

Initial setup and analysis options

Several 'pre-set' analyses can be performed across these buttons



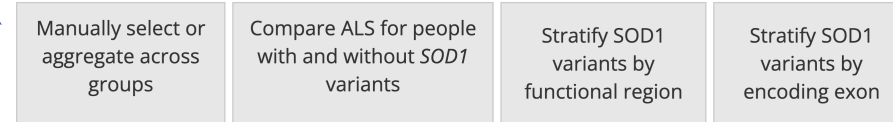
About

Welcome to the SOD1-ALS-Browser website. This tool allows analysis of trends in the clinical presentation of amyotrophic lateral sclerosis (ALS) across user-defined disease subgroups. It provides access to a large built-in dataset of people with and without mutations in SOD1.

Users are invited to apply these data, alone or with their own supplemental dataset, within the customisable analysis protocol available here. Several pre-defined analyses can be performed using the buttons below and user-defined analysis groups can be specified using individual SOD1 variants or by aggregating across multiple variants with the 'manually select or aggregate across groups' option. A non-SOD1 comparator group can be included in these analyses, along with any additional groups from the supplemental data.

We emphasise that since the quantity of data varies greatly by variant, robust analysis may require aggregation across select subgroups. While we hope that this tool is useful for research purposes, the results of analyses performed should not be interpreted as a reliable prognostic indicator for individuals living with or at risk of developing ALS.

Define analysis strata:

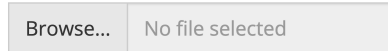


Set optional parameters (click to show options):

- [Dataset filtering](#)
- [Cox analysis configuration](#)
- [Customise figures](#)

Provide supplementary data

Import additional data in .csv format:



- Realign amino acid sequence of native dataset (details in data formatting guide)
- Append '_user' flag to records from supplementary dataset (i.e. distinguish between supplementary and native datasets)

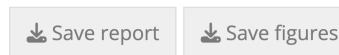
[Data formatting guide](#)

Tutorial

Tutorial slides, available [here](#), give an overview of analysis within this tool.

Save current analysis

The results of the most recent analysis can be downloaded here as a .html format report. The figures can also be downloaded as separate files with adjustable formatting and file type (default .pdf).



[Adjust formatting for saved figures](#)

Click here to display the options for a custom analysis (see "Running analysis ..." slides)

Various settings can be adjusted in this section. Click the buttons to see more!

Covariates used in multivariate analysis are adjusted here. Associations between analysis groups and available covariates can be tested once analysis groups

You can also select whether or not to use robust standard errors for Cox Proportional-Hazards models. Using robust standard errors ensures that variance estimates are accurate when the hazard ratio changes over time (achieved through bootstrapping)

Supplementary data can be included within analyses, and should be uploaded here. Instructions for formatting are provided on the site

SOD1 sequence numbering can be toggled between the old and new nomenclature (including or excluding methionine at the start of the sequence)

Running analysis across individual variants

Use these buttons to choose between analysis across individual variants (this slide) or across variant groups (see next)

Select strata (groups) for survival analysis

Compare individual variants Compare groups of variants

Define all variants and then press 'run analysis' to perform survival analysis.

The reference SOD1 amino acid sequence of the dataset integrated within this app is aligned with methionine as the first amino acid by default. Examples of SOD1 variants often linked to ALS include: A5V, D91A, and I114T. The dataset can be realigned to exclude the first methionine by checking the 'Realign amino acid sequence...' field to the right. The prior variant examples would then be coded: A4V, D90A, I113T.

Variants available for analysis can be selected here

Select or paste comma separated list of sampled variants

G94A G94C G94D G94R G94S G94V

Include an 'any other SOD1 variant' group

Click here to include all people harbouring any non-selected SOD1 variant within a single group

Run analysis

Once variants are selected, click here to run the analysis

Use these buttons to choose between analysis across individual variants (previous slide) or across variant groups (this slide)

Select strata (groups) for survival analysis

Compare individual variants Compare groups of variants

Define all variants within a group and press 'define group'. Press 'run analysis' once all groups are defined to perform survival analysis.

