THE LANCET
Digital Health

Supplementary appendix 3

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Faghri F, Brunn F, Dadu A, et al. Identifying and predicting amyotrophic lateral sclerosis clinical subgroups: a population-based machine-learning study. *Lancet Digit Health* 2022; published online March 24. https://doi.org/10.1016/S2589-7500(21)00274-0.

Appendix Supplementary Materials

Supplemental Methods

The following supplementary methodology is provided to support the study.

The Piedmont and Valle d'Aosta Registry (PARALS)

The Piedmont and Valle d'Aosta Registry for ALS (PARALS) has been in continuous operation since 1995. Details of the registry have been described elsewhere.[1,2,3] Briefly, the registry operates in Piemonte and Valle d'Aosta, two highly industrialized regions of Northwestern Italy. The study area's population was 4,393,609 in 2021, and the two regions encompass 11,066 square miles. The high level of the Italian health services means that patients with ALS are likely to be seen by a hospital neurologist at least once during their disease course. All twenty-six neurology departments of the two regions are involved in the study. In addition, there are two tertiary referral ALS centers, one located in Veruno, at the Salvatore Maugeri Center for Neurologic Rehabilitation, and the other in Turin, at the Department of Neuroscience, University of Turin.[1] One or more investigators act as study referent(s) in each department. All the rehabilitation and geriatrics departments in the region are regularly contacted and are asked to report all possible cases of ALS.

A patient is eligible for inclusion in the registry when they had resided in the study area at least two years before their diagnosis.[1] The diagnosis of ALS is based on the World Federation of Neurology diagnostic criteria[4], but after 2000, patients were also classified according to the revised El Escorial Criteria.[5] Patients are included in the registry if they meet the diagnosis of definite, probable, or probable laboratory-supported ALS at any stage of the disease.[2] When the diagnosis of a case is unclear, it is discussed during regular investigator meetings.[1] Uncertain diagnoses are verified at each follow-up visit. Patients with progressive muscular atrophy (PMA) and primary lateral sclerosis (PLS) variants of ALS are excluded.

The Piedmont regional government recognizes the Piemonte ALS Register as a register of high health interest.[2] Accordingly, the PARALS has the right to access all the existing databases owned by the regional administration and to obtain clinical information about patients with ALS from public and private hospitals and general practitioners.[2] A search is performed every six months in the Piemonte Central Regional Archive to ensure complete case ascertainment.[2] Clinical records of cases identified through the archive with the International Classification of Diseases (ICD-9) code 335.2 (motor neuron disease) are obtained from the hospitals. The relevant clinical information for each case is then analyzed to determine if the patient met the eligibility criteria.[1] Living patients are contacted by phone and visited by one of the neurologists involved in the study. Similar research is performed through the Central Regional Archive of Lombardy, a region bordering Piemonte, to identify cases resident in Piemonte or Valle d'Aosta but receiving their care in Lombardy.[1] An annual search of the Italian Statistical Bureau is also performed for mortality data (ICD-9 code 335.2).[2] Patients' date of death is obtained from the municipality offices where they resided.[2]

The investigators used a standard questionnaire to collect patients' demographic information, clinical history, neurologic and laboratory findings, ALS functional rating scale, and details of treatments.[1,2] Diagnostic electromyographic (EMG) examination is performed according to standard procedures. A clinical follow-up is performed at regular intervals (at least every 3 to 6 months).

Emilia Romagna Region Registry for ALS (ERRALS)

The Emilia Romagna Region Registry for ALS (ERRALS) was established in 2009. Details of the registry are described elsewhere.[6] Briefly, the Emilia Romagna Region consists of nine provinces (with 330 municipalities) and covers an area of 13,987 square miles. The population was 4,395,606 in 2009.[6] The prospective registry collects incident ALS cases among residents in the region using the revised El Escorial diagnostic criteria.[5]

Physicians collect a detailed phenotypic profile for each ALS patient.[6] This information includes age at symptom onset, age at diagnosis, gender, residency, employment history, site of onset, affected body regions, upper and lower motor neuron signs, El Escorial-revised classification, clinical phenotype, the presence of dementia or extrapyramidal signs, family history, diagnostic and laboratory tests (e.g., EMG, MRI), drugs use (including Riluzole), forced vital capacity, ALSFRS-R score, the use of enteral nutrition, the use of non-invasive or invasive ventilatory support, and the date, place and cause of death.

Clinical follow-up is performed at the seventeen neurological departments of the region.[6] These visits are used to collect information on the ALS clinical course, gastrostomy, ventilatory support, and survival. One or more investigators act as study referents in each department. The coordinating center (ALS Centre in Modena) regularly supervises the data to ensure record accuracy. The completeness of case ascertainment is augmented by cases attending the regional hospitals having an ICD-9 discharge code of 335.20, and by death certificates. Moreover, data were matched with data included in the Italian National Registry for Rare Diseases.[6] The clinical records of cases found through these secondary sources are reviewed accordingly.

Ethical approval

The PARALS study was approved by the ethics committees of Azienda Ospedaliera Universitaria City of Health and Science of Turin (number 004462, June 10, 2010). After that date, patients signed written informed consent. A waiver was obtained for patients included in the register before 2010. The Piedmont regional government also recognized PARALS as a 'Registry of High Sanitary Interest' (Regional Law, April 11, 2012, number 4). Accordingly, the PARALS can access databases owned by the regional administration to obtain clinical information about ALS patients from public and private hospitals, and general practitioners. The ERRALS study was approved by the Azienda Ospedaliero Universitaria of Modena (number 124/08, September 2, 2008). Patients enrolled in ERRALS signed written informed consent. All PARALS and ERRALS records were anonymized according to the Italian code for the protection of personal data, and data were treated following the UE 2016/679 General Data Protection Regulation (GDPR).

Supervised subtype prediction – model selection and hyperparameter tuning

The performance of the ensemble learning model was assessed by internal and external validation. The best algorithm and parameters to be used in the model were determined, followed by an evaluation of the effects of the different algorithms, algorithm parameters, and features. In multiple iterations, the cohort data was split into training and validation datasets. Hyperparameters were tuned on the training cohort using three-fold cross-validation. Two cross-validations were used for the optimization, and the third cross-validation was used to validate the results. A grid of possible parameter combinations and a randomized grid search were used to tune the hyperparameters and maximize accuracy.

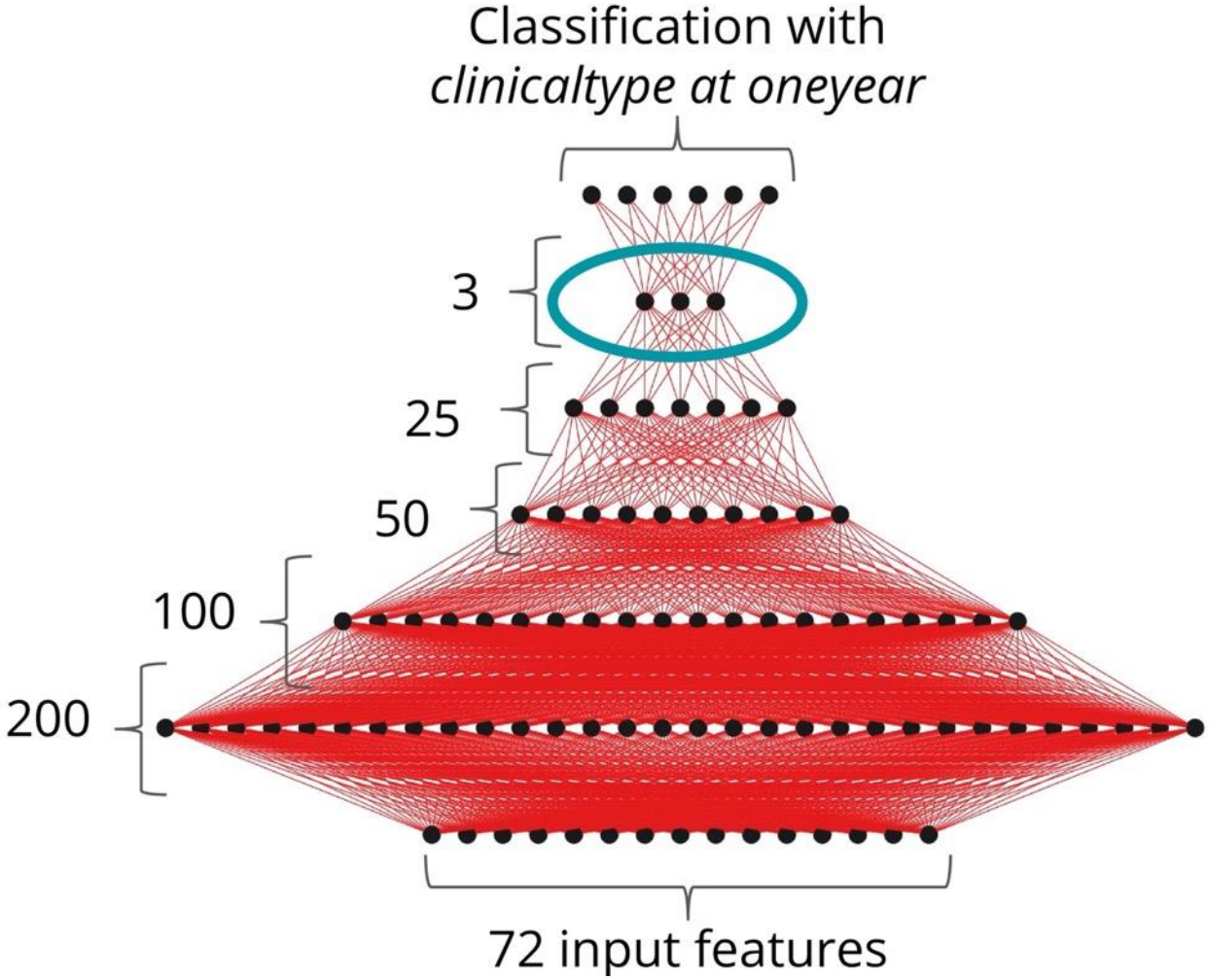
The features included in each iteration were updated using the variable importance obtained from the previous iteration as a probabilistic prior. The best-performing model and features are then passed to the feature evolution stage. For the final external validation, the model was trained using the complete discovery dataset and the optimal hyperparameter combination, and the resulting model was evaluated on the replication set.

References

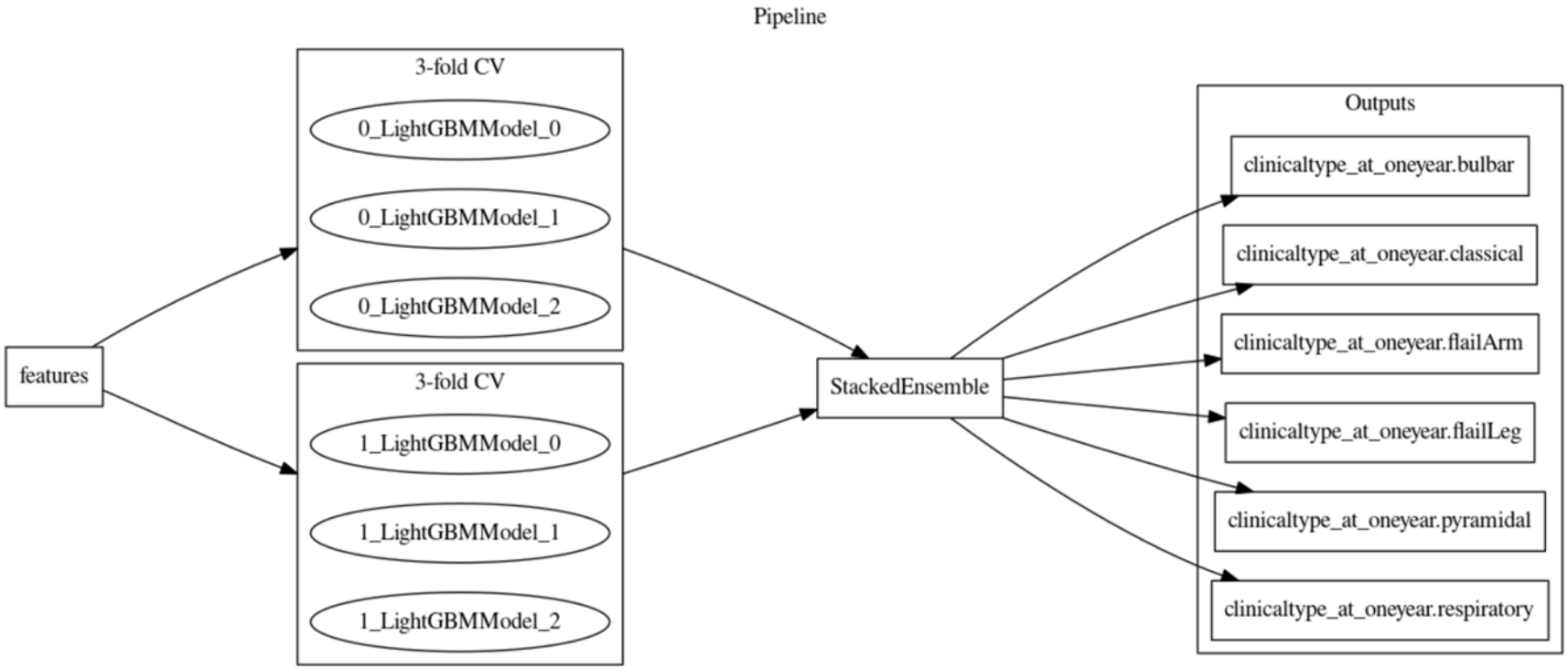
1. Piemonte and Valle d'Aosta Register for Amyotrophic Lateral Sclerosis (PARALS). Incidence of ALS in Italy: evidence for a uniform frequency in Western countries. *Neurology*. 2001; **56**: 239–44
2. Chiò A, Mora G, Moglia C, et al. Secular Trends of Amyotrophic Lateral Sclerosis: The Piemonte and Valle d'Aosta Register. *JAMA Neurol* 2017; **74**: 1097–104
3. Migliaretti G, Berchialla P, Dalmasso P, Cavallo F, Chiò A. Amyotrophic lateral sclerosis in Piedmont (Italy): a Bayesian spatial analysis of the incident cases. Amyotroph Lateral Scler Frontotemporal Degener 2013; 14: 58–65
4. Brooks BR. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. *J Neurol Sci* 1994; 124 Suppl: 96–107
5. Brooks BR, Miller RG, Swash M, Munsat TL. World Federation of Neurology Research Group on Motor Neuron Diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord* 2000; 1: 293–9
6. Mandrioli J, Biguzzi S, Guidi C, et al. Epidemiology of amyotrophic lateral sclerosis in Emilia Romagna Region (Italy): A population-based study. *Amyotroph Lateral Scler Frontotemporal Degener* 2014; **15**: 262–8.

7. Chiò A, Calvo A, Moglia C, et al. Phenotypic heterogeneity of amyotrophic lateral sclerosis: a population-based study. *J Neurol Neurosurg Psychiatry* 2011; **82**: 740–6.

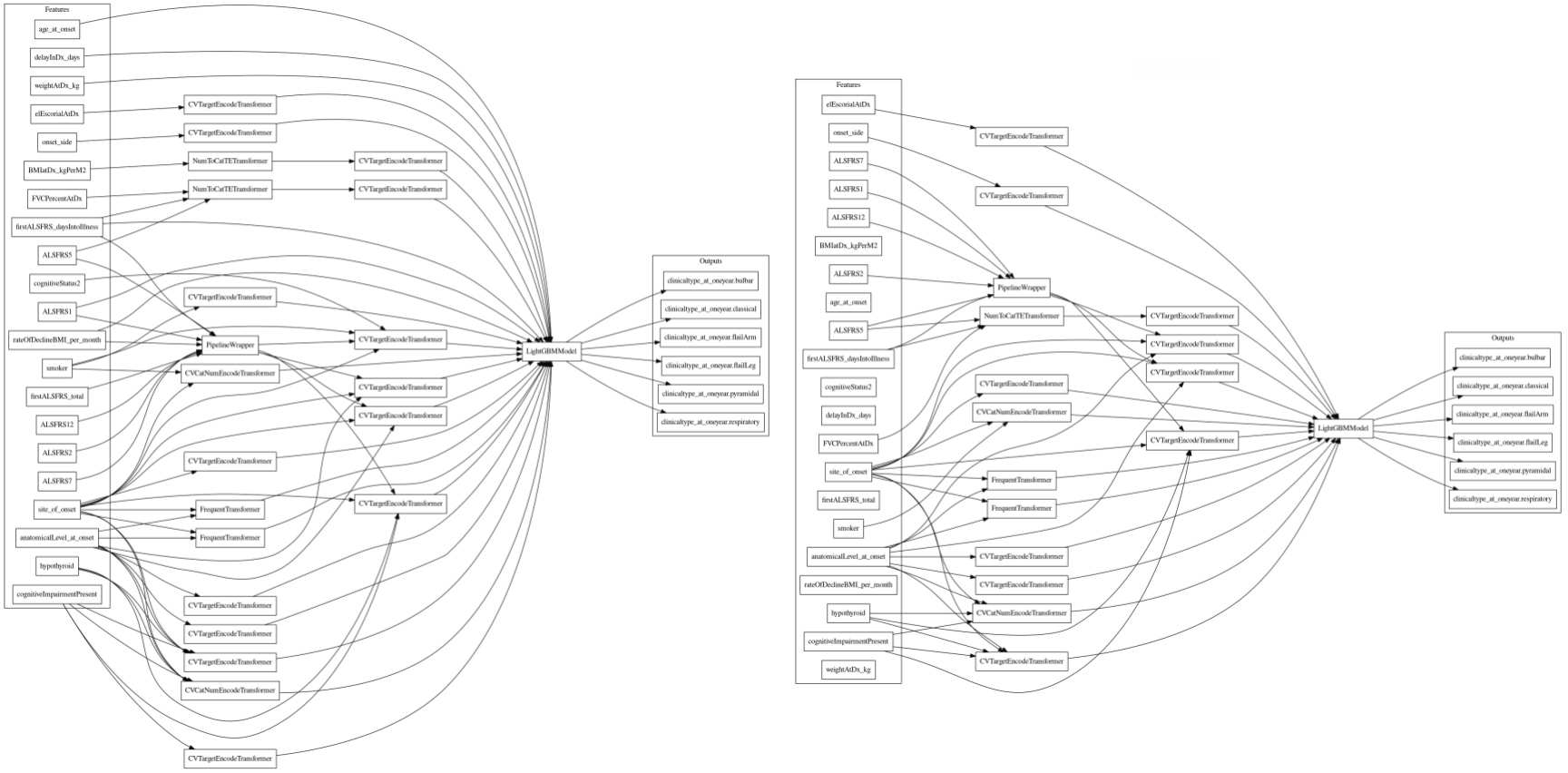
Supplemental Figure 1. The architecture of the multilayer perceptron neural network used for the semi-supervised subtype identification of ALS. The neural network consists of five hidden layers with 200, 100, 50, 25 and 3 neurons. The neural network acts as a dimension reduction technique, compressing the data into 3 dimensions. After training the network with ten-fold cross-validation, the activations of the last hidden layer are used as input for the UMAP algorithm.



