframes: one year (short follow-up) and two years (long follow-up).

Sample size. To address the impact of sample size, we generated three datasets of varying sizes but with a consistent two-year follow-up period. Specifically, the smaller and medium-sized datasets comprised random samples of 40 and 70 individuals, respectively, selected from the baseline dataset, while the larger dataset included all 105 individuals in the baseline dataset. Additionally, in a sensitivity analysis, to depict the sampling variability, we created two more datasets, each with a unique random sample of 60 individuals, enabling a comparison of random sampling effects.

Time of biospecimen sampling. To study the impact of sampling time relative to the diagnosis date, we included the 44 individuals who had donated multiple samples. We created two datasets: one with samples taken closest to the diagnosis date (within 3 months; $n = 32$) and a second with samples taken furthest from the diagnosis date (minimum of 3 months; $n = 44$). Survival time was calculated from the biomarker sampling date to death or initiation of invasive ventilation, or end of follow-up. We then matched individuals across these two datasets to allow a direct comparison of the same patients, isolating the effect of sampling time relative to diagnosis ($n = 32$).

Variable selection. Finally, we studied the impact of the choice of covariables in Cox model. Different research groups may use similar but not the same covariables, which also can have an impact on the obtained results. Using the baseline dataset (described above) including the baseline measurements from every patient ($N = 115$) and a follow-up duration of two years, we used five different models as follows:

Model 1: Biomarker + Age + Sex
Model 2: Biomarker + Age + Sex + Site of onset
Model 3: Biomarker + Age + Sex + Site of onset + BMI
Model 4: Biomarker + Age + Sex + Site of onset + BMI + ALSFRS-R + Diagnosis delay
Model 5: Biomarker + Age + Sex + Site of onset + BMI + Progression rate

These models were selected to illustrate how small choices in variable selection can influence results. For example, progression rate inherently incorporates information from both ALSFRS-R and time between symptom onset and ALSFRS-R measurement and is therefore not included together with these variables to avoid collinearity. As the purpose of all the analyses is to illustrate methodological choices, no adjustment was made for multiple comparisons, and no p values will be shown. The results from the models are presented as hazard ratios and the default 95% confidence intervals. The results are depicted on forest plots (Figure 2) and in Tables (Table 2 and Supplementary Tables 2 and 4). The PH assumption was evaluated using Schoenfeld's residual test (PH test). All analyses were performed using R 4.3.0 and the R packages "dplyr" (24), "tidyverse" (25)," ggplot2" (26)," survival" (27), "berryFunctions" (28), "survminer" (29).

Results

Descriptives. Table 1 presents the basic characteristics of the sample in our motivating case study. The last column describes the full sample, which is a representative of ALS patients in Stockholm County (30), while the other columns show two random subsets. Notably, taking a random sample affects some characteristics, such as sex distribution (women are overrepresented in