The reference SOD1 amino acid sequence of the dataset integrated within this app is aligned with methionine as the first amino acid by default. Examples of SOD1 variants often linked to ALS include: A5V, D91A, and I114T. The dataset can be realigned to exclude the first methionine by checking the 'Realign amino acid sequence...' field to the right. The prior variant examples would then be coded: A4V, D90A, I113T.

Declare the variants selected below as a new group or reset all groups with these buttons

Define group

Reset groups

Groups can be named here (otherwise are called Group 1,..., N)

Provide name for group (optional)

Any p.G94 variant

Select or paste comma separated list of sampled variants

G94A G94C G94D G94R G94S G94V

Variants to be included within a new group can be selected here

Include an 'any other SOD1 variant' group

Click here to include all people harbouring any non-selected SOD1 variant within a single group

Summary of variant groups

No variant groups have been defined, please select variants and press 'define group'.

Groups that have been already defined are summarised in a table shown here (see next slide for example)

Click here when all groups are defined to run the analysis

Run analysis

Example analysis groups

Select strata (groups) for survival analysis

Compare individual variants Compare groups of variants

Define all variants within a group and press 'define group'. Press 'run analysis' once all groups are defined to perform survival analysis.

The reference SOD1 amino acid sequence of the dataset integrated within this app is aligned with methionine as the first amino acid by default. Examples of SOD1 variants often linked to ALS include: A5V, D91A, and I114T. The dataset can be realigned to exclude the first methionine by checking the 'Realign amino acid sequence...' field to the right. The prior variant examples would then be coded: A4V, D90A, I113T.

Define group Reset groups

Provide name for group (optional)

Enter name

Select or paste comma separated list of sampled variants

Include an 'any other SOD1 variant' group

Summary of variant groups

p.G94/R/S/V	G94R, G94S, G94V
p.G94A	G94A
p.G94C	G94C
p.G94D	G94D

Run analysis

These groups will be used for the subsequent analysis example

They will be compared additionally to the 'OtherVariant' group

Example analysis: data summaries

Each row of this table shows descriptive statistics for one of the analysis groups (strata). See below the table for more details.

Strata	Total sample size	Records with age of onset	Records with disease duration [N censored]	Age of onset in years			Disease duration in months				
				Quartiles (0% 25% median (50%) 75% 100%)	Mean [SD / SE]	Median estimate [95% CI]	Quartiles in people not-censored (0% 25% median (50%) 75% 100%)	Mean in people not-censored [SD]	Restricted mean estimate [SE]	Median estimate [95% CI]	
OtherVariant	1320	1252	1034 [248]	13 41 49 57 85.52	49.07 [12.65 / 0.36]	49 [48, 50]	1 12 22.62 66 564.01	50 [64.7]	88.43 [4.83]	37.59 [30, 44]	
p.G94A	27	27	26 [1]	19 37 48 61 78	48.7 [16.77 / 3.17]	48 [43, 61]	12 16 22 32 172	33.44 [33.42]	33.44 [6.55]	22 [19, 32]	
p.G94C	14	14	9 [4]	31 34.25 37.5 44.88 65.98	40.46 [9.4 / 2.42]	37.5 [35, 51]	19.88 37.49 42.84 49.18 235.4	76.96 [89.24]	221.68 [86.97]	235.4 [42.84, NA]	
p.G94D	15	15	14 [5]	17 40.5 51 58.98 74	49.73 [15.17 / 3.78]	51 [45, 63]	9 27.04 46 74.88 94	49.4 [30.13]	55.56 [8.64]	53 [31, NA]	
p.G94R/S/V	7	7	3 [0]	30 35.5 39 44 46	39.14 [5.93 / 2.07]	39 [34, NA]	22 38.5 55 131 207	94.67 [98.67]	94.67 [46.51]	55 [22, NA]	

Overview of analysis strata. Descriptive statistics are provided based on all raw data in age of onset analysis and people who are not censored in disease duration analysis. 'Estimated' values for median and mean are also shown based on the survival curve calculated by survfit, which takes into account any censoring in data. SE (standard error) and 95% CI (confidence intervals) pertain to estimated median/restricted means from the survival analysis. SD (standard deviation) and quartiles are associated with the raw data.