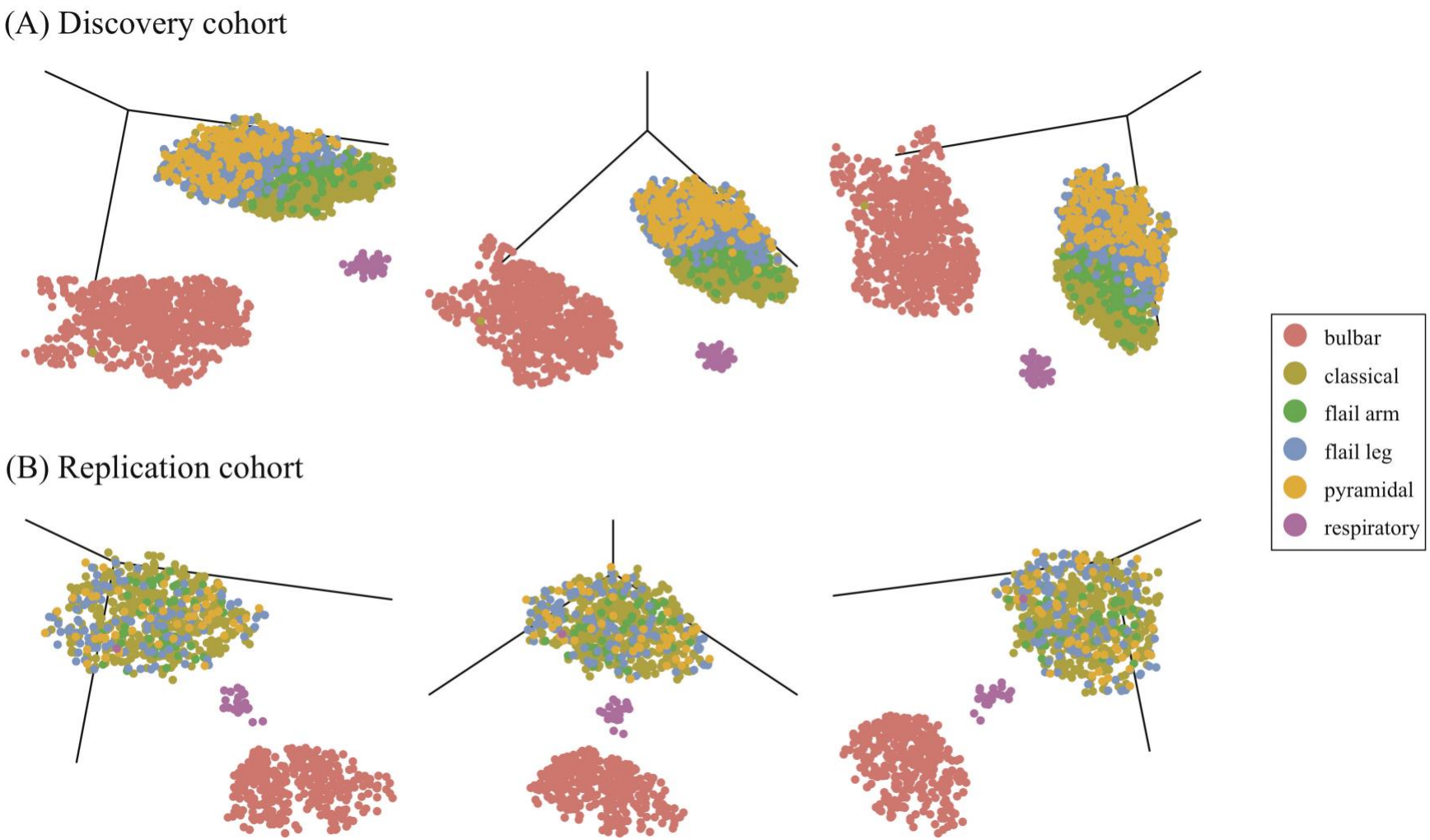
Supplemental Figure 2. Supervised learning model used for ALS subtype identification. The predictive model using all the clinical features was a stacked ensemble. The stacked ensemble parameters were: ensemble_level=2, transforming the 21 original features to 72 transformed features. Each of the three stacked models was fit on three internal holdout splits and were linearly combined. Similarly, the predictive model with top eleven features was a stacked ensemble. The stacked ensemble parameters were: ensemble_level=2, transforming the 11 original features to 34 transformed features. Each of the three stacked models was fit on three internal holdout splits and were linearly combined. The fitted features of the final model were the best features found during the feature engineering iterations.



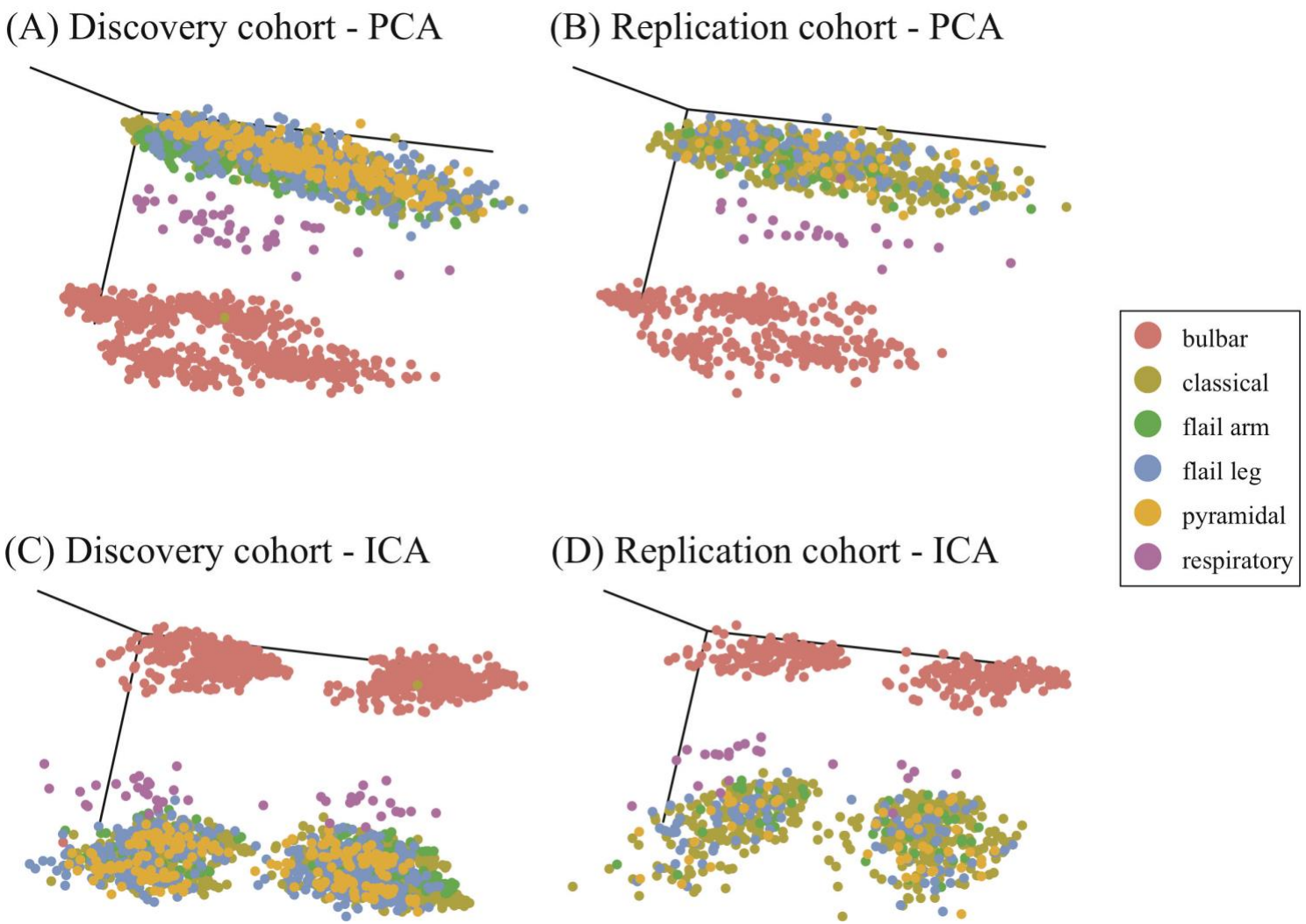
Supplemental Figure 3. Feature Transformation used for the supervised learning model. The effects of different types of algorithms, algorithm parameters, and features were evaluated during the Model and Feature Tuning Stage to determine the best algorithm and parameters to use. The Feature Transformation Stage uses a genetic algorithm to search the large feature engineering space. The graphs below show the Feature Transformations with the best performance from the stacked ensemble model. The left panel refers to model 0, and the right panel refers to model 1 in Figure S2.



Supplemental Figure 4. ALS subtypes identified by machine learning in the discovery and replication cohorts using UMAP alone. The top row (A) shows the three different 3D projections of the discovery cohort defined by the semi-supervised machine learning algorithm consisting of a UMAP algorithm alone. The same 3D projections of the replication cohort are shown in the bottom row (B). Each patient (dot) was color-coded after machine learning cluster generation according to the Chiò classification system.

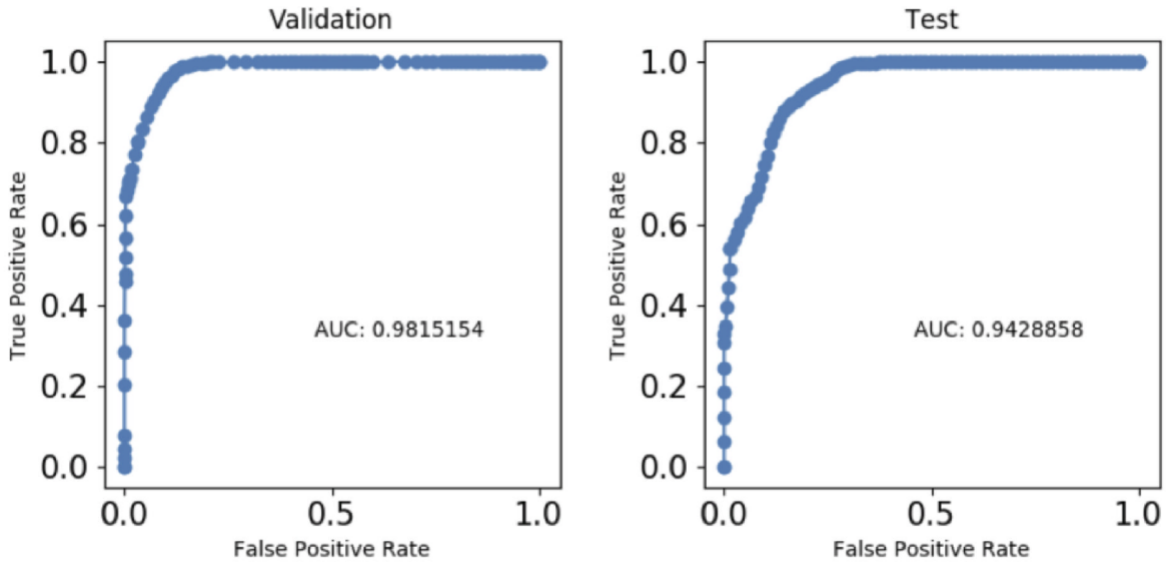


Supplemental Figure 5. ALS clustering using dimension reduction approaches principal component analysis (PCA) and independent component analysis (ICA). The top row shows a 3D projection of the discovery cohort (A) and the replication cohort (B) clustered using PCA. The bottom row shows the 3D projection of the discovery cohort (C) and the replication cohort (D) clustered using ICA. Each patient (dot) was color-coded after machine learning cluster generation according to the Chiò classification system.

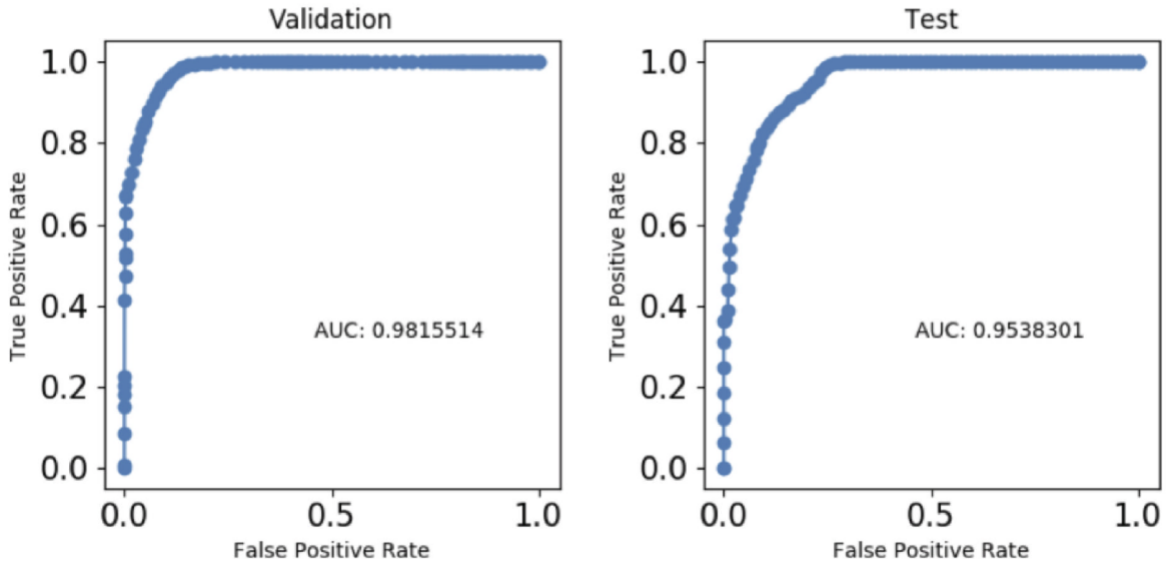


Supplemental Figure 6. The performance of ALS disease subtype prediction models based on receiver operating characteristic (ROC) curves. (A) ROC curves for the model that includes all features. The left panel shows the validation dataset, and the right panel shows the test dataset. (B) ROC curves for the model based on the eleven most important features. The left panel shows the validation dataset, and the right panel shows the test dataset.

(A)

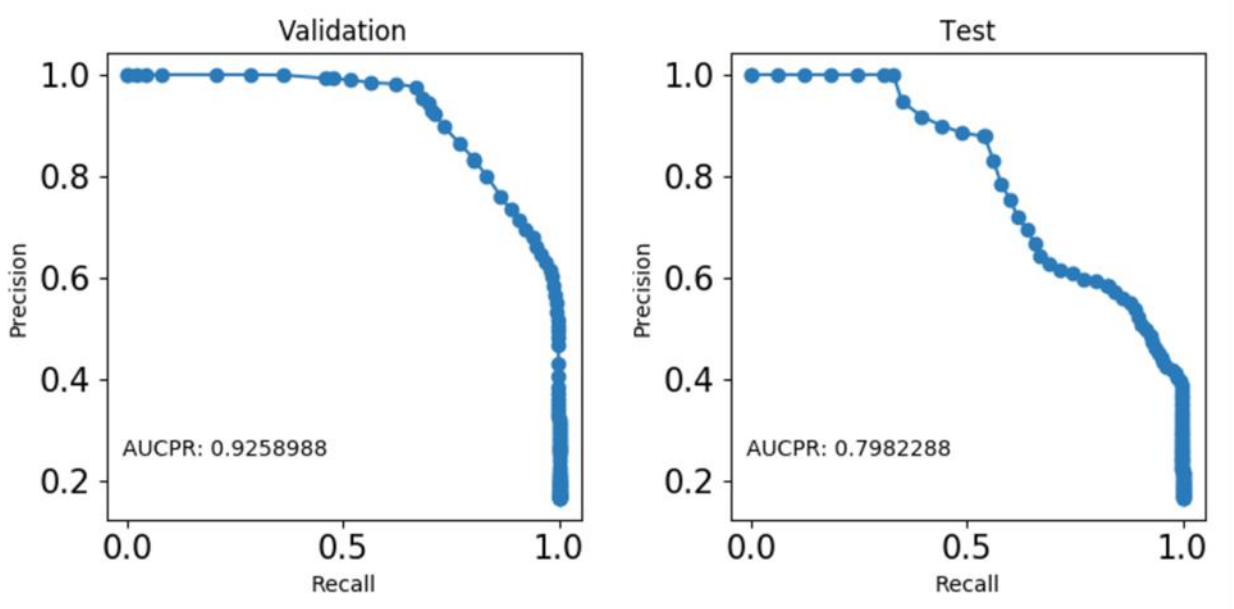


(B)

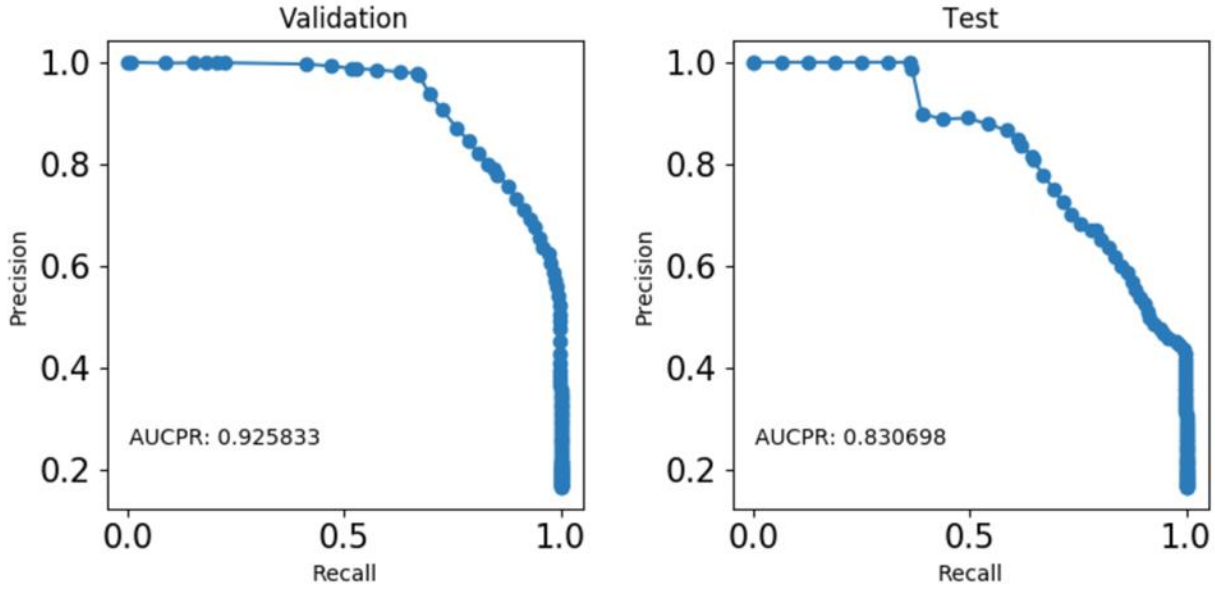


Supplemental Figure 7. Precision-Recall Curve for the performance of ALS disease subtype prediction models. The top row (**A**) shows the precision-recall curve for the prediction model based on all the clinical features. The bottom row (**B**) shows the precision-recall curve for the prediction model based on the eleven most important features.