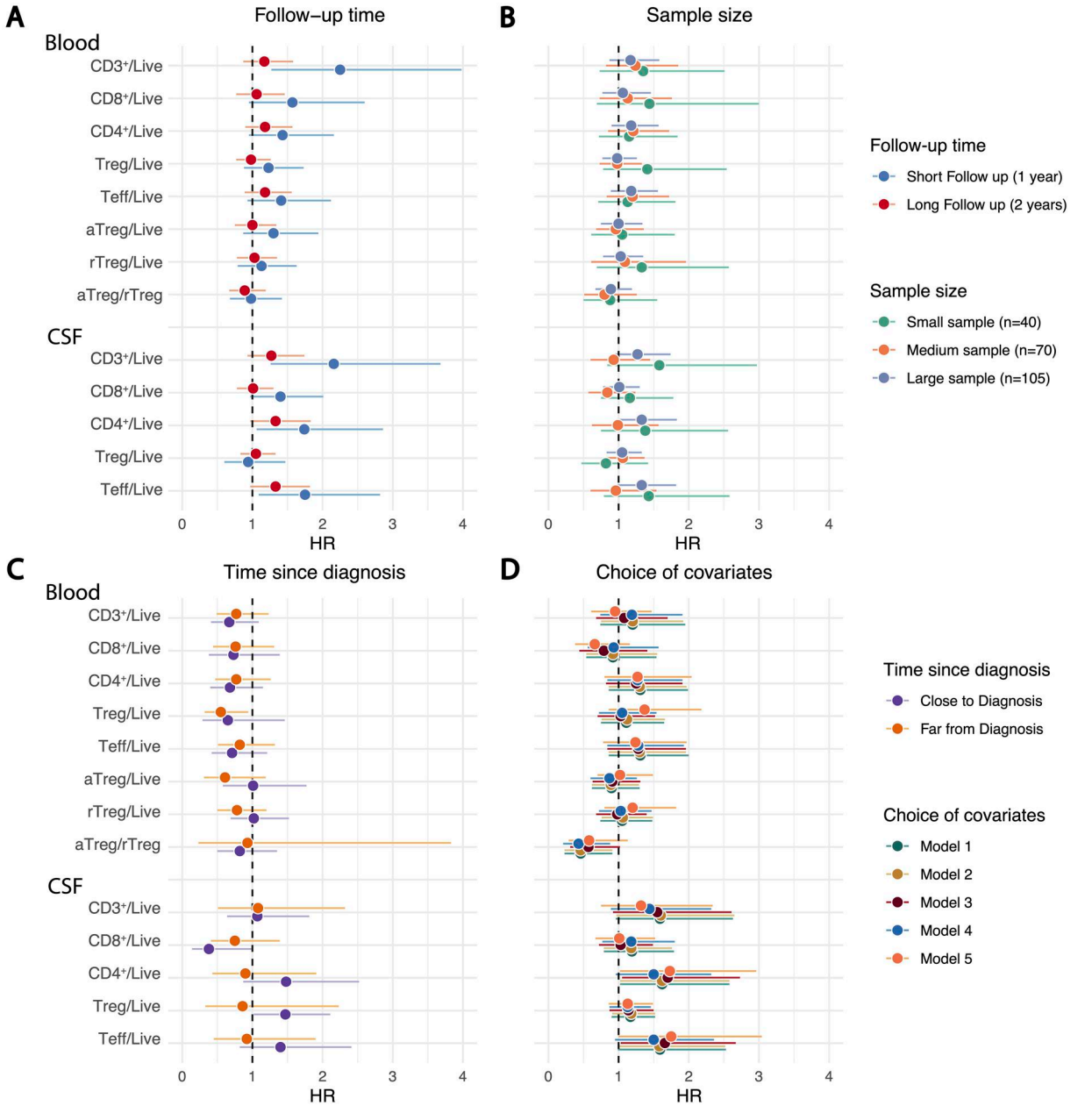


Figure 2. Hazard ratios (HR) and their 95% confidence intervals (CI) of death (or use of invasive ventilation) in relation to the biomarker frequency in blood and cerebrospinal fluid (CSF). The model is adjusted for age at sampling, sex, site of onset, diagnostic delay, weight difference between diagnosis and sampling, ALSFRS-R score closest to sampling, and time difference between ALSFRS-R score measurement and sampling. Model 4 as the original model, Model 1 including only biomarker age, and sex; Model 2 adding site of onset; Model 3 adding BMI; and Model 5 substituting progression rate for ALSFRS-R and diagnosis delay.

the smallest sample) and site of onset (bulbar onset is slightly overrepresented in the small and underrepresented in the medium sample). (Note, that this happens because of taking just a single random sample. If the process were to be repeated 100 times and the results averaged, the distributions would closely match those observed in the population.) Also note the generally larger standard deviations for the smaller samples (with the notable exception of diagnostic delay, which shows a different pattern). Despite these small discrepancies, even the smaller samples represent the population reasonably well due to the registry's high coverage.

Follow-up time. As anticipated, with the longer follow-up period the number of deaths (events) increased, leading to more precise estimates and smaller confidence intervals (CIs) (Figure 2(A) and Table 2). For all biomarkers, the hazard ratio (HR) was closer to 1 in the longer follow-up duration than in the shorter follow-up, implying smaller effect sizes.

Sample size. Larger sample sizes were associated with improved estimate precision and narrower confidence intervals (Figure 2(B) and Table 2). While some biomarkers showed more stable HRs in the larger samples, others – particularly CSF

Table 1. Clinical characteristics of the full sample of patients (last column) and the two smaller datasets of different sizes.