If your analysis includes strata which aggregate across subgroups, you can inspect boxplots in this section to understand the extent to which the survival curves differ across these subgroups.

Some quick comparisons between the strata and covariates available for use in Cox Proportional-Hazards models can be displayed by clicking here (the comparisons are shown in the tables/figures below)

Show median and restricted mean estimates for analysis of subgroups within analysis strata
View association between analysis strata and available Cox model covariates

Inferential statistics

	Test	Statistic	P-value
Diagnosis	Fisher's exact	-	0.2289
Family history	Fisher's exact	-	0.1064
Sex	Fisher's exact	-	0.01
Site of onset	Fisher's exact	-	0.8561
Region of origin	Fisher's exact	-	5e-04
Age of onset	ANOVA	F = 2.64	0.0325

Tests of difference between possible model covariates and strata of current survival analysis

Descriptive overview (select covariate)

Site of onset

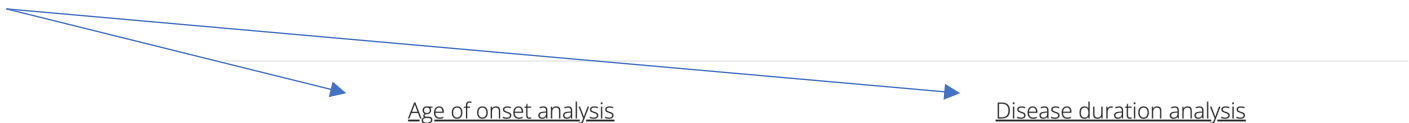
	Spinal	Bulbar	Mixed	Respiratory
OtherVariant	972	107	8	8
p.G94A	25	1	0	0
p.G94C	8	0	0	0
p.G94D	15	0	0	0
p.G94R/S/V	6	0	0	0

Crosstabulation with survival analysis strata

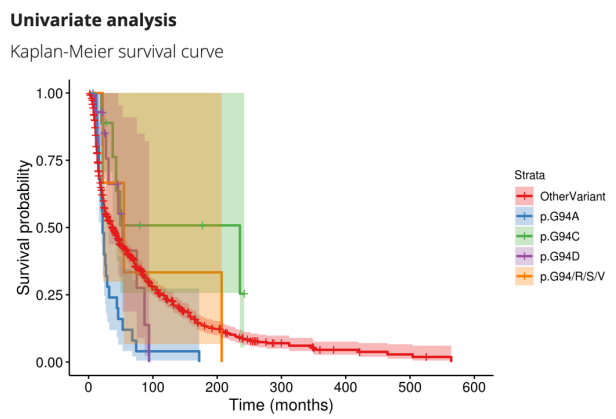
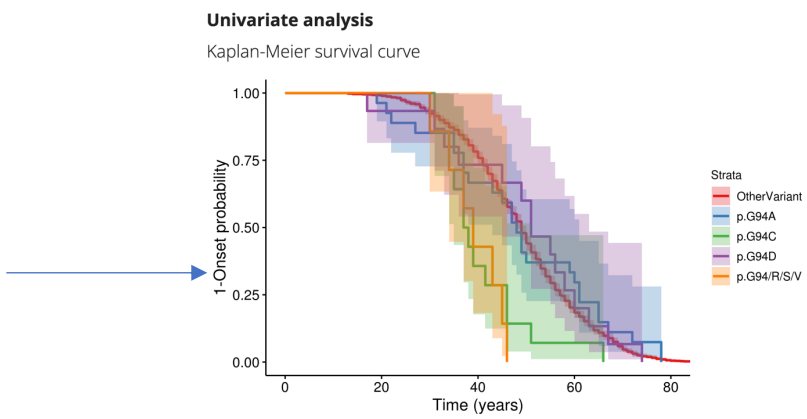
You can choose which variable to inspect from this menu