(A) Predictive model constructed using all clinical features



(B) Predictive model constructed using eleven clinical features



Supplemental Table 1. Clinical parameters of the discovery and replication ALS cohorts used in each stage of analysis.

	Prediction models	
	Unsupervised and semi-supervised	Supervised
Age at diagnosis	Yes	No
Age at symptom onset	Yes	Yes
ALSFRS1	Yes	Yes
ALSFRS2	Yes	No
ALSFRS3	Yes	No
ALSFRS4	Yes	No
ALSFRS5	Yes	No
ALSFRS6	Yes	No
ALSFRS7	Yes	No
ALSFRS8	Yes	No
ALSFRS9	Yes	No
ALSFRS10	Yes	No
ALSFRS11	Yes	No
ALSFRS12	Yes	No
First ALSFRS-R total	Yes	No
Anatomical level at onset	Yes	Yes
Ataxia	Yes	No
BMI 2 years prior to illness	Yes	No
BMI at diagnosis	Yes	No
C9orf72 status	No (data leakage)	No
Cancer	Yes	No
Cancer type	No (irrelevance)	No
Chorea	Yes	No
Clinical type at one year	No (data leakage)	No
Clinical type at onset	No (missingness)	No
Cognitive impairment present	No (missingness)	No
Cognitive status 1	No (missingness)	No
Cognitive status 2	No (missingness)	No
COPD	Yes	No
Delay in diagnosis (days)	Yes	No
Diabetes	Yes	No
Education	Yes	No
El Escorial category at diagnosis	Yes	Yes
El Escorial category at visit 2	No (irrelevance)	No
El Escorial category at visit 3	No (irrelevance)	No
Family history of ALS	Yes	No
FVC percent at diagnosis	Yes	Yes
Height	Yes	No
Hypertension	Yes	No
Hyperthyroid	Yes	No
Hypothyroid	Yes	No
Initial diagnosis was PLS	No (bias)	No
Marital status	Yes	No
Mutated gene	No (bias)	No
Mutation amino acid change	No (bias)	No
Mutation present	No (bias)	No
NIPPV	No (data leakage)	No

Onset side	Yes	Yes
Parkinsonism	Yes	No
PEG inserted	No (data leakage)	No
Place of birth	No	No
Place of residence	No	No
Rate of decline ALSFRS-R (per month)	Yes	No
Rate of decline BMI (per month)	Yes	Yes
Sex	Yes	No
Site of symptom onset	Yes	Yes
Smoker	No (missingness)	Yes
Survival (days)	No (data leakage)	No
Time of first ALSFRS-R (days into illness)	Yes	Yes
Time of NIPPV (days into illness)	No (data leakage)	No
Time of PEG (days into illness)	No (data leakage)	No
Time of tracheostomy (days into illness)	No (data leakage)	No
Tracheostomy	No (data leakage)	No
Vital status	No (data leakage)	No
Weight 2 years prior to illness	Yes	No
Weight at diagnosis	Yes	Yes

ALSFRS-R, ALS Functional Rating Scale – Revised; ALSFRS1 refers to the first question on the scale. ALSFRS2, second question; BMI, body mass index measured in kg per m^2; COPD, chronic obstructive pulmonary disease; PEG, percutaneous endoscopic gastrostomy; NIPPV, non-invasive positive pressure ventilation.

Supplemental Table 2. Demographics and clinical features of the ALS registries analyzed in this study

	PARALS (n = 2,858)	ERRALS (n=1,097)
Age at diagnosis (years)	66.8 (11.1)	67.8 (11.3)
Age at symptom onset (years)	65.9 (11.1)	66.6 (11.4)
ALSFRS1	3.3 (0.9)	3.1 (1.1)
ALSFRS2	3.7 (0.6)	3.5 (0.8)
ALSFRS3	3.4 (0.8)	3.2 (1)
ALSFRS4	3.3 (0.9)	3.2 (1)
ALSFRS5	3.2 (0.9)	2.6 (1.4)
ALSFRS6	3.2 (1)	2.7 (1.2)
ALSFRS7	3.4 (0.9)	3.2 (1.1)
ALSFRS8	3.1 (0.9)	2.9 (1.1)
ALSFRS9	2.7 (1.4)	2.3 (1.5)
ALSFRS10	3.7 (0.7)	3.6 (0.9)
ALSFRS11	3.8 (0.6)	3.7 (0.7)
ALSFRS12	3.9 (0.5)	3.8 (0.6)
First ALSFRS-R total	40.8 (6.4)	38.3 (7.9)
Anatomical level at onset (n)		
bulbar	1,009	376
lower limb distal	908	270
lower limb proximal	115	77
respiratory	48	23
upper limb distal	617	196
upper limb proximal	161	65
Ataxia (n)	1	0
BMI 2 years prior to illness (kg/m2)	26.0 (4.1)	25.8 (3.9)
BMI at diagnosis (kg/m2)	24.5 (4.2)	24.4 (4)
C9orf72 status (n)	100	22
Cancer (n)	136	122
Cancer type (n)		
acoustic neuroma	0	2
adrenal	0	2
adrenal, breast	0	1
adrenal, ovary, thyroid	0	1
bile duct	1	0
bile duct, melanoma	0	1
bladder	12	5
bladder, prostate	0	1
bladder, rectum	0	1
breast	21	22
breast, colon	0	1
breast, melanoma	0	1
breast, myeloma	0	1
breast, ovary	0	2
breast, uterus	0	2
chronic lymphoid leukemia	5	0
chronic myeloid leukemia	2	0
colon	16	9
colon, prostate	0	1
esophagus	3	0
gallbladder	1	0
larynx	5	0
larynx, stomach	0	1
liver	1	0
lung	23	1
lung, thyroid	0	1
lymphoma	4	2
lymphoma, cutaneous	1	0
lymphoma, melanoma	0	1
lymphoma, non-Hodgkin's	1	1

lymphoma, non-Hodgkin's, thyroid	0	1
melanoma	1	0
meningioma	0	2
meningioma, prostate	0	1
mesothelioma	1	0
myeloma	1	2
myeloma, stomach	0	1
myelophthisis	1	0
nasopharyngeal	0	4
ovary	2	1
ovary, uterus	1	1
pancreas	1	3
pituitary	0	4
pharynx	1	0
plasmacytoma	1	0
pleura	2	0
prostate	13	21
prostate, renal	0	1
rectum	3	0
renal	4	3
schwannoma	0	1
skin cancer	0	2
stomach	2	1
thymoma	1	1
thyroid	1	3
uterus	2	8
vulva	1	0
Chorea (n)	1	2
Clinical type at one year (n)		
bulbar	1,009	376
classical	841	462
flail arm	188	52
flail leg	531	109
pyramidal	240	50
respiratory	48	24
Clinical type at onset (n)		
bulbar	893	NA
classical	1,030	NA
LMN predominant	501	NA
UMN predominant	434	NA
Cognitive impairment present (n)	557	97
Cognitive status 1 (n)		
ALSbi	33	NA
ALSci	176	NA
FTD	287	97
non-executive impairment	22	NA
normal	456	1,000
Cognitive status 2 (n)		
ALSbi	61	NA
ALScbi	38	NA
ALSci	129	NA
FTD	171	NA
non-executive impairment	16	NA
normal	380	NA
COPD (n)	202	97
Delay in diagnosis (days)	347.8 (306.4)	423.5 (429.6)
Diabetes (n)	302	85
Education (n)		
less than 5 years	0	47
5 years	1,240	280
8 years	705	260
11 years	184	28
13 years	359	223
18 years	132	51
El Escorial category at diagnosis (n)		222
possible	359	345
probable	880	139
probable–lab supported	169	336

definite	1,335	
El Escorial category at visit 2 (n)		
possible	71	NA
probable	260	NA
probable–lab supported	39	NA
definite	372	NA
El Escorial category at visit 3 (n)		
possible	8	NA
probable	260	NA
probable–lab supported	39	NA
definite	372	NA
Family history of ALS (n)	187	51
FVC percent at diagnosis	84.6 (25.7)	84.5 (24.9)
Height (cm)	164.5 (9.3)	167.9 (8.8)
Hypertension (n)	1,207	473
Hyperthyroid (n)	44	19
Hypothyroid (n)	204	56
Initial diagnosis was PLS (n)	26	0 (0)
Marital status (n)		
divorced	64	10
married	2,034	246
never married	245	25
separated	52	3
widowed	366	49
Mutated gene (n)		
C9orf72	99	22
C9orf72 + TARDBP	1	0
FUS	11	7
OPTN	1	0
SOD1	34	14
TARDBP	21	3
TUBA4A	2	0
Mutation amino acid change	NA	NA
Mutation present (n)	169	46
NIPPV (n)	862	263
Onset side (n)		
both	549	52
left	438	407
right	789	207
Parkinsonism (n)	52	30
PEG inserted (n)	890	233
Place of birth	NA	NA
Place of residence	NA	NA
Rate of decline ALSFRS-R (per month)	0.9 (1.3)	1 (1.4)
Rate of decline BMI (per month)	0.04 (0.1)	0 (0.1)
Sex (female)	1,318	485
Site of symptom onset (n)		
bulbar	1,009	376
respiratory	48	23
spinal	1,801	674
Smoker (n)		
current	406	92
former	640	20
never	957	61
Survival (days)	1,319.6 (1234.5)	1271.7 (964.4)
Time of first ALSFRS-R (days into illness)	402.1 (385.4)	484.1 (430.4)
Time of NIPPV (days into illness)	868.9 (719.6)	733 (491.6)
Time of PEG (days into illness)	833.6 (598.1)	738.3 (453.5)
Time of tracheostomy (days into illness)	907.7 (617)	764.1 (480.2)
Tracheostomy (n)	363	299
Vital status (deceased)	2,549	663

Weight 2 years prior to illness (kg)	70.3 (12.4)	73 (13.4)
Weight at diagnosis (kg)	66.2 (12.8)	69.1 (13.4)

Numerical values are presented as means (standard deviations); All other columns are presented as counts; ALSFRS-R, ALS Functional Rating Scale – Revised; ALSFRS1 refers to the first question on the scale. ALSFRS2, second question; BMI, body mass index measured in kg per m^2; COPD, chronic obstructive pulmonary disease; PEG, percutaneous endoscopic gastrostomy; NIPPV, non-invasive positive pressure ventilation.

Supplementary Table 3. Performance of the models used to predict ALS clinical subtypes. The tables show the metrics used to evaluate the performance of (A) the model based on all of the available clinical features and (B) the model based on only the most important clinical features.

(A) Predictive model based on all available clinical features

Scorer	Optimized	Better score is	Final ensemble scores on validation (internal)	Final ensemble standard deviation on validation (internal)	Final test scores (external)	Final test standard deviation
ACCURACY		higher	0·8150798	0·01275987	0·6439888	0·01293791
AUC		higher	0·9815154	0·002010137	0·9428858	0·003738604
AUCPR		higher	0·9258988	0·00742477	0·7982288	0·01136424
F05		higher	0·8150798	0·01275987	0·6439888	0·01293791
F1		higher	0·8150798	0·01275987	0·6439888	0·01293791
F2		higher	0·8150798	0·01275987	0·6439888	0·01293791
GINI		higher	0·9630309	0·004020274	0·8857717	0·007477209
LOGLOSS	*	lower	0·4072223	0·02122787	0·7757664	0·03125351
MACROAUC		higher	0·9585233	0·003836152	0·8972699	0·006460401
MCC		higher	0·7780957	0·01531184	0·5727866	0·01552549

(B) Predictive model based on most important clinical features

Scorer	Optimized	Better score is	Final ensemble scores on validation (internal)	Final ensemble standard deviation on validation (internal)	Final test scores (external)	Final test standard deviation
ACCURACY		higher	0·8145004	0·009715746	0·7148183	0·01396052
AUC		higher	0·9815514	0·001332268	0·9538301	0·003653627
AUCPR		higher	0·925833	0·005046326	0·830698	0·01184665
F05		higher	0·8145004	0·009715746	0·7148183	0·01396052
F1		higher	0·8145004	0·009715746	0·7148183	0·01396052
F2		higher	0·8145004	0·009715746	0·7148183	0·01396052
GINI		higher	0·9631028	0·002664535	0·9076601	0·007307254
LOGLOSS	*	lower	0·411088	0·0143602	0·7687869	0·03917954
MACROAUC		higher	0·9595809	0·002420295	0·9051199	0·006400558
MCC		higher	0·7774005	0·0116589	0·6577819	0·01675263

Supplementary Table 4. Validation Confusion Matrix of the models used to predict ALS clinical subtypes. The tables show the validation confusion matrix of (A) the model based on all of the available clinical features and (B) the model based on only the most important clinical features.

(A) Predictive model based on all available clinical features

	Predicted: bulbar	**Predicted: classical**	**Predicted: flail arm**	**Predicted: flail leg**	**Predicted: pyramidal**	**Predicted: respiratory**	**error**
Actual: bulbar	1,007	1	0	1	0	0	0%
Actual: classical	1	601	1	223	15	0	29%
Actual: flail arm	0	28	160	0	0	0	15%
Actual: flail leg	0	34	0	470	27	0	11%
Actual: pyramidal	0	11	0	189	40	0	83%
Actual: respiratory	0	0	0	0	0	48	0%

(B) Predictive model based on most important clinical features

	Predicted: bulbar	**Predicted: classical**	**Predicted: flail arm**	**Predicted: flail leg**	**Predicted: pyramidal**	**Predicted: respiratory**	**error**
Actual: bulbar	1,007	1	0	1	0	0	0%
Actual: classical	1	657	1	156	26	0	22%
Actual: flail arm	0	28	160	0	0	0	15%
Actual: flail leg	0	91	0	397	43	0	25%
Actual: pyramidal	0	46	0	137	57	0	76%
Actual: respiratory	0	0	0	0	0	48	0%

Supplementary Table 5. Test Confusion Matrix of the models used to predict ALS clinical subtypes. The tables show the test confusion matrix of (A) the model based on all of the available clinical features and (B) the model based on only the most important clinical features.

(A) Predictive model based on all available clinical features

	Predicted: bulbar	**Predicted: classical**	**Predicted: flail arm**	**Predicted: flail leg**	**Predicted: pyramidal**	**Predicted: respiratory**	**error**
Actual: bulbar	376	0	0	0	0	0	0%
Actual: classical	0	191	122	143	6	0	59%
Actual: flail arm	0	28	23	0	1	0	56%
Actual: flail leg	0	5	26	75	3	0	31%
Actual: pyramidal	0	3	13	31	3	0	94%
Actual: respiratory	0	0	1	0	0	23	4%

(B) Predictive model based on most important clinical features

	Predicted: bulbar	**Predicted: classical**	**Predicted: flail arm**	**Predicted: flail leg**	**Predicted: pyramidal**	**Predicted: respiratory**	**error**
Actual: bulbar	376	0	0	0	0	0	0%
Actual: classical	0	284	51	119	8	0	39%
Actual: flail arm	0	39	13	0	0	0	75%
Actual: flail leg	0	34	0	69	6	0	37%
Actual: pyramidal	0	18	0	30	2	0	96%
Actual: respiratory	0	0	1	0	0	23	4%

Supplementary Table 6. Detailed confusion matrix statistics on validation data of the models used to predict ALS clinical subtypes. The tables show the detailed confusion matrix statistics of (A) the model based on all of the available clinical features and (B) the model based on only the most important clinical features.

(A) Predictive model based on all available clinical features

	Bulbar	Classical	Flail arm	Flail leg	Pyramidal	Respiratory
Threshold (max F1 score)	argmax	argmax	argmax	argmax	argmax	argmax
Population	2,857	2,857	2,857	2,857	2,857	2,857
P: Condition positive	1,009	841	188	531	240	48
N: Condition negative	1,848	2,016	2,669	2,326	2,617	2,809
Test outcome positive	1,008	675	161	883	82	48
Test outcome negative	1,849	2,182	2,696	1,974	2,775	2,809
TP: True Positive	1,007	601	160	470	40	48
TN: True Negative	1,847	1,942	2,668	1,913	2,575	2,809
FP: False Positive	1	74	1	413	42	0
FN: False Negative	2	240	28	61	200	0
TPR: (Sensitivity, hit rate, recall)	0·998017839	0·714625446	0·85106383	0·885122411	0·166666667	1
TNR=SPC: (Specificity)	0·999458874	0·963293651	0·999625328	0·82244196	0·983951089	1
PPV: Pos Pred Value (Precision)	0·999007937	0·89037037	0·99378882	0·532276331	0·487804878	1
NPV: Neg Pred Value	0·998918334	0·890009166	0·989614243	0·969098278	0·927927928	1
FPR: False-out	0·000541126	0·036706349	0·000374672	0·17755804	0·016048911	0
FDR: False Discovery Rate	0·000992063	0·10962963	0·00621118	0·467723669	0·512195122	0
FNR: Miss Rate	0·001982161	0·285374554	0·14893617	0·114877589	0·833333333	0
ACC: Accuracy	0·998949947	0·890094505	0·989849492	0·834091705	0·915295765	1
F1 score	0·998512643	0·792875989	0·916905444	0·664780764	0·248447205	1
MCC: Matthew's correlation coefficient	0·997701467	0·727347366	0·914642183	0·595612969	0·250233375	1
Informedness	0·997476714	0·677919097	0·850689158	0·707564371	0·150617756	1
Markedness	0·997926271	0·780379536	0·983403063	0·501374608	0·415732806	1
Prevalence	0·353167658	0·294364718	0·06580329	0·185859293	0·0840042	0·01680084
LR+: Positive likelihood ratio	1844·336967	19·46871485	2271·489362	4·984975126	10·38492063	inf
LR-: Negative likelihood ratio	0·001983234	0·296248765	0·148991993	0·139678658	0·846925566	0
DOR: Diagnostic odds ratio	929964·5	65·71745495	15245·71429	35·68888183	12·26190476	inf
FOR: False omission rate	0·001081666	0·109990834	0·010385757	0·030901722	0·072072072	0

(B) Predictive model based on most important clinical features

	Bulbar	Classical	Flail arm	Flail leg	Pyramidal	Respiratory
Threshold (max F1 score)	argmax	argmax	argmax	argmax	argmax	argmax
Population	2,857	2,857	2,857	2,857	2,857	2,857
P: Condition positive	1,009	841	188	531	240	48
N: Condition negative	1,848	2,016	2,669	2,326	2,617	2,809
Test outcome positive	1,008	823	161	691	126	48
Test outcome negative	1,849	2,034	2,696	2,166	2,731	2,809
TP: True Positive	1,007	657	160	397	57	48
TN: True Negative	1,847	1,850	2,668	2,032	2,548	2,809
FP: False Positive	1	166	1	294	69	0
FN: False Negative	2	184	28	134	183	0
TPR: (Sensitivity, hit rate, recall)	0·998017839	0·781212842	0·85106383	0·747645951	0·2375	1
TNR=SPC: (Specificity)	0·999458874	0·91765873	0·999625328	0·873602752	0·973633932	1
PPV: Pos Pred Value (Precision)	0·999007937	0·798298906	0·99378882	0·574529667	0·452380952	1
NPV: Neg Pred Value	0·998918334	0·909537856	0·989614243	0·938134811	0·932991578	1
FPR: False-out	0·000541126	0·08234127	0·000374672	0·126397248	0·026366068	0
FDR: False Discovery Rate	0·000992063	0·201701094	0·00621118	0·425470333	0·547619048	0
FNR: Miss Rate	0·001982161	0·218787158	0·14893617	0·252354049	0·7625	0
ACC: Accuracy	0·998949947	0·877493875	0·989849492	0·85019251	0·91179559	1
F1 score	0·998512643	0·789663462	0·916905444	0·649754501	0·31147541	1
MCC: Matthew's correlation coefficient	0·997701467	0·703339883	0·914642183	0·56435108	0·28524589	1
Informedness	0·997476714	0·698871572	0·850689158	0·621248703	0·211133932	1
Markedness	0·997926271	0·707836763	0·983403063	0·512664478	0·385372531	1
Prevalence	0·353167658	0·294364718	0·06580329	0·185859293	0·0840042	0·01680084
LR+: Positive likelihood ratio	1844·336967	9·487500537	2271·489362	5·915049259	9·007789855	inf
LR-: Negative likelihood ratio	0·001983234	0·238418871	0·148991993	0·288865904	0·783148548	0
DOR: Diagnostic odds ratio	929964·5	39·79341278	15245·71429	20·47679968	11·50201948	inf
FOR: False omission rate	0·001081666	0·090462144	0·010385757	0·061865189	0·067008422	0

Supplementary Table 7. Detailed confusion matrix statistics on the test data of the models used to predict ALS clinical subtypes. The tables show the detailed confusion matrix statistics of (A) the model based on all of the available clinical features and (B) the model based on only the most important clinical features.