Characteristics	Small (N = 40)	Medium (N = 70)	Large (N = 105) Full sample
Sex, N (%)			
Female	17 (42.5%)	29 (41.4%)	41 (39%)
Male	23 (57.5%)	41 (58.6%)	64 (61%)
Site of onset, N (%)			
Spinal	24 (60%)	44 (62.9%)	62 (59%)
Bulbar	16 (40%)	25 (35.7%)	41 (39%)
Missing	0	1 (1.4%)	2 (2%)
Age at sampling			
Median	62.26	65.78	66.54
Mean (SD)	63.52 (14.42)	64.18 (12.17)	65.18 (11.61)
BMI			
Median	22.53	23.82	24.09
Mean (SD)	23.65 (4.07)	23.95 (3.35)	24.15 (3.57)
Diagnostic delay in days			
Median	303	338	341
Mean (SD)	384 (277)	448 (494)	475 (526)
ALSFRS-R closest to sampling			
Median	41.5	42	41
Mean (SD)	38.7 (8.7)	39.4 (7)	39 (7)

markers – exhibited considerable variability across different sample sizes. Notably, fluctuations were more pronounced among CSF markers. In part, this heterogeneity is due to taking a random sample (and not just sample size), since there may be considerable differences between random samples.

Time of biospecimen sampling. We further examined the impact of the time interval between biospecimen sampling and ALS diagnosis on our results. Differences in HRs were observed for most biomarkers, with particularly pronounced variability among CSF variables (Figure 2(C) and Supplementary Table 2). This variability may partly reflect the hypothesis that the influence of biomarkers on mortality risk evolves over the disease course. Additionally, only patients who survived at least three months (the minimum interval we set between biospecimen samplings corresponding to the time between patient visits) could provide multiple samples. As shown in Supplementary Table 3, patients in the second dataset were older, with lower BMI and ALSFRS-R scores compared to those in the first dataset.

Variable selection. In the panel of Figure 2 (Figure 2(D) and Supplementary Table 4), the impact of the choice of covariables is displayed for five different models. As expected, the difference between the models are relatively small, however, even such a small difference in models can have a large impact in ALS research: while the HR point estimates did not differ largely between the different models, but in some models the exposure's effect on survival was statistically significant while in other models it was not, leading to potentially different conclusions.

Discussion

In this study, we illustrated how small decisions related to the study design and data analysis could impact the results of survival data analysis in ALS using Cox regression. While generally it is expected that a larger sample size will lead to more precise estimates or that the length of follow-up will have no or only minor impact on the hazard ratio estimates, using real-world data, we demonstrated that, in the case of a rare and heterogenous disease like ALS, the situation might be more complex.

Follow-up time. We observed differences between the analyses with shorter and longer follow-up periods. Various reasons might underlie such pattern. First, a biomarker measured close to the time of diagnosis may be more indicative of risk of death in the period immediately following diagnosis than later in the disease course. An attenuation effect over time may have thus a biological explanation. Second, shorter follow-up likely captures deaths primarily among patients with rapidly progressing disease – a more homogeneous subgroup. In contrast, longer follow-up includes deaths among patients with slower progression, introducing greater heterogeneity. Third, the proportional hazards assumption may not be realistic as it implies that the effect of a biomarker measured at a single time point remains constant over one month, one year, or even ten years which is a biologically unlikely scenario. This may be very improbable. We suggest that, when interpreting results from survival analysis in progressive diseases, researchers should always emphasize the

Table 2. The HR estimates corresponding confidence intervals and the assumption check for all models in panels a and B of Figure 2.

	Short (1 year)			Long (2 years)/Large sample (N = 105)			Medium sample (N = 70)			Small sample (N = 40)		
	N (Events)	HR (95% CI)	PH-test*	N (Events)	HR (95% CI)	PH-test*	N (Events)	HR	PH-test*	N (Events)	HR	PH-test*
Blood												
CD3+/Live	98 (27)	**2.25 (1.27–3.98)**	0.82	98 (54)	1.17 (0.87–1.58)	0.14	68 (37)	1.24 (0.82–1.85)	0.17	37 (20)	1.35 (0.73–2.51)	*0.05*
CD8+/Live	86 (22)	1.57 (0.95–2.60)	0.69	86 (48)	1.06 (0.77–1.46)	0.15	60 (33)	1.13 (0.73–1.76)	0.10	33 (19)	1.44 (0.69–3.00)	*0.04*
CD4+/Live	97 (27)	1.43 (0.95–2.16)	0.89	97 (54)	1.18 (0.90–1.57)	0.54	67 (37)	1.21 (0.85–1.72)	0.74	37 (20)	1.15 (0.72–1.84)	0.24
CSF												
CD3+/Live	98 (28)	**2.16 (1.26–3.68)**	0.75	98 (54)	1.27 (0.93–1.74)	0.18	66 (36)	0.93 (0.60–1.45)	0.15	37 (21)	1.58 (0.84–2.97)	0.24
CD8+/Live	85 (23)	1.40 (0.97–2.01)	0.81	85 (48)	1.01 (0.78–1.30)	0.25	58 (32)	0.84 (0.57–1.24)	0.14	33 (20)	1.16 (0.75–1.78)	0.40
CD4+/Live	98 (28)	**1.74 (1.06–2.86)**	0.67	98 (54)	1.33 (0.97–1.83)	0.34	66 (36)	0.99 (0.62–1.57)	0.49	37 (21)	1.38 (0.75–2.56)	0.36