Example analysis: univariate results

Analyses using survival analysis methodologies are shown across 2 columns. Age of symptom onset on the left, and disease duration (from onset until death or censoring) on the right



Kaplan-Meier plots display the survival curve for each analysis group



A significant log-rank test ($p < 0.5$) suggests a difference in the median age of onset (or disease duration) across the analysis groups

Differences between strata
Global log-rank test: Chi-Squared = 23.7, $p = 9.17e-05$

Strata	Unadjusted Hazard Ratio [95% Confidence Interval]
OtherVariant	-
p.G94A	0.857 [0.561, 1.31]
p.G94C	2.4 [1.21, 4.76]
p.G94D	0.891 [0.57, 1.39]
p.G94/R/S/V	3.49 [2.27, 5.36]

Unadjusted hazard ratio for age of onset in analysis strata relative to the reference group, within Cox Proportional Hazards model excluding covariates.

Differences between strata
Global log-rank test: Chi-Squared = 11.58, $p = 0.0208$

Strata	Unadjusted Hazard Ratio [95% Confidence Interval]
OtherVariant	-
p.G94A	1.75 [1.35, 2.28]
p.G94C	0.446 [0.197, 1.01]
p.G94D	0.89 [0.589, 1.35]
p.G94/R/S/V	0.85 [0.415, 1.74]

Unadjusted hazard ratio for disease duration in analysis strata relative to the reference group, within Cox Proportional Hazards model excluding covariates.

Here are the hazard ratios from a Cox Proportional-Hazards model including no covariates (relative to the reference group – which has no estimate shown)

Show pairwise log-rank results

If testing 3+ strata, pairwise log-rank tests can be viewed by clicking here

Multivariate analysis
Cox proportional-hazards model
Model formula

```
Surv(`Age of onset`) ~ strata + Sex
```

Select reference category for strata variable in Cox model

OtherVariant

Display advanced options for Cox model

Show pairwise log-rank results

Multivariate analysis
Cox proportional-hazards model
Model formula

```
Surv(`Disease duration`, status) ~ strata + Sex + `Age of
```

Select reference category for strata variable in Cox model

OtherVariant

Display advanced options for Cox model

Example analysis: multivariate results (Cox Proportional-Hazards models)

Summaries of multivariate analyses, Cox Proportional-Hazards models, are next provided for analyses of age of onset (left) and disease duration (right)

You can select the reference group to use in the analyses here. The analysis will re-run which takes a few seconds

Multivariate analysis
Cox proportional-hazards model
Model formula
`Surv(`Age of onset`) ~ strata + Sex`
Select reference category for strata variable in Cox model
OtherVariant
 Display advanced options for Cox model
Model summary Coefficients Raw output

Multivariate analysis
Cox proportional-hazards model
Model formula
`Surv(`Disease duration`, status) ~ strata + Sex + `Age of`
Select reference category for strata variable in Cox model
OtherVariant
 Display advanced options for Cox model
Model summary Coefficients Raw output

The 'model summary' tab contains general statistics about the fit of the model and returns the test of the proportional-hazards assumption

The 'coefficients' tab contains a table summarising all variables in the analysis, along with a forest plot visualising hazard ratios

This tab prints summaries directly from the R output for log-rank and Cox Proportional-Hazards analysis

The 'advanced' options allow you to specify models with time-dependent coefficients or to limit the time-interval to consider in the survival analysis (e.g. only 12 months from symptom onset). More details on these options are within.