(A) Predictive model based on all available clinical features

	Bulbar	Classical	Flail arm	Flail leg	Pyramidal	Respiratory
Threshold (max F1 score)	argmax	argmax	argmax	argmax	argmax	argmax
Population	1,073	1,073	1,073	1,073	1,073	1,073
P: Condition positive	376	462	52	109	50	24
N: Condition negative	697	611	1021	964	1,023	1,049
Test outcome positive	376	227	185	249	13	23
Test outcome negative	697	846	888	824	1,060	1,050
TP: True Positive	376	191	23	75	3	23
TN: True Negative	697	575	859	790	1,013	1,049
FP: False Positive	0	36	162	174	10	0
FN: False Negative	0	271	29	34	47	1
TPR: (Sensitivity, hit rate, recall)	1	0·413419913	0·442307692	0·688073394	0·06	0·958333333
TNR=SPC: (Specificity)	1	0·941080196	0·841332027	0·819502075	0·990224829	1
PPV: Pos Pred Value (Precision)	1	0·841409692	0·124324324	0·301204819	0·230769231	1
NPV: Neg Pred Value	1	0·679669031	0·967342342	0·958737864	0·955660377	0·999047619
FPR: False-out	0	0·058919804	0·158667973	0·180497925	0·009775171	0
FDR: False Discovery Rate	0	0·158590308	0·875675676	0·698795181	0·769230769	0
FNR: Miss Rate	0	0·586580087	0·557692308	0·311926606	0·94	0·041666667
ACC: Accuracy	1	0·7138863	0·821994408	0·806150979	0·946877912	0·999068034
F1 score	1	0·554426705	0·194092827	0·418994413	0·095238095	0·978723404
MCC: Matthews correlation coefficient	1	0·429793514	0·161246109	0·36323619	0·096764638	0·978478735
Informedness	1	0·35450011	0·28363972	0·507575469	0·050224829	0·958333333
Markedness	1	0·521078722	0·091666667	0·259942683	0·186429608	0·999047619
Prevalence	0·350419385	0·4305685	0·048462255	0·101584343	0·046598322	0·022367195
LR+: Positive likelihood ratio	inf	7·016654642	2·787630579	3·812084783	6·138	inf
LR-: Negative likelihood ratio	0	0·623305101	0·662868273	0·380629427	0·949279368	0·041666667
DOR: Diagnostic odds ratio	inf	11·25717507	4·205406556	10·01521298	6·465957447	inf
FOR: False omission rate	0	0·320330969	0·032657658	0·041262136	0·044339623	0·000952381

(B) Predictive model based on most important clinical features

	Bulbar	Classical	Flail arm	Flail leg	Pyramidal	Respiratory
Threshold (max F1 score)	argmax	argmax	argmax	argmax	argmax	argmax
Population	1,073	1,073	1,073	1,073	1,073	1,073
P: Condition positive	376	462	52	109	50	24
N: Condition negative	697	611	1,021	964	1,023	1,049
Test outcome positive	376	375	65	218	16	23
Test outcome negative	697	698	1,008	855	1,057	1,050
TP: True Positive	376	284	13	69	2	23
TN: True Negative	697	520	969	815	1,009	1,049
FP: False Positive	0	91	52	149	14	0
FN: False Negative	0	178	39	40	48	1
TPR: (Sensitivity, hit rate, recall)	1	0·614718615	0·25	0·633027523	0·04	0·958333333
TNR=SPC: (Specificity)	1	0·85106383	0·94906954	0·845435685	0·986314761	1
PPV: Pos Pred Value (Precision)	1	0·757333333	0·2	0·316513761	0·125	1
NPV: Neg Pred Value	1	0·744985673	0·961309524	0·953216374	0·954588458	0·999047619
FPR: False-out	0	0·14893617	0·05093046	0·154564315	0·013685239	0
FDR: False Discovery Rate	0	0·242666667	0·8	0·683486239	0·875	0
FNR: Miss Rate	0	0·385281385	0·75	0·366972477	0·96	0·041666667
ACC: Accuracy	1	0·749301025	0·915191053	0·823858341	0·94221808	0·999068034
F1 score	1	0·678614098	0·222222222	0·422018349	0·060606061	0·978723404
MCC: Matthew's correlation coefficient	1	0·483705876	0·179197692	0·359243575	0·045764082	0·978478735
Informedness	1	0·465782445	0·19906954	0·478463208	0·026314761	0·958333333
Markedness	1	0·502319007	0·161309524	0·269730136	0·079588458	0·999047619
Prevalence	0·350419385	0·4305685	0·048462255	0·101584343	0·046598322	0·022367195
LR+: Positive likelihood ratio	inf	4·127396413	4·908653846	4·095560618	2·922857143	inf
LR-: Negative likelihood ratio	0	0·452705628	0·790247678	0·434063151	0·973320119	0·041666667
DOR: Diagnostic odds ratio	inf	9·11717496	6·211538462	9·435402685	3·00297619	inf
FOR: False omission rate	0	0·255014327	0·038690476	0·046783626	0·045411542	0·000952381

CONSORTIA

The members of the PARALS Consortium are Adriano Chiò[1,2], Andrea Calvo[1,2], Cristina Moglia[1,2], Antonio Canosa[1,2], Umberto Manera[1], Rosario Vasta[1], Francesca Palumbo[1], Alessandro Bombaci[1], Maurizio Grassano[1], Maura Brunetti[1], Federico Casale[1], Giuseppe Fuda[1], Paolina Salamone[1], Barbara Iazzolino[1], Laura Peotta[1], Paolo Cugnasco[1], Giovanni De Marco[3], Maria Claudia Torrieri[1], Salvatore Gallone[3], Marco Barberis[4], Luca Sbaiz[4], Salvatore Gentile[5], Alessandro Mauro[1,6], Letizia Mazzini[7,8], Fabiola De Marchi[7,8], Lucia Corrado[9,8], Sandra D'Alfonso[9,8], Antonio Bertolotto[10], Daniele Imperiale[11], Marco De Mattei[12], Salvatore Amarù[13], Cristoforo Comi[14,15], Carmelo Labate[16], Fabio Poglio[16], Luigi Ruiz[17], Lucia Testa[18], Eugenia Rota[19], Paolo Ghiglione[20], Nicola Launaro[21], and Alessia Di Sapio[22]

1. "Rita Levi Montalcini" Department of Neuroscience, ALS Centre, University of Torino, Turin, Italy
2. SC Neurologia 1, Azienda Ospedaliero Universitaria Città della Salute e della Scienza, Torino, Italy
3. Neurologia 1, Azienda Ospedaliero Universitaria Città della Salute e della Scienza, Torino, Italy
4. Department of Medical Genetics, Azienda Ospedaliero Universitaria Città della Salute e della Scienza, Torino, Italy
5. Neurologia 3, Azienda Ospedaliero Universitaria Città della Salute e della Scienza di Torino, Torino, Italy
6. Istituto Auxologico Italiano, IRCCS, Piancavallo, Italy
7. Department of Neurology, 'Amedeo Avogadro' University of Piemonte Orientale, Novara, Italy
8. Azienda Ospedaliero Universitaria 'Maggiore della Carità', Novara, Italy
9. Department of Haed Sciences, 'Amedeo Avogadro' University of Piemonte Orientale, Novara, Italy
10. Department of Neurology and Multiple Sclerosis Center, Azienda Ospedaliero Universitaria San Luigi, Orbassano, Italy
11. Department of Neurology, Ospedale Maria Vittoria, ASL Città di Torino, Torino, Italy
12. Department of Neurology, Ospedale 'Santa Croce' di Moncalieri, ASL Torino 5, Moncaliari, Italy
13. Department of Neurology, Presidio Ospedaliero di Rivoli, ASL Torino3, Rivoli, Italy
14. Department of Neurology, Ospedale 'Sant' Andrea' di Vercelli, ASL Vercelli, Vercelli, Italy
15. Department of Clinical and Experimental Medicine, 'Amedeo Avogadro' University of Piemonte Orientale, Novara, Italy
16. Department of Neurology, Ospedale Civile 'Edoardo Agnelli' di Pinerolo, ALS Torino 2, Pinerolo, Italy
17. Department of Neurology, Azienda Ospedaliera 'Santi Antonio e Biagio' di Alesssandria, Alessandria, Italy
18. Department of Neurology, Ospedale 'Santo Spirito' di Casale Monferrato, ASL Alessandria, Casale Monferrato, Italy
19. Department of Neurology, Ospedale 'San Giacomo' di Novi Ligure, ASL Alesssandria, Novi Ligure, Italy
20. Department of Neurology, Ospedale 'Maggiore Santissima Annuziata' di Savigliano, ASL Cuneo 1, Savigliano, Italy
21. Department of Anestesiology, Ospedale 'Maggiore Santissima Annuziata' di Savigliano, ASL Cuneo 1, Savigliano, Italy
22. Department of Neurology, Ospedale 'Regina Montis Regalis' di Mondovì, ASL Cuneo 1, Italy

The members of the ERRALS Consortium are: Jessica Mandrioli[1,2], Nicola Fini[1], Ilaria Martinelli[1,2], Elisabetta Zucchi[1,2], Giulia Gianferrari[2], Cecilia Simonini[1], Stefano Meletti[1-2], Rocco Liguori[3], Veria Vacchiano[3], Fabrizio Salvi[4], Ilaria Bartolomei[4], Roberto Michelucci[4], Pietro Cortelli[3,5], Rita Rinaldi[5], Anna Maria Borghi[6], Andrea Zini[6], Elisabetta Sette[7], Valeria Tugnoli[7], Maura Pugliatti[8], Elena Canali[9], Luca Codeluppi[9], Franco Valzania[9], Lucia Zinno[10], Giovanni Pavesi[10], Doriana Medici[11], Giovanna Pilurzi[11], Emilio Terlizzi[12], Donata Guidetti[12], Silvia De Pasqua[13], Mario Santangelo[13], Patrizia De Massis[14], Martina Bracaglia[14], Mario Casmiro[15], Pietro Querzani[15], Simonetta Morresi[16], Marco Longoni[16-17], Alberto Patuelli[17], Susanna Malagù[17], Marco Currò Dossi[18], Simone Vidale[18], and Salvatore Ferro[19]

1. Department of Neurosciences, Azienda Ospedaliero Universitaria di Modena, Modena, Italy
2. Department of Biomedical, Metabolic and Neurosciences, University of Modena and Reggio Emilia, Modena, Italy
3. Dipartimento di Scienze Biomediche e Neuromotorie, University of Bologna, and IRCCS Istituto delle Scienze Neurologiche di Bologna, Bellaria Hospital, Bologna, Italy

4. IRCCS Istituto delle Scienze Neurologiche di Bologna, Bellaria Hospital, Bologna, Italy
5. IRCCS Istituto delle Scienze Neurologiche di Bologna, UOC Interaziendale Clinica Neurologica Metropolitana (NeuroMet), Bologna, Italy
6. IRCCS Istituto delle Scienze Neurologiche di Bologna, Department of Neurology and Stroke Center, Maggiore Hospital, Bologna, Italy
7. Department of Neurosciences and Rehabilitation, St Anna Hospital, Ferrara, Italy
8. Department of Neuroscience, University of Ferrara, Ferrara, Italy
9. Department of Neurology, IRCCS Arcispedale Santa Maria Nuova, Reggio Emilia
10. Department of Neuroscience, University of Parma, Parma, Italy
11. Department of Neurology, Fidenza Hospital, Parma, Italy
12. Department of Neurology, G. Da Saliceto Hospital, Piacenza, Italy
13. Department of Neurology, Carpi Hospital, Modena, Italy
14. Department of Neurology, Imola Hospital, Bologna, Italy
15. Department of Neurology, Faenza and Ravenna Hospital, Ravenna, Italy
16. Department of Neurology, Bufalini Hospital, Cesena, Italy
17. Department of Neurology, Forlì Hospital, Forlì, Italy
18. Department of Neurology, Infermi Hospital, Rimini, Italy
19. Department of Hospital Services, Emilia Romagna Regional Health Authority, Bologna, Italy

Table 1 The MI-CLAIM checklist

From: Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist

Fagri et al, tldigitalhealth-D-21-00297

Before paper submission			
Study design (Part 1)	**Completed: page number**		**Notes if not completed**
The clinical problem in which the model will be employed is clearly detailed in the paper.	☒	8-9	
The research question is clearly stated.	☒	8-9	
The characteristics of the cohorts (training and test sets) are detailed in the text.	☒	9-10	
The cohorts (training and test sets) are shown to be representative of real-world clinical settings.	☒	9-10	
The state-of-the-art solution used as a baseline for comparison has been identified and detailed.	☒	10	
Data and optimization (Parts 2, 3)	**Completed: page number**		**Notes if not completed**
The origin of the data is described and the original format is detailed in the paper.	☒	9-10	
Transformations of the data before it is applied to the proposed model are described.	☒	10-11	
The independence between training and test sets has been proven in the paper.	☒	10	
Details on the models that were evaluated and the code developed to select the best model are provided.	☒	11-14	
Is the input data type structured or unstructured?	☒ Structured ☐ Unstructured		
Model performance (Part 4)	**Completed: page number**		**Notes if not completed**
The primary metric selected to evaluate algorithm performance			

The primary metric selected to evaluate algorithm performance (e.g., AUC, F-score, etc.), including the justification for selection, has been clearly stated.	☒	11-13	
The primary metric selected to evaluate the clinical utility of the model (e.g., PPV, NNT, etc.), including the justification for selection, has been clearly stated.	☒	13	
The performance comparison between baseline and proposed model is presented with the appropriate statistical significance.	☒	12	
Model examination (Part 5)	**Completed:**	**page number**	**Notes if not completed**
Examination technique 1[a]	☒	15	
Examination technique 2[a]	☒	16	
A discussion of the relevance of the examination results with respect to model/algorithm performance is presented.	☒	17-21	
A discussion of the feasibility and significance of model interpretability at the case level if examination methods are uninterpretable is presented.	☒	17-21	
A discussion of the reliability and robustness of the model as the underlying data distribution shifts is included.	☒	17-21	
Reproducibility (Part 6): choose appropriate tier of transparency			**Notes**
Tier 1: complete sharing of the code	☒		
Tier 2: allow a third party to evaluate the code for accuracy/fairness; share the results of this evaluation	☐		
Tier 3: release of a virtual machine (binary) for running the code on new data without sharing its details	☐		
Tier 4: no sharing	☐		

1. PPV, positive predictive value; NNT, numbers needed to treat.
2. [a]Common examination approaches based on study type: for studies involving exclusively structured data, coefficients and sensitivity analysis are often appropriate; for studies involving unstructured data in the domains of image analysis or natural language processing, saliency maps (or equivalents) and sensitivity analyses are often appropriate.

Back to article page ›

Nature Medicine ISSN 1546-170X (online)

Faghri et al, tldigitalhealth-D-21-00297

STROBE (Strengthening The Reporting of OBservational Studies in Epidemiology) Checklist

A checklist of items that should be included in reports of observational studies. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Section and Item	Item No.	Recommendation	Reported on Page No.
Title and Abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/Rationale	2	Explain the scientific background and rationale for the investigation being reported	8-9
Objectives	3	State specific objectives, including any prespecified hypotheses	9,11
Methods		**Text**	
Study Design	4	Present key elements of study design early in the paper	9-13
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	9-10
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	9-10
		Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	NA
		Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants	NA
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed	NA
		Case-control study—For matched studies, give matching criteria and the number of controls per case	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	11

Section and Item	Item No.	Recommendation	Reported on Page No.
Data Sources/ Measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	9-10
Bias	9	Describe any efforts to address potential sources of bias	10-11
Study Size	10	Explain how the study size was arrived at	9-10
Quantitative Variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	10-11
Statistical Methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	10-13
		(*b*) Describe any methods used to examine subgroups and interactions	10-13
		(*c*) Explain how missing data were addressed	10-11
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed	10-11
		Case-control study—If applicable, explain how matching of cases and controls was addressed	NA
		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	13
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9,14,15
		(b) Give reasons for non-participation at each stage	9-11
		(c) Consider use of a flow diagram	Figure 1
Descriptive Data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	9-11, Table S2
		(b) Indicate number of participants with missing data for each variable of interest	11, Table S2
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	9-10
Outcome Data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	9,10,14,15
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	NA
		Cross-sectional study—Report numbers of outcome events or summary measures	NA

Section and Item	Item No.	Recommendation	Reported on Page No.
Main Results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	16
		(*b*) Report category boundaries when continuous variables were categorized	NA
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other Analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	16, Tables S3 to S5
Discussion			
Key Results	18	Summarise key results with reference to study objectives	21
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	20
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	21
Generalisability	21	Discuss the generalisability (external validity) of the study results	20
Other Information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	14

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.