*The *p* value of the PH-test (proportional hazards test) to test the PH assumption. If non-significant, the hazards are assumed to be proportional.

length of follow-up, and HRs should be interpreted as applying to that specific time frame.

Sample size. Having a larger sample size will often lead to greater statistical power in statistical analysis. However, in survival analysis, the number of events (driven by sample size, follow-up duration and the probability of an event) is more important. Since ALS is a progressive disease, a longer follow-up time will increase the number of events for the same sample automatically as seen in our example.

In addition, it is also important to look at the composition of the patients and their clinical characteristics. Site of onset, sex, delay in diagnosis, etc. are all important pieces of information about how the patients will progress over time. Ensuring a representative sample that is large enough to perform the desired hypothesis tests is key. Performing a sample size calculation or power analysis before data collection could reduce the number of non-reproducible results and contradictory reports. Moreover, while the suggestion to acquire larger samples sizes may seem straightforward, it presents a challenge when working with rare diseases like ALS, where ongoing sample collection may span several additional years. An alternative is to apply Firth's correction to the regular Cox regression to increase power and reduce small sample bias (31,32).

Furthermore, survival analyses typically require an adequate number of events per variable (EPV) – often cited as at least 10 events per predictor or in case of categorical predictors have unbalanced distributions even higher EPV (e.g. 20–50+) may be needed (33,34) – to ensure model stability and reliable estimates. In practice, these thresholds are rarely met in ALS studies, and the number of events is often not reported at all, making it difficult to assess the robustness of published findings. This tension between ideal methodological standards and real-world constraints is a core issue we aim to highlight. Therefore, it is important to find realistic solutions for smaller sample sizes. Potential solutions for small sample sizes include identifying more homogeneous subgroup of patients for whom predicting survival may be more straightforward, correcting for small sample bias (32) or collaborating with other ALS research groups (recent examples are the PRECISION-ALS (35) or ALS Natural History cohorts (36). We would like to advocate for the latter not just because it allows to include all kinds of ALS patients but also because this could result in more robust and replicable studies. Collaborations may also allow for the development and use of more advanced analysis methods (36,37).

Time of biospecimen sampling. Third, when we divided the patients based on when their

biospecimen was sampled since diagnosis, we used the same follow-up period of two years for both groups. It is important to acknowledge that the sample sizes ($n = 32$ and $n = 44$) are too small for a meaningful analysis; while they serve to illustrate our methodological point, analyses with such small sample sizes should not be used to derive clinical insights. Despite this limitation, using biomarkers measured closer to diagnosis (within three months) or later (three months or more) as the exposure yielded different results, highlighting the potential influence of sampling time on observed associations.

This discrepancy might be attributable to several reasons. First, biomarkers measured close to diagnosis might indeed be more predictive of immediate risk of death during the year to follow, compared to biomarkers measured more than three months after ALS diagnosis. Second, patients that had a measurement of biomarkers at least three months after diagnosis would have to have survived at least three months after diagnosis while it is not necessary for patients that had a measurement within three months of diagnosis. Therefore, there is an interaction between progression rate and diagnostic delay (which directly impacts biospecimen sampling). The fact that the confidence intervals are smaller in the group with a measurement further away from diagnosis could be due to greater homogeneity in this group, consisting of patients who survived longer and may experience slower progression. Having a protocol that harmonizes what data should be collected when may reduce heterogeneity arising from a long sampling period.

Variable selection. Finally, we compared five models differing only by one covariable. For example, one adjusted for ALSFRS-R score and diagnosis delay while another used progression rate or further adjusting for site of onset and BMI. The choice could depend on the research question or what is available in the dataset or registry, among others. In our real-data example, these alternatives led to only slight changes in the estimates and the confidence intervals when models only differed by one variable. Even so, a

Table 3. Checklist for transparent reporting of survival analysis in ALS.