TRIPOD Checklist: Prediction Model Development

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	3
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	8-9
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	9
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	9-10
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	9-10
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	9-10
	5b	Describe eligibility criteria for participants.	9-10
	5c	Give details of treatments received, if relevant.	NA
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	11-12
	6b	Report any actions to blind assessment of the outcome to be predicted.	11-12
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	11–12
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	11-12
Sample size	8	Explain how the study size was arrived at.	9-10
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	10-11
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	10-13
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	11-13
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	11,13
Risk groups	11	Provide details on how risk groups were created, if done.	NA
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	9,10,14, 15,fig 1
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	9,10,14,15, table S2
Model development	14a	Specify the number of participants and outcome events in each analysis.	9,10
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	NA
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	17
	15b	Explain how to the use the prediction model.	17
Model performance	16	Report performance measures (with CIs) for the prediction model.	17
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	20-21
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	19-21
Implications	20	Discuss the potential clinical use of the model and implications for future research.	17-19, 21
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	16, 21
Funding	22	Give the source of funding and the role of the funders for the present study.	14,21

We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

Exploratory Data Analysis Report on the PARALS Registry

Report by dlookr package

2021-07-23

1 Introduction

The EDA Report provides exploratory data analysis information on objects that inherit data.frame and data.frame.

1.1 Information of Dataset

The dataset that generated the EDA Report is an **'data.frame'** object. It consists of **2,858 observations** and **66 variables**.

1.2 Information of Variables

The variable information of the data set that generated the EDA Report is shown in the following table.:

Information of Variables

variables	types	missing_count	missing_percent	unique_count	unique_rate
Age at diagnosis	numeric	0	0.0000000	865	0.3026592

Age at symptom onset	numeric	0	0.0000000	884	0.3093072
ALSFRS1	integer	486	17.0048985	6	0.0020994
ALSFRS2	integer	486	17.0048985	6	0.0020994
ALSFRS3	integer	486	17.0048985	6	0.0020994
ALSFRS4	integer	486	17.0048985	6	0.0020994
ALSFRS5	integer	486	17.0048985	6	0.0020994
ALSFRS6	integer	486	17.0048985	6	0.0020994
ALSFRS7	integer	486	17.0048985	6	0.0020994
ALSFRS8	integer	486	17.0048985	6	0.0020994
ALSFRS9	integer	486	17.0048985	6	0.0020994
ALSFRS10	integer	486	17.0048985	6	0.0020994
ALSFRS11	integer	486	17.0048985	6	0.0020994
ALSFRS12	integer	486	17.0048985	6	0.0020994
First ALSFRS-R total	integer	486	17.0048985	42	0.0146956
Anatomical level at onset	character	0	0.0000000	6	0.0020994
Ataxia	character	0	0.0000000	2	0.0006998
BMI 2 years prior to illness	numeric	900	31.4905528	1003	0.3509447
BMI at diagnosis	numeric	853	29.8460462	1087	0.3803359
C9orf72 status	character	1437	50.2799160	3	0.0010497
Cancer	character	0	0.0000000	2	0.0006998
Cancer type	character	2724	95.3114066	32	0.0111966
Chorea	character	0	0.0000000	2	0.0006998
Clinical type at one year	character	1	0.0349895	7	0.0024493
Clinical type at onset	character	0	0.0000000	4	0.0013996
Cognitive impairment present	character	1884	65.9202239	3	0.0010497
Cognitive status 1	character	1884	65.9202239	6	0.0020994
Cognitive status 2	character	2063	72.1833450	7	0.0024493
COPD	character	0	0.0000000	2	0.0006998
Delay in diagnosis (days)	integer	0	0.0000000	191	0.0668300
Diabetes	character	42	1.4695591	3	0.0010497
Education	character	145	5.0734780	7	0.0024493
El Escorial category at diagnosis	character	0	0.0000000	5	0.0017495
El Escorial category at visit 2	character	2116	74.0377887	5	0.0017495

El Escorial category at visit 3	character	2676	93.6319104	5	0.0017495
Family history of ALS	character	0	0.0000000	2	0.0006998
FVC percent at diagnosis	numeric	919	32.1553534	138	0.0482855
Height	integer	842	29.4611617	58	0.0202939
Hypertension	character	42	1.4695591	3	0.0010497
Hyperthyroid	character	0	0.0000000	2	0.0006998
Hypothyroid	character	0	0.0000000	2	0.0006998
Initial diagnosis was PLS	character	0	0.0000000	2	0.0006998
Marital status	character	97	3.3939818	6	0.0020994
Mutated gene	character	2689	94.0867740	8	0.0027992
Mutation amino acid change	character	2788	97.5507348	37	0.0129461
Mutation present	character	0	0.0000000	2	0.0006998
NIPPV	character	279	9.7620714	3	0.0010497
Onset side	character	1082	37.8586424	4	0.0013996
Parkinsonism	character	0	0.0000000	2	0.0006998
PEG inserted	character	298	10.4268719	3	0.0010497
Place of birth	logical	2858	100.0000000	1	0.0003499
Place of residence	logical	2858	100.0000000	1	0.0003499
Rate of decline ALSFRS-R (per month)	numeric	491	17.1798460	1632	0.5710287
Rate of decline BMI (per month)	numeric	900	31.4905528	1380	0.4828551
Sex	character	1	0.0349895	3	0.0010497
Site of symptom onset	character	0	0.0000000	3	0.0010497
Smoker	character	855	29.9160252	4	0.0013996
Survival (days)	integer	5	0.1749475	1598	0.5591323
Time of first ALSFRS-R (days into illness)	integer	483	16.8999300	812	0.2841148
Time of NIPPV (days into illness)	integer	1998	69.9090273	257	0.0899230
Time of PEG (days into illness)	integer	1971	68.9643107	665	0.2326802
Time of tracheostomy (days into illness)	integer	2495	87.2988104	330	0.1154654
Tracheostomy	character	83	2.9041288	3	0.0010497
Vital status	character	4	0.1399580	3	0.0010497

Weight 2 years prior to illness	numeric	868	30.3708887	108	0.0377887
Weight at diagnosis	numeric	810	28.3414976	126	0.0440868

The target variable of the data is **'Clinical type at one year'**, and the data type of the variable is **character**.

1.3 About EDA Report

EDA reports provide information and visualization results that support the EDA process. In particular, it provides a variety of information to understand the relationship between the target variable and the rest of the variables of interest.

2 Univariate Analysis

2.1 Descriptive Statistics

edaData

66 Variables 2858 Observations

Age at diagnosis

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
2858	0	865	1	66.82	12.23	45.99	52.14	60.51	68.18	74.49	79.58	82.67

lowest : 16.91781 20.49315 24.16438 24.75342 25.91781 , highest: 89.58904 89.67123 90.58904 91.00000 91.91781

Age at symptom onset

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
2858	0	884	1	65.86	12.25	45.07	51.23	59.58	67.25	73.51	78.59	81.75

lowest : 15.91781 20.32877 21.34247 21.67123 24.16438 , highest: 89.00000 89.24658 89.75342 90.00000 91.42466

ALSFRS1

n	missing	distinct	Info	Mean	Gmd
2372	486	5	0.822	3.309	0.855

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     52    53   205   863  1199
Proportion 0.022 0.022 0.086 0.364 0.505
```

ALSFRS2

n	missing	distinct	Info	Mean	Gmd
2372	486	5	0.567	3.68	0.51

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency      5    26   104   452  1785
Proportion 0.002 0.011 0.044 0.191 0.753
```

ALSFRS3

n	missing	distinct	Info	Mean	Gmd
2372	486	5	0.797	3.382	0.8017

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     43    29   214   779  1307
Proportion 0.018 0.012 0.090 0.328 0.551
```

ALSFRS4

n	missing	distinct	Info	Mean	Gmd
2372	486	5	0.83	3.297	0.8436

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     33    67   223   889  1160
Proportion 0.014 0.028 0.094 0.375 0.489
```

ALSFRS5

n	missing	distinct	Info	Mean	Gmd
2372	486	5	0.849	3.224	0.9397

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     39   100   292   801  1140
Proportion 0.016 0.042 0.123 0.338 0.481
```

ALSFRS6

n	missing	distinct	Info	Mean	Gmd
2372	486	5	0.846	3.209	0.9902

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     53   114   298   727  1180
Proportion 0.022 0.048 0.126 0.306 0.497
```

ALSFRS7

n	missing	distinct	Info	Mean	Gmd
2372	486	5	0.77	3.382	0.8645

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     49    80   200   631  1412
Proportion 0.021 0.034 0.084 0.266 0.595
```

ALSFRS8

n	missing	distinct	Info	Mean	Gmd
2372	486	5	0.878	3.121	0.9685

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     34    91   419   837   991
Proportion 0.014 0.038 0.177 0.353 0.418
```

ALSFRS9

n	missing	distinct	Info	Mean	Gmd
2372	486	5	0.916	2.691	1.492

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency    231   323   364   484   970
Proportion 0.097 0.136 0.153 0.204 0.409
```

ALSFRS10

n	missing	distinct	Info	Mean	Gmd
2372	486	5	0.405	3.735	0.4703

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     48    11    85   234  1994
Proportion 0.020 0.005 0.036 0.099 0.841
```

ALSFRS11

n	missing	distinct	Info	Mean	Gmd
2372	486	5	0.347	3.812	0.3376

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     14    10    69   222  2057
Proportion 0.006 0.004 0.029 0.094 0.867
```

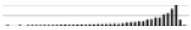
ALSFRS12

n	missing	distinct	Info	Mean	Gmd
2372	486	5	0.107	3.916	0.1635

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     24     8    23    33  2284
Proportion 0.010 0.003 0.010 0.014 0.963
```

First ALSFRS-R total



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
2372	486	41	0.99	40.76	6.333	27	32	39	43	45	46	46

lowest : 5 8 10 11 12 , highest: 44 45 46 47 48

Anatomical level at onset

n	missing	distinct
2858	0	6

lowest :	bulbar	lower_limbs_distal	lower_limbs_proximal	respiratory	upper_limbs_distal
highest:	lower_limbs_distal	lower_limbs_proximal	respiratory	upper_limbs_distal	upper_limbs_proximal

```
Value                   bulbar   lower_limbs_distal lower_limbs_proximal
Frequency                 1009                  908                  115
Proportion               0.353                0.318                0.040

Value               respiratory   upper_limbs_distal upper_limbs_proximal
Frequency                    48                  617                  161
Proportion                0.017                0.216                0.056
```

Ataxia

n	missing	distinct
2858	0	2

```
Value        no  yes
Frequency  2857    1
Proportion    1    0
```

BMI 2 years prior to illness

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
1958	900	1002	1	25.96	4.477	20.08	21.26	23.26	25.39	28.23	31.22	33.33

lowest : 14.53287 15.88697 16.00000 16.02307 16.67585 , highest: 42.43663 42.46114 43.28824 43.42857 43.50039

BMI at diagnosis

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
2005	853	1086	1	24.45	4.611	18.27	19.37	21.77	24.09	26.83	29.78	31.63

lowest : 13.23605 13.38797 13.84083 14.17769 14.19686 , highest: 41.01562 41.08839 41.40163 41.64931 43.42857

C9orf72 status

n	missing	distinct
1421	1437	2

```
Value       EXP   WT
Frequency   100 1321
Proportion 0.07 0.93
```

Cancer

n	missing	distinct
2858	0	2

```
Value         no   yes
Frequency   2722   136
Proportion 0.952 0.048
```

Cancer type

n	missing	distinct
134	2724	31

lowest :	bile_duct	bladder	breast	chronic_lymphoid_leukemia	chronic_myeloid_leukemia
highest:	stomach	thymoma	thyroid	uterus	vulva

Chorea

n	missing	distinct
2858	0	2

Value	no	yes
Frequency	2857	1
Proportion	1	0

Clinical type at one year

n	missing	distinct
2857	1	6

lowest :	bulbar	classical	flailArm	flailLeg	pyramidal
highest:	classical	flailArm	flailLeg	pyramidal	respiratory

Value	bulbar	classical	flailArm	flailLeg	pyramidal	respiratory
Frequency	1009	841	188	531	240	48
Proportion	0.353	0.294	0.066	0.186	0.084	0.017

Clinical type at onset

n	missing	distinct
2858	0	4

Value	bulbar	classical	LMNpredominant	UMNpredominant
Frequency	893	1030	501	434
Proportion	0.312	0.360	0.175	0.152

Cognitive impairment present

n	missing	distinct
974	1884	2

Value	no	yes
Frequency	417	557
Proportion	0.428	0.572

Cognitive status 1

n	missing	distinct
974	1884	5

lowest :	ALSbi	ALSci	FTD	non-executive_impairment	normal
highest:	ALSbi	ALSci	FTD	non-executive_impairment	normal

Value	ALSbi	ALSci
Frequency	33	176

```
Proportion                    0.034                    0.181

Value                           FTD non-executive_impairment
Frequency                       287                       22
Proportion                    0.295                    0.023

Value                      normal
Frequency                     456
Proportion                  0.468
```

Cognitive status 2

n	missing	distinct
795	2063	6

lowest :	ALSbi	ALScbi	ALSci	FTD	non-executive_impairment
highest:	ALScbi	ALSci	FTD	non-executive_impairment	normal

```
Value                        ALSbi                    ALScbi
Frequency                       61                        38
Proportion                   0.077                     0.048

Value                        ALSci                       FTD
Frequency                      129                       171
Proportion                   0.162                     0.215

Value      non-executive_impairment                    normal
Frequency                       16                       380
Proportion                   0.020                     0.478
```

COPD

n	missing	distinct
2858	0	2

```
Value         no   yes
Frequency   2656   202
Proportion 0.929 0.071
```

Delay in diagnosis (days)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
2858	0	191	0.999	347.8	283	61	92	153	273	396	730	914

lowest : 28 29 30 31 59 , highest: 2191 2374 2585 3226 3652

Diabetes

n	missing	distinct

2816	42	2

```
Value         no   yes
Frequency   2514   302
Proportion 0.893 0.107
```

Education

n	missing	distinct
2713	145	6

lowest :	11years	13years	18years	5years	8years
highest:	13years	18years	5years	8years	lessthan5years

```
Value            11years 13years 18years  5years
Frequency            184     359     132    1240
Proportion         0.068   0.132   0.049   0.457

Value             8years lessthan5years
Frequency            705            93
Proportion         0.260         0.034
```

El Escorial category at diagnosis

n	missing	distinct
2858	0	5

lowest :	definite	possible	probable	probable_labSupported	suspected
highest:	definite	possible	probable	probable_labSupported	suspected

```
Value                definite possible probable
Frequency                1335      359      880
Proportion              0.467    0.126    0.308

Value      probable_labSupported suspected
Frequency                    169       115
Proportion                 0.059     0.040
```

El Escorial category at visit 2

n	missing	distinct
742	2116	4

```
Value                definite possible probable
Frequency                 372       71      260
Proportion              0.501    0.096    0.350
```

Value	probable_labSupported
Frequency	39
Proportion	0.053

El Escorial category at visit 3

n	missing	distinct
182	2676	4

Value	definite	possible	probable
Frequency	146	8	24
Proportion	0.802	0.044	0.132

Value	probable_labSupported
Frequency	4
Proportion	0.022

Family history of ALS

n	missing	distinct
2858	0	2

Value	no	yes
Frequency	2671	187
Proportion	0.935	0.065

FVC percent at diagnosis

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
1939	919	137	1	84.61	28.95	37	48	68	88	102	114	125

lowest : 10 17 20 21 22 , highest: 153 154 157 159 160

Height

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
2016	842	57	0.998	164.5	10.48	150	152	158	165	171	176	180

lowest : 136 137 138 139 140 , highest: 188 189 190 196 198

Hypertension

n	missing	distinct
2816	42	2

Value	no	yes
Frequency	1609	1207
Proportion	0.571	0.429

Hyperthyroid

n	missing	distinct
2858	0	2

```
Value         no   yes
Frequency   2814    44
Proportion 0.985 0.015
```

Hypothyroid

n	missing	distinct
2858	0	2

```
Value         no   yes
Frequency   2654   204
Proportion 0.929 0.071
```

Initial diagnosis was PLS

n	missing	distinct
2858	0	2

```
Value         no   yes
Frequency   2832    26
Proportion 0.991 0.009
```

Marital status

n	missing	distinct
2761	97	5

lowest :	divorced	married	neverMarried	separated	widowed
highest:	divorced	married	neverMarried	separated	widowed

```
Value          divorced      married neverMarried    separated      widowed
Frequency            64         2034          245           52          366
Proportion        0.023        0.737        0.089        0.019        0.133
```

Mutated gene

n	missing	distinct

169	2689	7

lowest :	C9ORF72	C9ORF72+TARDBP	FUS	OPTN	SOD1
highest:	FUS	OPTN	SOD1	TARDBP	TUBA4A

Value	C9ORF72	C9ORF72+TARDBP	FUS	OPTN
Frequency	99	1	11	1
Proportion	0.586	0.006	0.065	0.006

Value	SOD1	TARDBP	TUBA4A
Frequency	34	21	2
Proportion	0.201	0.124	0.012

Mutation amino acid change

n	missing	distinct
70	2788	36

lowest : A382T A4V D109Y D438N D90A_het , highest: R521G S135N S393L T349S V47F

Mutation present

n	missing	distinct
2858	0	2

Value	no	yes
Frequency	2689	169
Proportion	0.941	0.059

NIPPV

n	missing	distinct
2579	279	2

Value	no	yes
Frequency	1717	862
Proportion	0.666	0.334

Onset side

n	missing	distinct
1776	1082	3

Value	both	left	right
Frequency	549	438	789
Proportion	0.309	0.247	0.444

Parkinsonism

n	missing	distinct
2858	0	2

```
Value         no   yes
Frequency   2806    52
Proportion 0.982 0.018
```

PEG inserted

n	missing	distinct
2560	298	2

```
Value         no   yes
Frequency   1670   890
Proportion 0.652 0.348
```

Rate of decline ALSFRS-R (per month)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
2367	491	1631	1	0.8659	0.8923	0.1066	0.1587	0.2748	0.5389	1.0332	1.8249	2.5048

lowest :	0.000000000	0.008283542	0.009431045	0.027752502	0.030439122
highest:	11.090909091	13.401515152	14.233333333	14.724137931	30.500000000

Rate of decline BMI (per month)

```
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25
    1958      900     1379    0.978  0.04479  0.06829 -0.01681  0.00000  0.00000
     .50      .75      .90      .95
 0.02511  0.06849  0.13791  0.18421
```

lowest :	-0.4680072	-0.1600308	-0.1465404	-0.1253496	-0.1227047
highest:	0.3707365	0.3716188	0.3719334	0.4214638	0.4431558

Sex

n	missing	distinct
2857	1	2

```
Value          F     M
Frequency   1318  1539
Proportion 0.461 0.539
```

Site of symptom onset

n	missing	distinct
2858	0	3

Value	bulbar	respiratory	spinal
Frequency	1009	48	1801
Proportion	0.353	0.017	0.630