Checklist item	Details	Suggested reading/reference
Study Design and Sample	Clearly define inclusion/exclusion criteria.	
	Describe clinical characteristics of the cohort (e.g. site of onset, age, sex, genetic status).	
	Specify recruitment source and sampling time point (e.g. diagnosis, baseline visit).	
	Report total sample size and number of events (not just % deceased).	(7)
Follow-up Definition	Define and justify start of follow-up (e.g. symptom onset, diagnosis, sampling date). Discuss the impact of the starting point.	
	Define end point (e.g. death, tracheostomy, permanent assisted ventilation). Consider the consequences of composite endpoints	(7)
	Report median follow-up time and range.	
	Clarify censoring rules and handling of lost-to-follow-up cases.	(8)
Covariate Selection and Justification	Provide rationale for covariate inclusion (clinical relevance, prior literature, statistical criteria).	
	Address potential collinearity, overfitting and EPV.	
	Consider time-dependent covariates if applicable.	
Statistical Methods	Specify model type (e.g. Cox proportional hazards, parametric models).	For an overview, see Supplementary Table 5
	Report test point estimates with confidence intervals.	
	Report assumption check(s).	
	Consider alternative metrics (e.g. restricted mean survival time) when PH assumption is violated.	(38)
Sample Size and Power	Include power calculations or justification for sample size.	(7)
	Discuss implications of event count on model stability and statistical power.	
	Apply correction for small samples if appropriate.	(31) (32)
	Report total sample size but also number of events used in each model.	(7)
Reporting and Reproducibility	Provide a clear description of all analytical steps.	
	Share code or analysis scripts (if possible).	
	Present results in table or forest plots (include estimates in a supplementary table)	
	Encourage replication in independent cohorts.	

slightly broader confidence interval could lead to a non-significant effect, which in turn results in a potentially different conclusion. Moreover, results could differ substantially when models are more dissimilar or when the sample size is smaller. Finally, it is impossible to know in advance how these changes will affect a multivariable model. Hence, researchers should carefully consider which variables to include in their study and unambigously report which variables were included in each model and why.

In addition, variable selection in ALS research presents a tradeoff: it must balance the risk of low EPV, which can lead to overfitting or unstable models, against the need to adjust for important clinical characteristics that account for sample heterogeneity. A model with fewer variables may mitigate EPV-related issues but at the cost of omitting relevant covariates, potentially resulting in biased or spurious findings.

Reporting and transparency. Throughout this study, we have emphasized that methodological decisions, including follow-up duration, time of sampling, and covariate selection, can significantly influence survival analysis results. However, sometimes published ALS survival studies do not clearly report key design elements such as the number of events, the rationale for covariate selection, or the definition of follow-up start and end points. In our review of the ALS survival studies mentioned in the Introduction, we observed that such details were occasionally missing or inconsistently reported. We believe that incomplete reporting due to a lack of reporting guidelines may partly explain the variability in findings across the literature. To support more transparent and reproducible research, we have included a checklist in Table 3 summarizing key methodological considerations, along with references to guide researchers in making and reporting these decisions.

Conclusion

Due to the complexities of analyzing survival data in progressive diseases, ALS researchers must exercise caution when interpreting and generalizing results. Observed differences between groups could be due to true biological effects (e.g. a biomarker might be more relevant to an event soon after measurement), variations in study populations, analytical choices, or even chance alone.

Differences in biomarker results between CSF and blood may reflect underlying biological factors such as compartment-specific immune activity (39) or the lower number of immune cells in CSF which can amplify measurement variability. These biological characteristics may also influence the effectiveness or interpretability of certain statistical methods, including Cox regression, particularly in contexts with small sample sizes, limited events, or time-dependent effects that challenge the proportional hazards assumption.

We hypothesize that some of this lack of reproducibility in ALS may be attributed to the methodological difficulties such as the ones discussed in this study. In particular, the number of events rather than just the overall sample size, is a critical determinant of model stability and statistical power. Yet, this information is often underreported in the literature, making it difficult to assess the robustness of published findings.

We therefore emphasize the importance of thoughtful study design and advocate for greater transparency in reporting of ALS survival analysis. This includes clearly outlining study design, data collection methods, and specific statistical analyses including the rationale behind covariate selection. Additionally, ALS research would benefit from the development of field-specific guidelines for conducting and reporting survival analyses, as well as from more replication studies. Replication of findings in independent studies is crucial to establish robust and reliable conclusions. Ultimately, rigorous methodology and transparent reporting are essential tools for advancing our understanding of ALS and developing effective therapies for this devastating disease.

Acknowledgements

We would like to thank the two anonymous reviewers for their thoughtful comments and feedback on the previous versions of this tutorial.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Funding

This is methodological article received no funding, and data was used only to illustrate methodological issues in survival analysis.

ORCID

Solmaz Yazdani http://orcid.org/0000-0001-8729-394X
John Andersson http://orcid.org/0000-0003-2799-6349
Caroline Ingre http://orcid.org/0000-0001-5327-7204
Fang Fang http://orcid.org/0000-0002-3310-6456

Anikó Lovik http://orcid.org/0000-0002-6397-5011

Data availability

Data from the MND Registry and the ALSrisc study cannot be shared publicly due to sensitive patient information and may not leave the secure servers where they are kept. Interested researchers should contact Prof. Fang Fang or Prof. Caroline Ingre to discuss conditions for accessing data (which may include obtaining ethical approval).