Smoker

n	missing	distinct
2003	855	3

Value	current	former	never
Frequency	406	640	957
Proportion	0.203	0.320	0.478

Survival (days)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
2853	5	1597	1	1320	1138	261.0	363.2	581.0	932.0	1552.0	2737.0	3866.0

lowest : 18 33 44 64 80 , highest: 8249 8305 8584 8614 8645

Time of first ALSFRS-R (days into illness)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
2375	483	811	1	402.1	334.3	88.0	119.0	184.0	301.0	463.5	774.6	1077.3

lowest : 0 9 11 15 18 , highest: 3652 3682 3813 3820 3839

Time of NIPPV (days into illness)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
860	1998	256	1	868.9	669.4	183.9	273.0	426.0	699.5	1066.2	1614.0	2100.1

lowest : 0 60 61 62 89 , highest: 4717 5114 5783 6574 7152

Time of PEG (days into illness)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
887	1971	664	1	833.6	574.6	202.0	287.0	474.5	696.0	988.5	1506.0	1870.1

lowest : 44 80 87 94 108 , highest: 3428 3564 4372 4541 6372

Time of tracheostomy (days into illness)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
363	2495	329	1	907.7	644.8	199.0	256.8	459.0	787.0	1225.0	1638.0	2007.7

lowest : 18 33 44 92 106 , highest: 3084 3086 3160 3165 4505

Tracheostomy

n	missing	distinct
2775	83	2

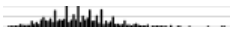
```
Value         no   yes
Frequency   2412   363
Proportion 0.869 0.131
```

Vital status

n	missing	distinct
2854	4	2

```
Value      alive  dead
Frequency    305  2549
Proportion 0.107 0.893
```

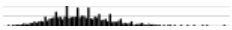
Weight 2 years prior to illness



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
1990	868	107	0.999	70.28	13.88	52	55	61	70	78	85	91

lowest : 40 41 42 43 44 , highest: 114 118 126 132 133

Weight at diagnosis



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
2048	810	125	0.999	66.22	14.37	47	50	57	65	75	83	88

lowest : 33 35 36 37 38 , highest: 116 119 123 125 133

Variables with all observations missing:

`Place of birth`, `Place of residence`

2.2 Normality Test of Numerical Variables

2.2.1 Statistics and Visualization of (Sample) Data

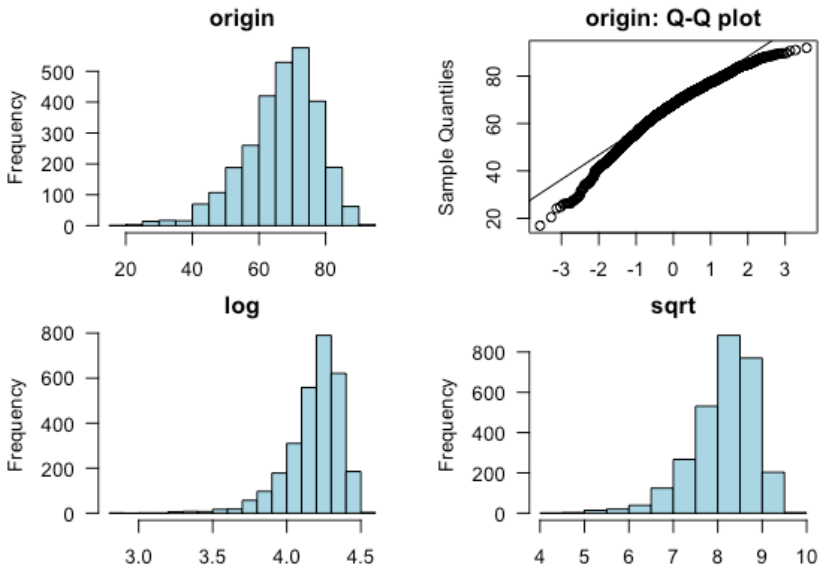
[Age at diagnosis]

normality test : Shapiro-Wilk normality test

```
statistic : 0.9697,  p-value : 4.08035E-24
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.7433808	3.787688
log transformation	-1.5836750	7.471166
sqrt transformation	-1.1069448	5.045099



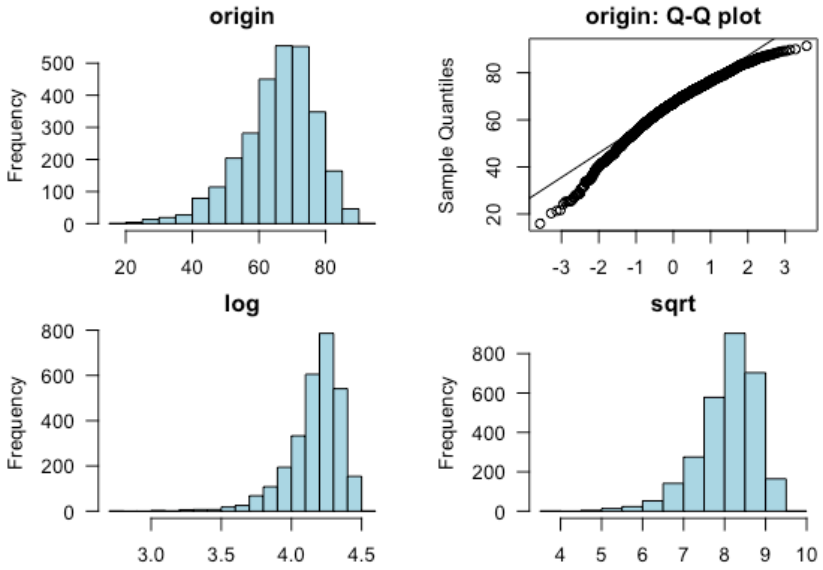
[Age at symptom onset]

normality test : Shapiro-Wilk normality test

```
statistic : 0.97055,  p-value : 8.57396E-24
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.7349793	3.789471
log transformation	-1.6099153	7.707313
sqrt transformation	-1.1101840	5.100604



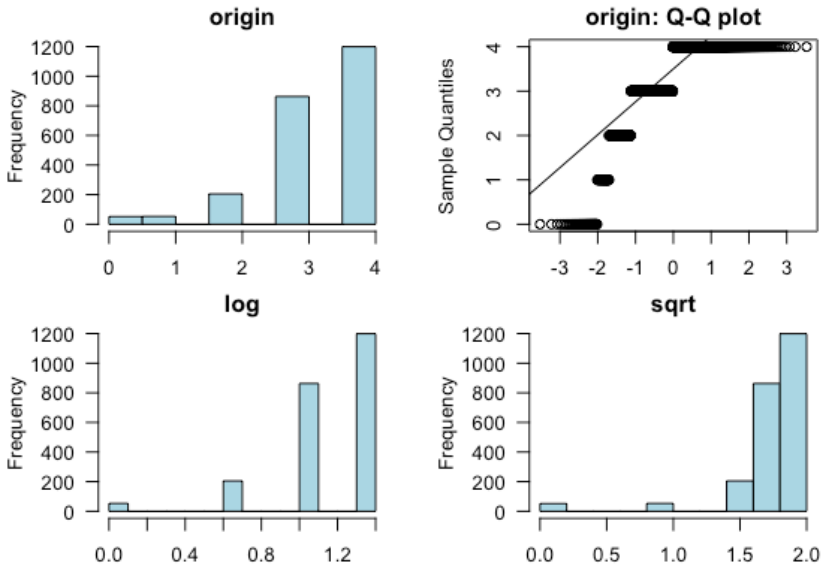
[ALSFRS1]

normality test : Shapiro-Wilk normality test

```
statistic : 0.73165,  p-value : 3.98662E-52
```

skewness and kurtosis

type	skewness	kurtosis
original	-1.595538	5.922016
log transformation	NaN	NaN
sqrt transformation	-3.256496	16.290301



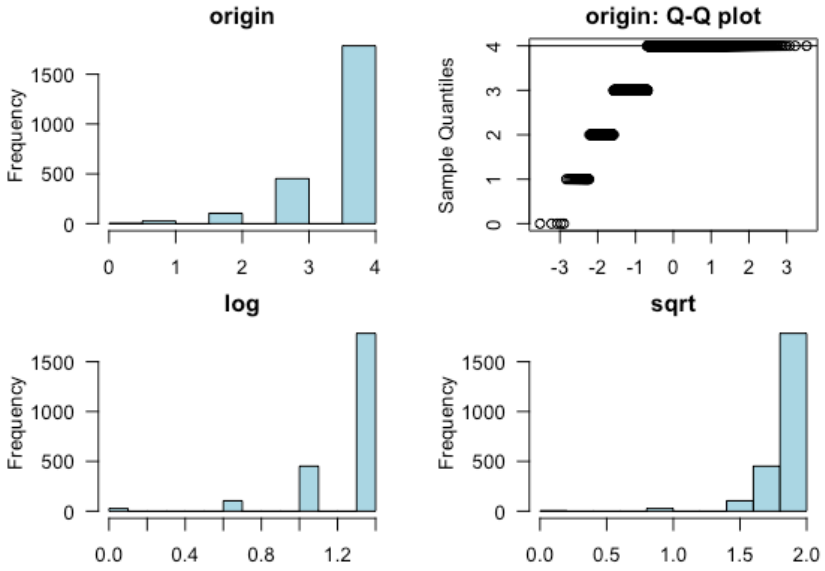
[ALSFRS2]

normality test : Shapiro-Wilk normality test

```
statistic : 0.55904,  p-value : 7.12812E-61
```

skewness and kurtosis

type	skewness	kurtosis
original	-2.244431	8.606259
log transformation	NaN	NaN
sqrt transformation	-3.618126	24.298742



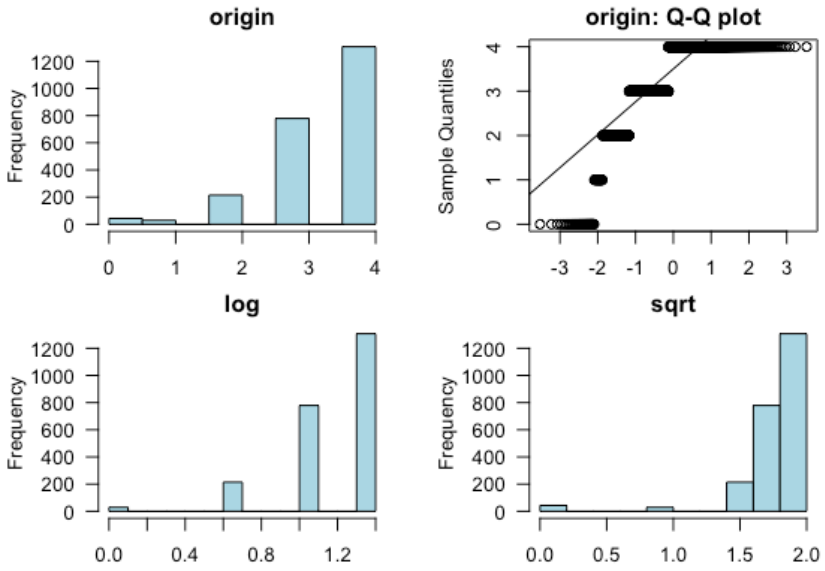
[ALSFRS3]

normality test : Shapiro-Wilk normality test

```
statistic : 0.71181,  p-value : 2.43915E-53
```

skewness and kurtosis

type	skewness	kurtosis
original	-1.668627	6.357821
log transformation	NaN	NaN
sqrt transformation	-3.492396	18.983734



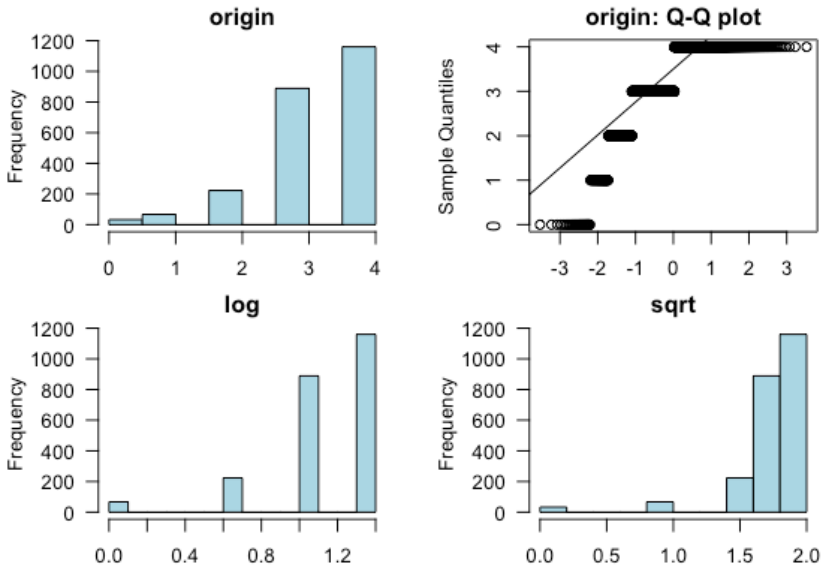
[ALSFRS4]

normality test : Shapiro-Wilk normality test

```
statistic : 0.75214,  p-value : 8.60098E-51
```

skewness and kurtosis

type	skewness	kurtosis
original	-1.410584	5.242386
log transformation	NaN	NaN
sqrt transformation	-3.042965	16.150241



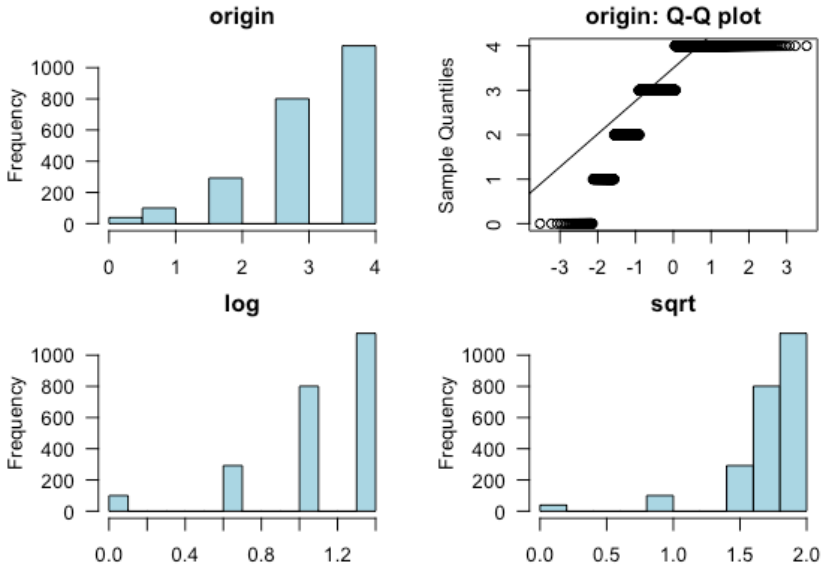
[ALSFRS5]

normality test : Shapiro-Wilk normality test

```
statistic : 0.77326,  p-value : 2.56673E-49
```

skewness and kurtosis

type	skewness	kurtosis
original	-1.255686	4.303177
log transformation	NaN	NaN
sqrt transformation	-2.652390	12.666443



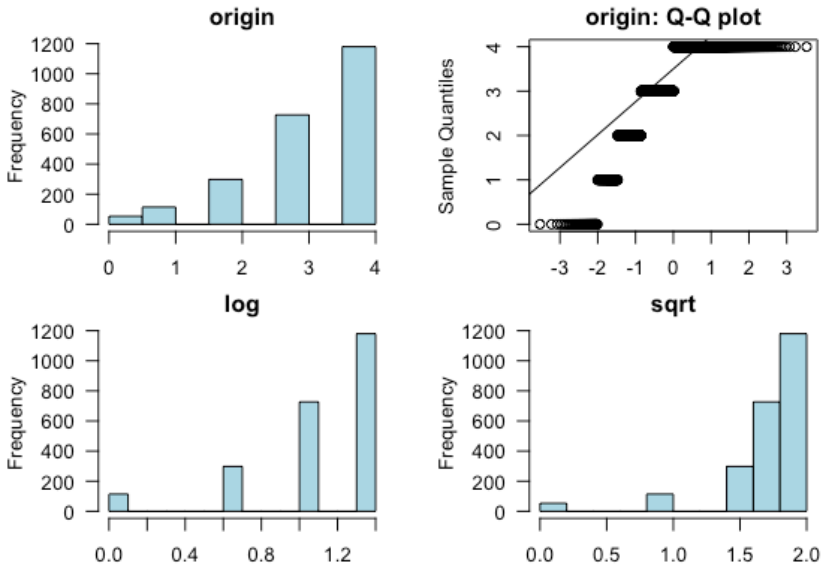
[ALSFRS6]

normality test : Shapiro-Wilk normality test

```
statistic : 0.76647,  p-value : 8.38756E-50
```

skewness and kurtosis

type	skewness	kurtosis
original	-1.286043	4.200638
log transformation	NaN	NaN
sqrt transformation	-2.605095	11.556253



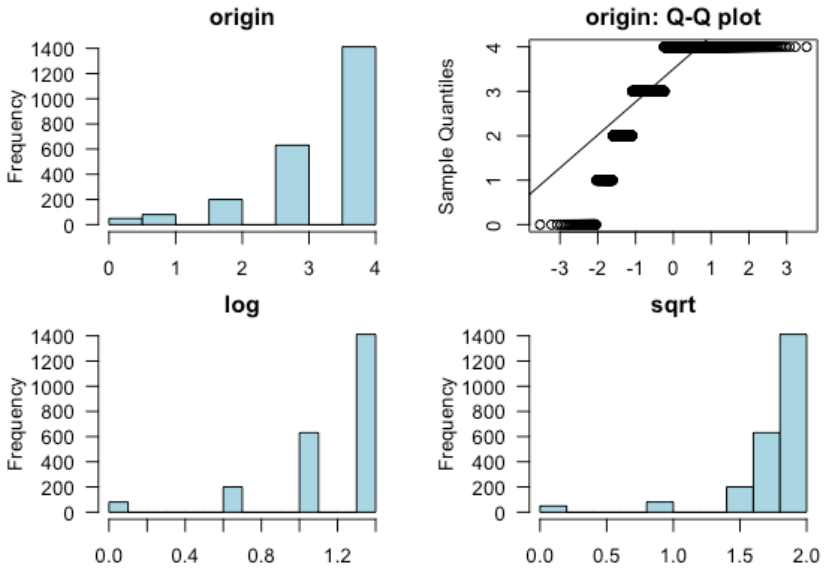
[ALSFRS7]

normality test : Shapiro-Wilk normality test

```
statistic : 0.69116,  p-value : 1.57169E-54
```

skewness and kurtosis

type	skewness	kurtosis
original	-1.708816	5.728035
log transformation	NaN	NaN
sqrt transformation	-3.116444	14.837037