References

1. Longinetti E, Regodón Wallin A, Samuelsson K, Press R, Zachau A, Ronnevi L-O, et al. The Swedish motor neuron disease quality registry. Amyotroph Lateral Scler Frontotemporal Degener. 2018;19:528–37.
2. Rojas P, Ramírez AI, Fernández-Albarral JA, López-Cuenca I, Salobrar-García E, Cadena M, et al. Amyotrophic lateral sclerosis: a neurodegenerative motor neuron disease with ocular involvement. Front Neurosci. 2020;14:566858.
3. Imrell S, Fang F, Ingre C, Sennfält S. Increased incidence of motor neuron disease in Sweden: a population-based study during 2002–2021. J Neurol. 2024;271:2730–5.
4. Al-Chalabi A, Hardiman O, Kiernan MC, Chiò A, Rix-Brooks B, van den Berg LH, et al. Amyotrophic lateral sclerosis: moving towards a new classification system. Lancet Neurol. 2016;15:1182–94.
5. Logroscino G, Urso D, Tortelli R. The challenge of amyotrophic lateral sclerosis descriptive epidemiology: to estimate low incidence rates across complex phenotypes in different geographic areas. Curr Opin Neurol. 2022;35:678–85.
6. Schober P, Vetter TR. Survival analysis and interpretation of time-to-event data: the tortoise and the hare. Anesth Analg. 2018;127:792–8.
7. Gordon PH, Corcia P, Lacomblez L, Pochigaeva K, Abitbol J-L, Cudkowicz M, et al. Defining survival as an outcome measure in amyotrophic lateral sclerosis. Arch Neurol. 2009;66:758–61.
8. van Eijk RPA, Nikolakopoulos S, Roes KCB, Middelkoop BM, Ferguson TA, Shaw PJ, et al. Critical design considerations for time-to-event endpoints in amyotrophic lateral sclerosis clinical trials. J Neurol Neurosurg Psychiatry. 2019;90:1331–7.
9. Lu C-H, Macdonald-Wallis C, Gray E, Pearce N, Petzold A, Norgren N, et al. Neurofilament light chain: a prognostic biomarker in amyotrophic lateral sclerosis. Neurol. 2015;84:2247–57.
10. Verde F, Steinacker P, Weishaupt JH, Kassubek J, Oeckl P, Halbgebauer S, et al. Neurofilament light chain in serum for the diagnosis of amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2019;90:157–64.
11. Thouvenot E, Demattei C, Lehmann S, Maceski-Maleska A, Hirtz C, Juntas-Morales R, et al. Serum neurofilament light chain at time of diagnosis is an independent prognostic factor of survival in amyotrophic lateral sclerosis. Eur J Neurol. 2020;27:251–7.
12. De Schaepdryver M, Lunetta C, Tarlarini C, Mosca L, Chio A, Van Damme P, et al. Neurofilament light chain and C reactive protein explored as predictors of survival in amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 2020;91:436–7.
13. Cox DR. Regression models and life-tables. J R Stat Soc B. 1972;34:187–202.
14. Collett D. Modelling survival data in medical research. New York: Chapman and Hall/CRC; 2014.
15. Hernán MA. The hazards of hazard ratios. Epidemiology. 2010;21:13–5.
16. Hiller L, Marshall A, Dunn J. Assessing violations of the proportional hazards assumption in Cox regression: does the chosen method matter? Trials. 2015;16:P134.
17. Staats KA, Borchelt DR, Tansey MG, Wymer J. Blood-based biomarkers of inflammation in amyotrophic lateral sclerosis. Mol Neurodegener. 2022;17:11.
18. Vijayakumar UG, Milla V, Cynthia Stafford MY, Bjourson AJ, Duddy W, Duguez SM-R, et al. A systematic review of suggested molecular strata, biomarkers and their tissue sources in ALS. Front Neurol. 2019;10:400.
19. Irwin KE, Sheth U, Wong PC, Gendron TF. Fluid biomarkers for amyotrophic lateral sclerosis: a review. Mol Neurodegener. 2024;19:9.
20. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med. 2019;38:1276–96.
21. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49:1373–9.
22. Kleinbaum DG, Klein M. Survival analysis: a self-learning text. New York, NY: Springer; 2012.
23. Yazdani S, Seitz C, Cui C, Lovik A, Pan L, Piehl F, et al. T cell responses at diagnosis of amyotrophic lateral sclerosis predict disease progression. Nat Commun. 2022;13:6733.
24. Wickham H, François R, Henry L, Müller K, Vaughan D. dplyr: a grammar of data manipulation. 2023; Available at: https://cran.r-project.org/web/packages/dplyr/index.html
25. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. JOSS. 2019;4:1686.
26. Wickham H. ggplot2. WIREs Comput Stats. 2011;3:180–5.
27. Therneau TM. survival: survival analysis; 2024. Available at: https://cran.r-project.org/web/packages/survival/index.html
28. Boessenkool B. berryFunctions: function collection related to plotting and hydrology; 2024. Available at: https://cran.r-project.org/web/packages/berryFunctions/index.html
29. Kassambara A, Kosinski M, Biecek P. survminer: drawing survival curves using 'ggplot2'. R package version 0.5.0; 2024. Available at: https://rpkgs.datanovia.com/survminer/index.html.10.32614/CRAN.package.survminer
30. Chourpiliadis C, Seitz C, Lovik A, Joyce EE, Pan L, Hu Y, et al. Lifestyle and medical conditions in relation to ALS risk and progression – an introduction to the Swedish ALSrisc Study. J Neurol. 2024;271:5447–59.
31. Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80:27–38.
32. Jóźwiak K, Nguyen VH, Sollfrank L, Linn SC, Hauptmann M. Cox proportional hazards regression in small studies of predictive biomarkers. Sci Rep. 2024;14:14232.
33. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. J Clin Epidemiol. 2016;76:175–82.
34. Austin PC, Allignol A, Fine JP. The number of primary events per variable affects estimation of the subdistribution hazard competing risks model. J Clin Epidemiol. 2017;83:75–84.
35. Caravaca Puchades A, McDonough HE, Al-Chalabi A, Chiò A, Corcia P, Galvin M, et al. Mapping the natural history of amyotrophic lateral sclerosis: time-to-event analysis of clinical milestones in the pan-European, population-based PRECISION-ALS cohort. Amyotroph Lateral Scler Frontotemporal Degener. 2025;26:8–19.

36. Schneck D, Arguedas A, Xenopoulos-Oddsson A, Arcila-Londono X, Lunetta C, Wymer J, et al. Time-to-event prediction in ALS using a landmark modeling approach, using the ALS Natural History Consortium dataset. Amyotroph Lateral Scler Frontotemporal Degener. 2025; 26:417–25.
37. Arguedas A, Schneck D, Cui E, Xenopoulos-Oddsson A, Arcila-Londono X, Lunetta C, et al. Risk prediction for ALS using semi-competing risk models with applications to the ALS Natural History Consortium dataset. Amyotroph Lateral Scler Frontotemporal Degener. 2025; 26:550–7.
38. Li Y, Ross JL, Smith AM, Miller DP. Restricted mean survival time estimate using covariate adjusted pseudovalue regression to improve precision. arXiv preprint arXiv:2208.04495. 2022; Aug 9.
39. Yazdani S, Lovik A, Seitz C, Ingre C, Fang F, Andersson J, et al. T cell subset composition differs between blood and cerebrospinal fluid in amyotrophic lateral sclerosis. Clin Immunol. 2024;265:110270.