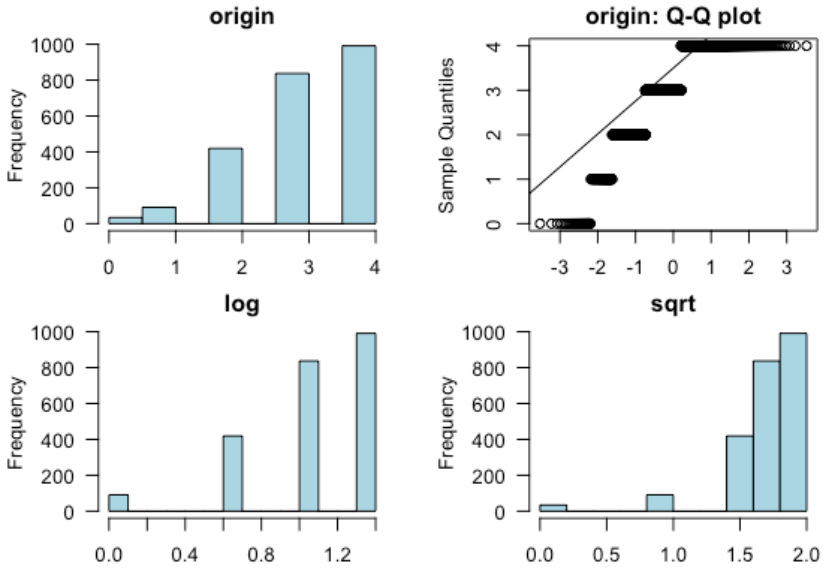
[ALSFRS8]

normality test : Shapiro-Wilk normality test

```
statistic : 0.81361,  p-value : 3.76053E-46
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.9601278	3.586815
log transformation	NaN	NaN
sqrt transformation	-2.3705254	11.684958



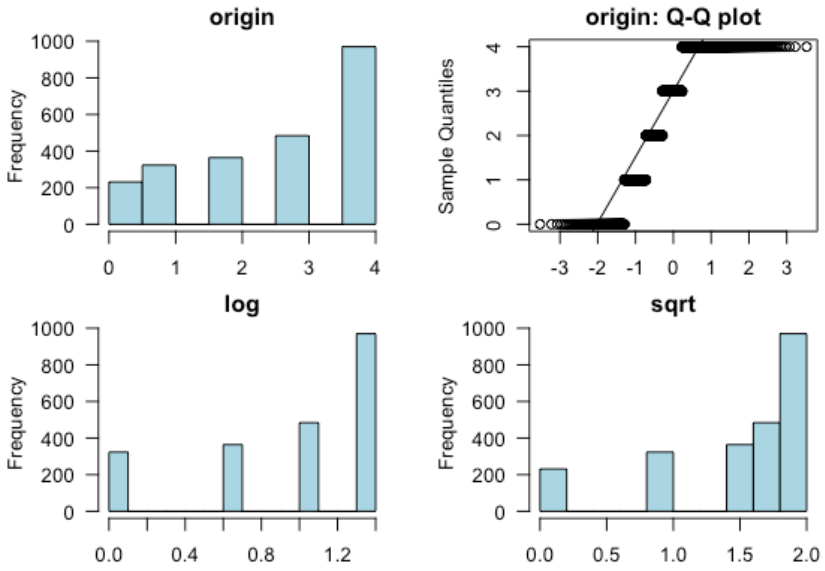
[ALSFRS9]

normality test : Shapiro-Wilk normality test

```
statistic : 0.82716,  p-value : 5.82704E-45
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.6488974	2.092024
log transformation	NaN	NaN
sqrt transformation	-1.4357580	4.142034



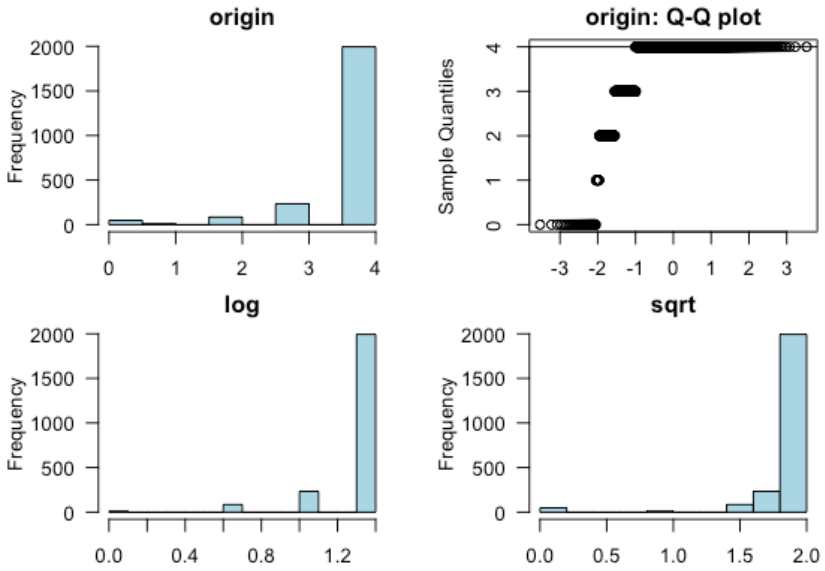
[ALSFRS10]

normality test : Shapiro-Wilk normality test

```
statistic : 0.41113,  p-value : 2.73028E-66
```

skewness and kurtosis

type	skewness	kurtosis
original	-3.453364	15.78238
log transformation	NaN	NaN
sqrt transformation	-4.947365	29.41125



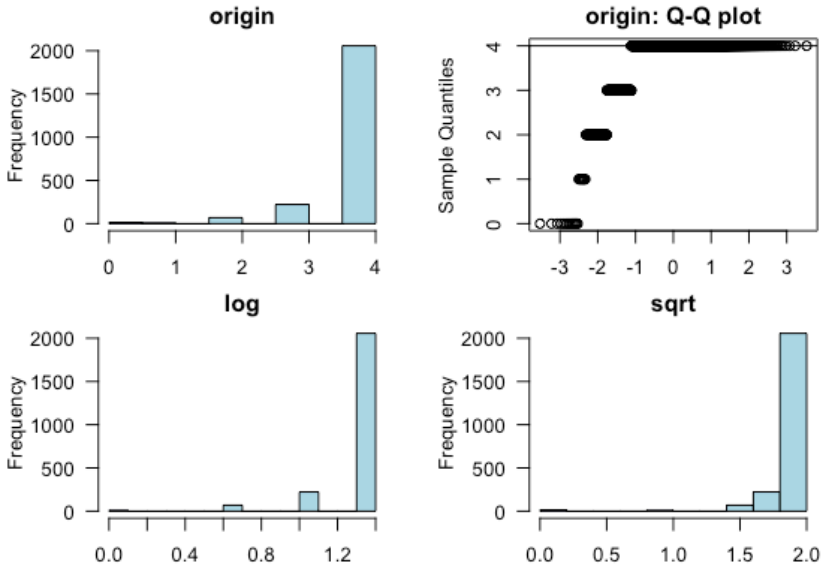
[ALSFRS11]

normality test : Shapiro-Wilk normality test

```
statistic : 0.38354,  p-value : 3.60784E-67
```

skewness and kurtosis

type	skewness	kurtosis
original	-3.751613	19.79364
log transformation	NaN	NaN
sqrt transformation	-6.195364	52.97576



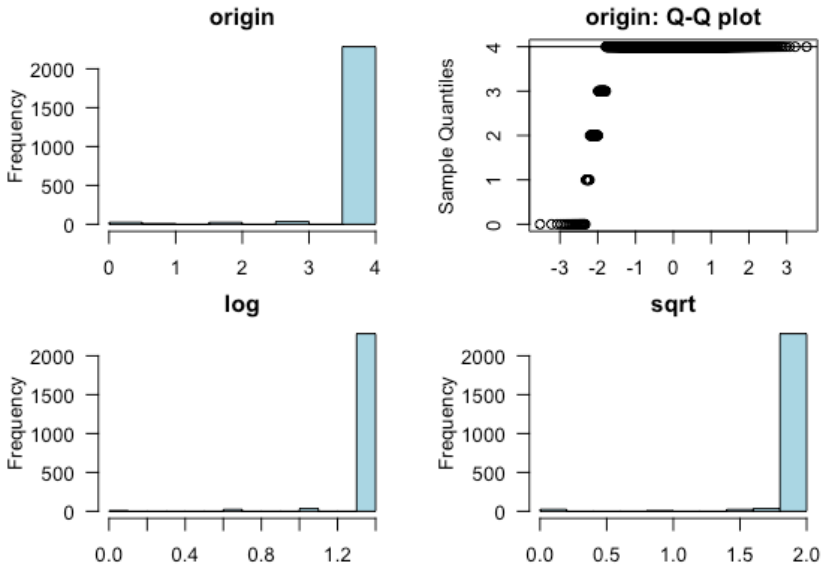
[ALSFRS12]

normality test : Shapiro-Wilk normality test

```
statistic : 0.16376,  p-value : 3.55133E-73
```

skewness and kurtosis

type	skewness	kurtosis
original	-6.632627	48.83758
log transformation	NaN	NaN
sqrt transformation	-8.010665	70.08808



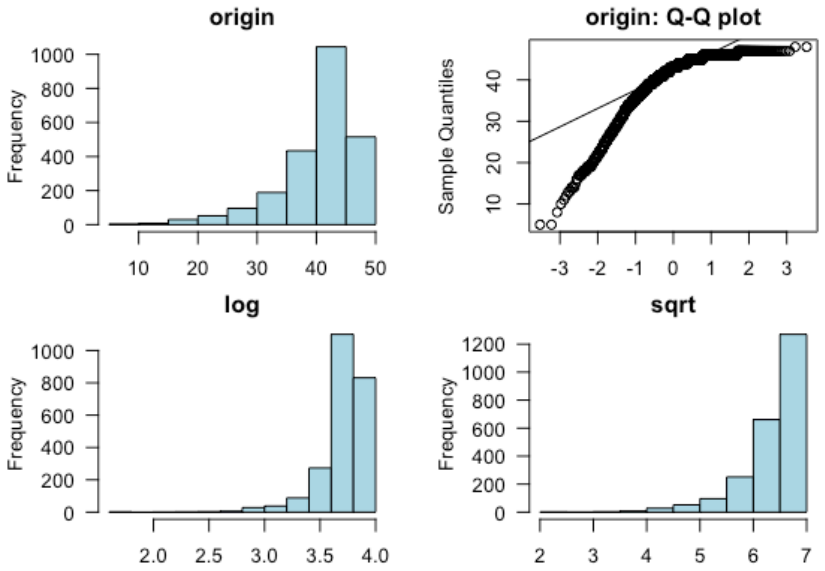
[First ALSFRS-R total]

normality test : Shapiro-Wilk normality test

```
statistic : 0.80567,  p-value : 8.13619E-47
```

skewness and kurtosis

type	skewness	kurtosis
original	-1.838538	6.865773
log transformation	-3.286342	20.611139
sqrt transformation	-2.349192	10.526823



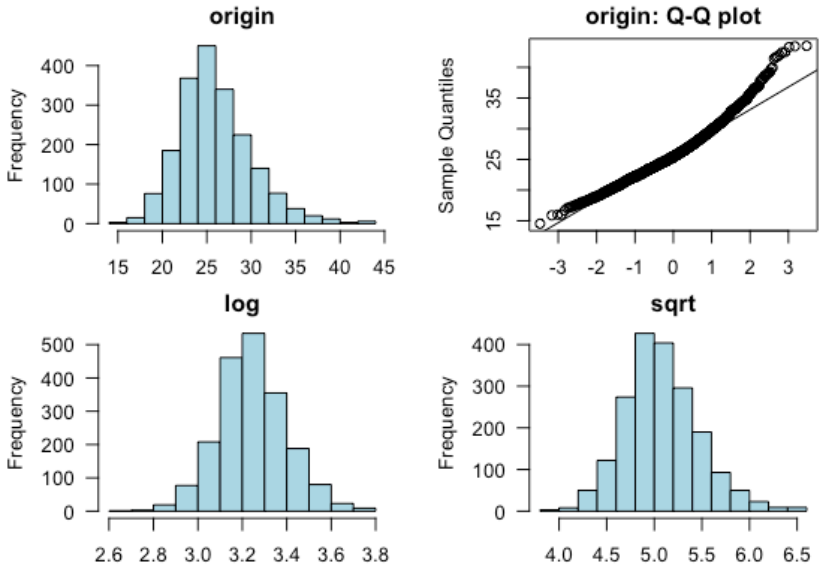
[BMI 2 years prior to illness]

normality test : Shapiro-Wilk normality test

```
statistic : 0.97062,  p-value : 1.39736E-19
```

skewness and kurtosis

type	skewness	kurtosis
original	0.7524503	4.165585
log transformation	0.2031673	3.373713
sqrt transformation	0.4735973	3.636518



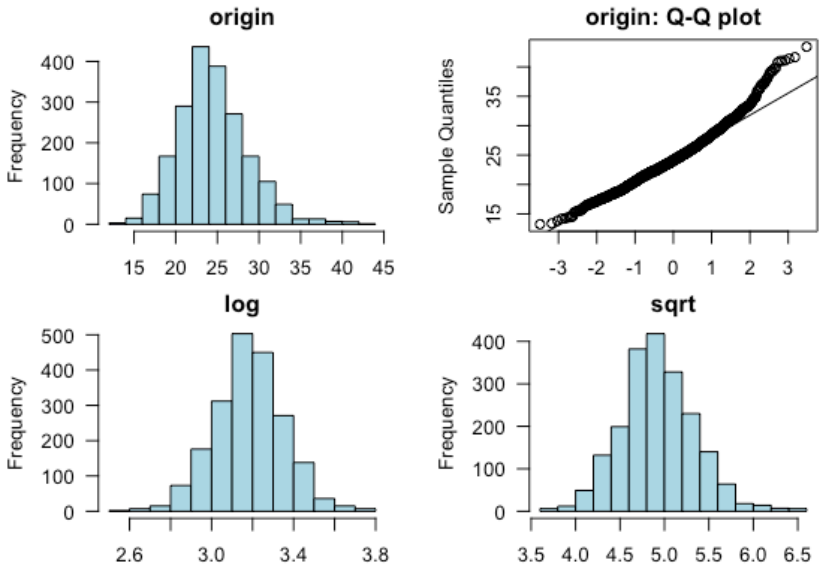
[BMI at diagnosis]

normality test : Shapiro-Wilk normality test

```
statistic : 0.97913,  p-value : 1.55442E-16
```

skewness and kurtosis

type	skewness	kurtosis
original	0.6233883	4.079254
log transformation	0.0049483	3.385226
sqrt transformation	0.3091934	3.562802



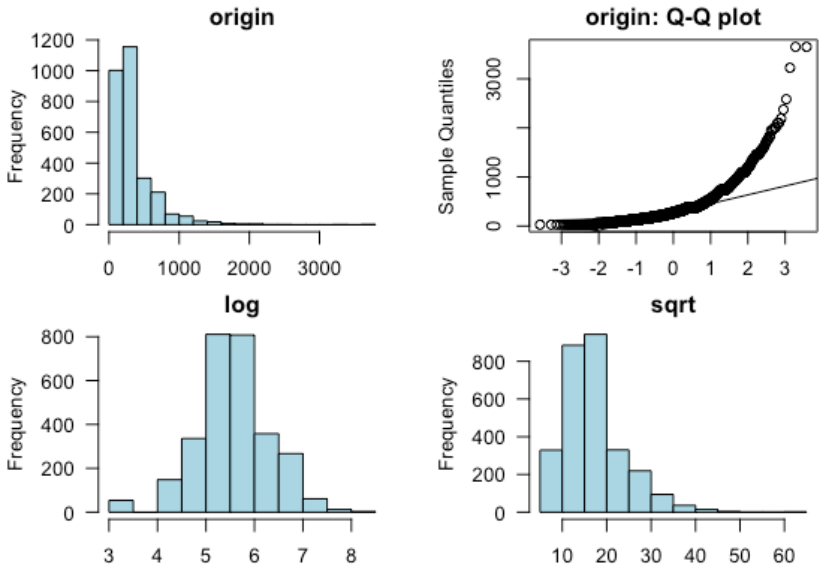
[Delay in diagnosis (days)]

normality test : Shapiro-Wilk normality test

```
statistic : 0.73979,  p-value : 3.55768E-55
```

skewness and kurtosis

type	skewness	kurtosis
original	3.1429618	21.041965
log transformation	-0.1505645	3.298884
sqrt transformation	1.2129054	5.753846



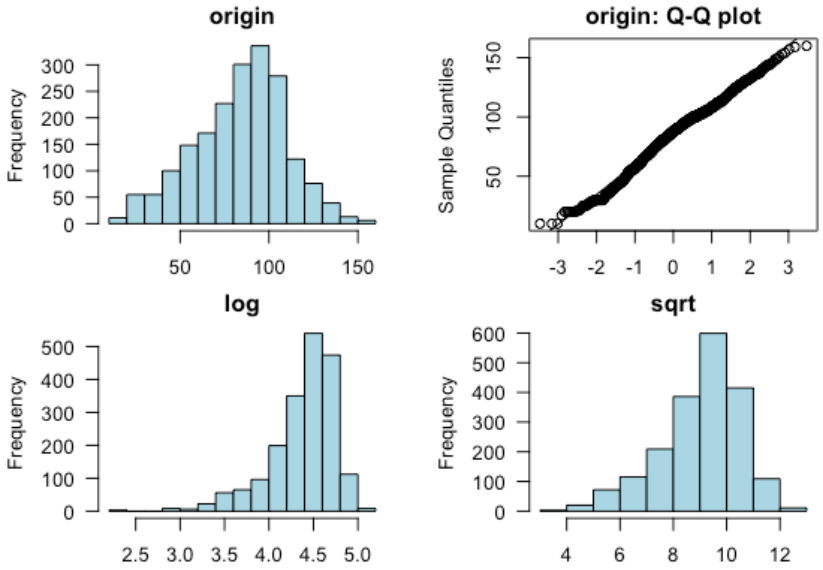
[FVC percent at diagnosis]

normality test : Shapiro-Wilk normality test

```
statistic : 0.98882,  p-value : 4.12542E-11
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.2783920	2.848114
log transformation	-1.4014099	5.699046
sqrt transformation	-0.7686291	3.521085



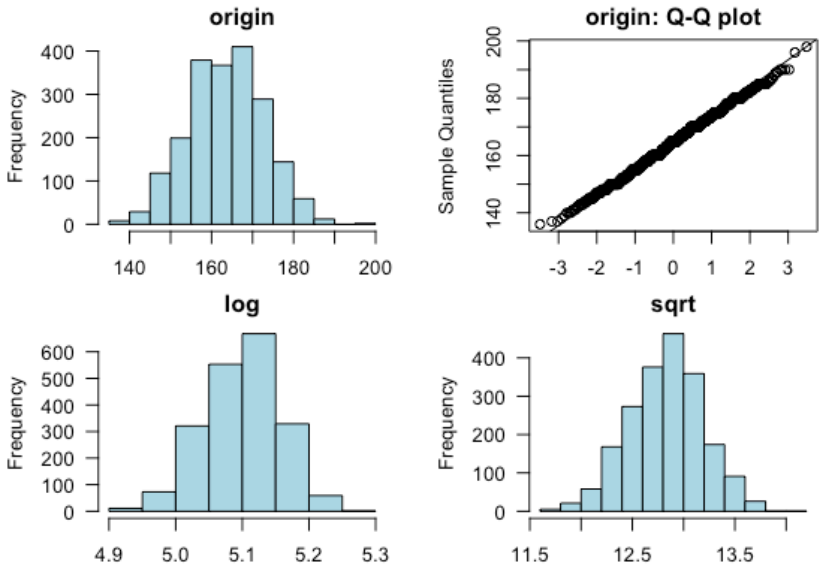
[Height]

normality test : Shapiro-Wilk normality test

```
statistic : 0.99702,  p-value : 0.000610463
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.0171470	2.805405
log transformation	-0.1699096	2.838068
sqrt transformation	-0.0934284	2.811543



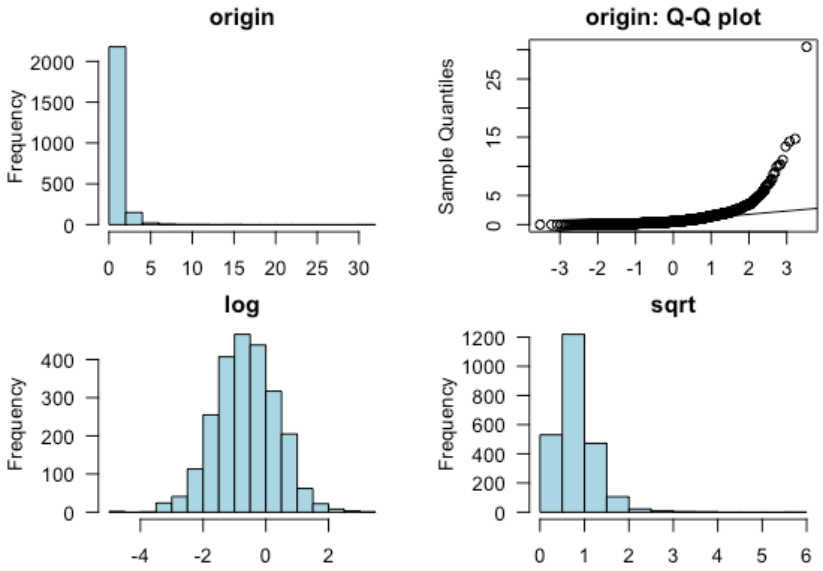
[Rate of decline ALSFRS-R (per month)]

normality test : Shapiro-Wilk normality test

```
statistic : 0.48661,  p-value : 1.20193E-63
```

skewness and kurtosis

type	skewness	kurtosis
original	9.048915	159.79566
log transformation	NaN	NaN
sqrt transformation	2.151759	13.77023



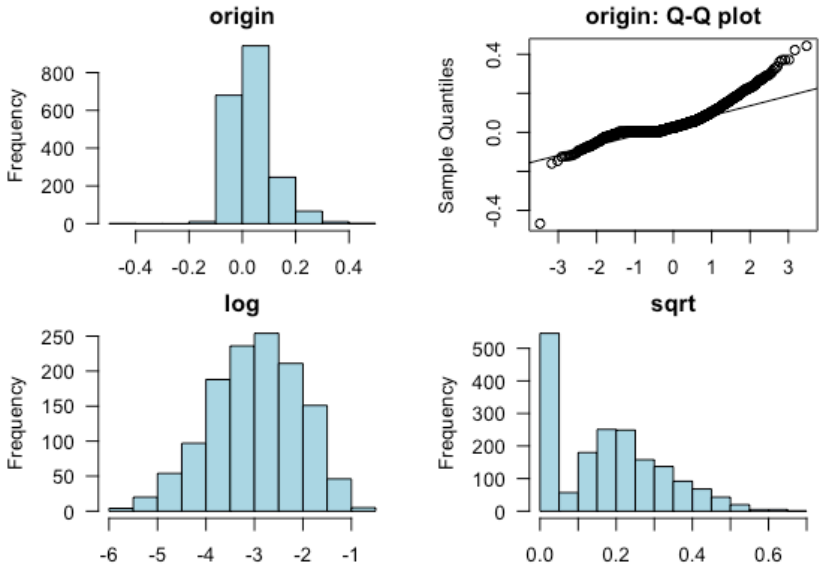
[Rate of decline BMI (per month)]

normality test : Shapiro-Wilk normality test

```
statistic : 0.84826,  p-value : 8.53916E-40
```

skewness and kurtosis

type	skewness	kurtosis
original	1.2863473	7.976611
log transformation	NaN	NaN
sqrt transformation	0.4194388	2.404911



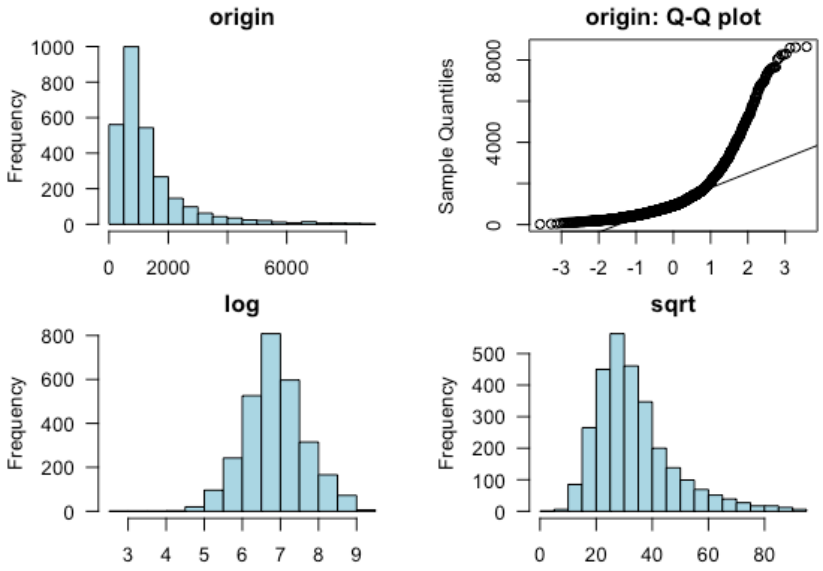
[Survival (days)]

normality test : Shapiro-Wilk normality test

```
statistic : 0.73311,  p-value : 1.38024E-55
```

skewness and kurtosis

type	skewness	kurtosis
original	2.5468653	11.028408
log transformation	0.0426648	3.364052
sqrt transformation	1.3072292	5.106014



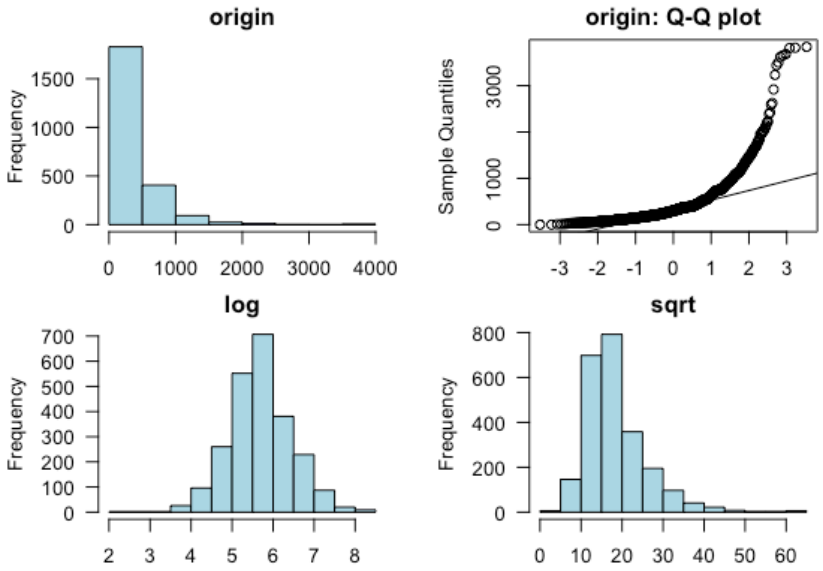
[Time of first ALSFRS-R (days into illness)]

normality test : Shapiro-Wilk normality test

```
statistic : 0.67426,  p-value : 1.76102E-55
```

skewness and kurtosis

type	skewness	kurtosis
original	3.749677	25.143865
log transformation	NaN	NaN
sqrt transformation	1.552927	7.385901



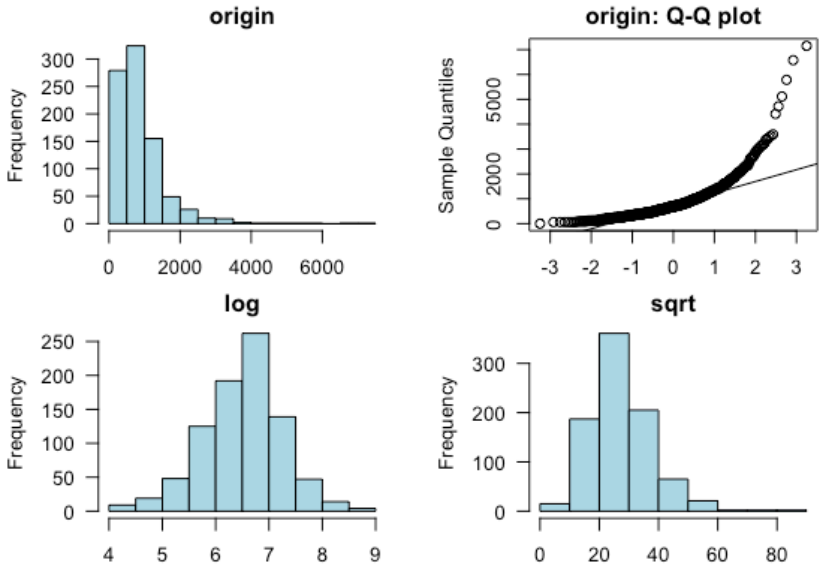
[Time of NIPPV (days into illness)]

normality test : Shapiro-Wilk normality test

```
statistic : 0.75086,  p-value : 4.79887E-34
```

skewness and kurtosis

type	skewness	kurtosis
original	3.132622	19.853501
log transformation	NaN	NaN
sqrt transformation	1.135598	5.958096



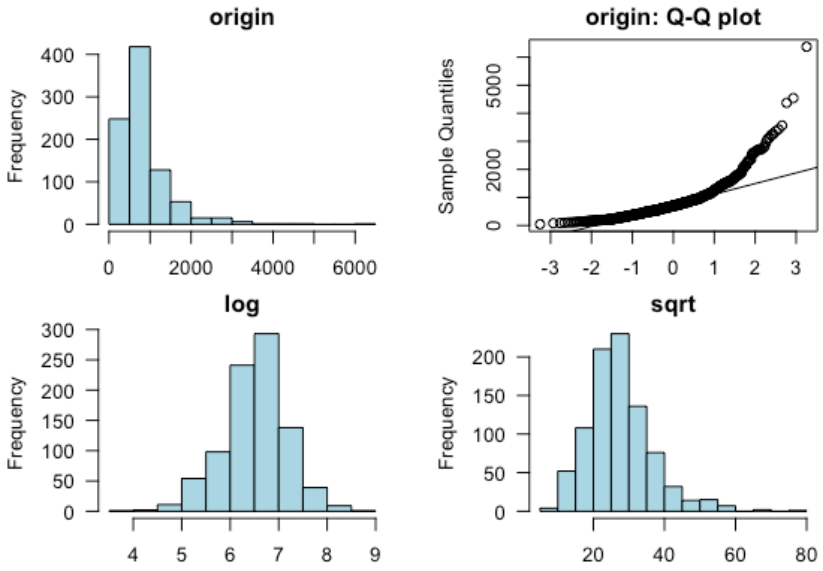
[Time of PEG (days into illness)]

normality test : Shapiro-Wilk normality test

```
statistic : 0.79249,  p-value : 3.87343E-32
```

skewness and kurtosis

type	skewness	kurtosis
original	2.6914360	16.253751
log transformation	-0.2379703	3.609555
sqrt transformation	1.0384974	5.448605



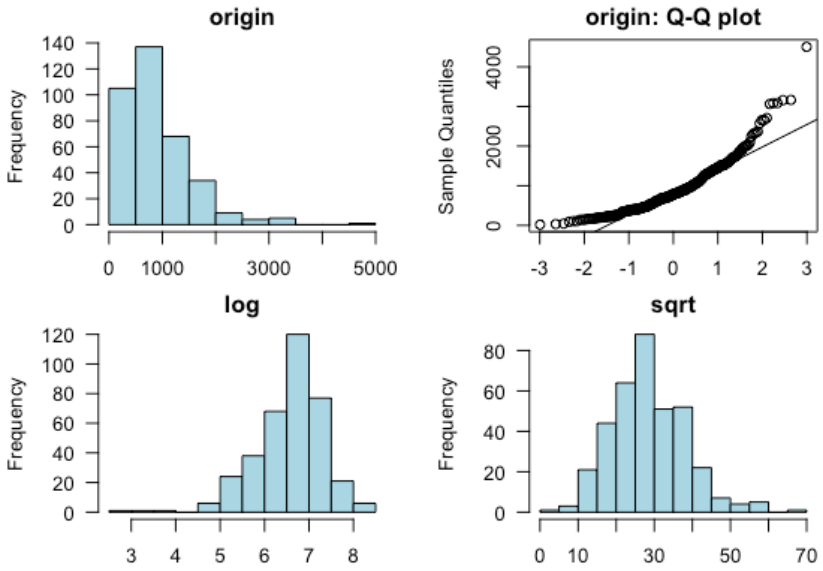
[Time of tracheostomy (days into illness)]

normality test : Shapiro-Wilk normality test

```
statistic : 0.88447,  p-value : 6.78908E-16
```

skewness and kurtosis

type	skewness	kurtosis
original	1.6167809	7.304867
log transformation	-0.8653495	5.052483
sqrt transformation	0.4817271	3.495637



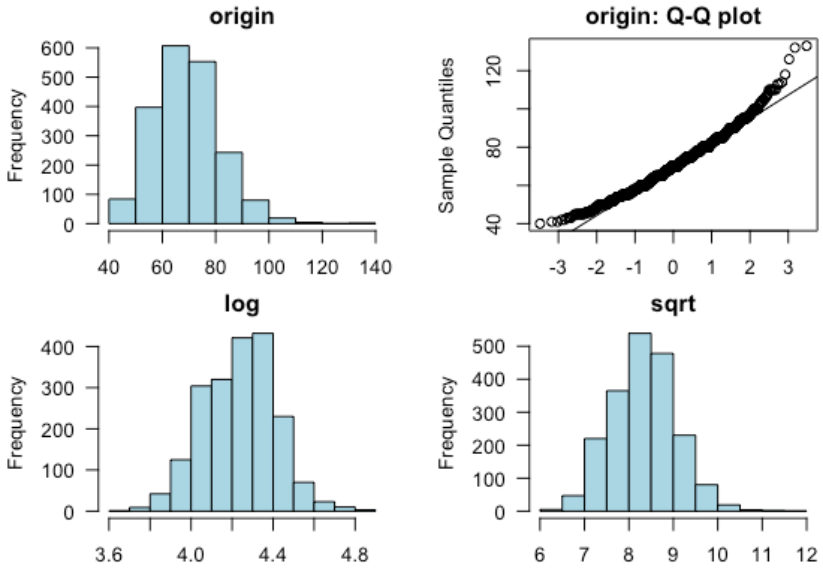
[Weight 2 years prior to illness]

normality test : Shapiro-Wilk normality test

```
statistic : 0.98356,  p-value : 2.29724E-14
```

skewness and kurtosis

type	skewness	kurtosis
original	0.5196063	3.784897
log transformation	-0.0467443	3.012962
sqrt transformation	0.2260243	3.230385



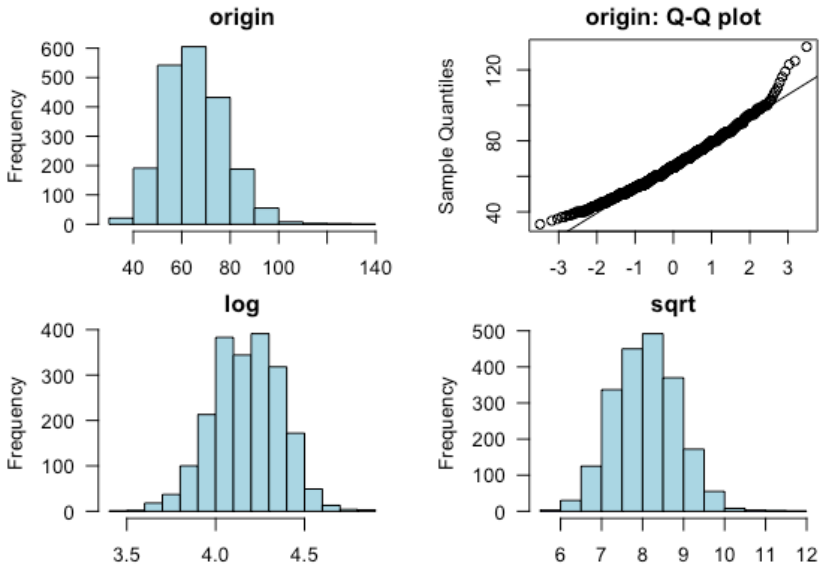
[Weight at diagnosis]

normality test : Shapiro-Wilk normality test

```
statistic : 0.98521,  p-value : 1.00215E-13
```

skewness and kurtosis

type	skewness	kurtosis
original	0.5006059	3.657525
log transformation	-0.1020962	2.997880
sqrt transformation	0.1909131	3.136565



3 Relationship Between Variables

3.1 Correlation Coefficient

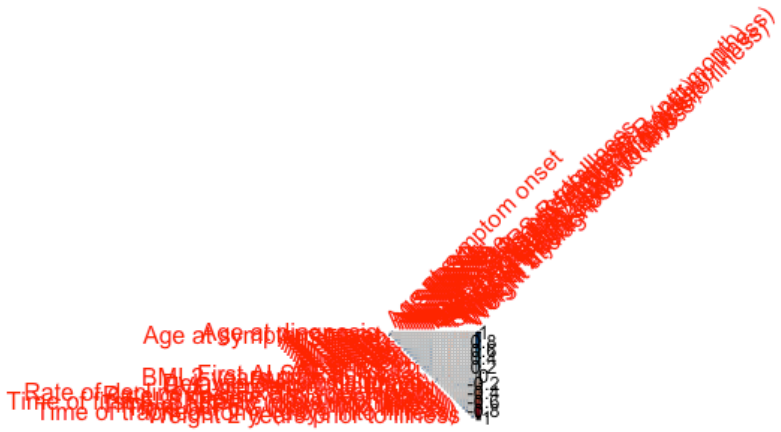
3.1.1 Correlation Coefficient by Variable Combination

Table of correlation coefficients (0.5 or more)

Variable1	Variable2	Correlation Coefficient
Time of tracheostomy (days into illness)	Survival (days)	0.9995719
Age at symptom onset	Age at diagnosis	0.9971255
Time of tracheostomy (days into illness)	Time of PEG (days into illness)	0.9202645
ALSFRS9	ALSFRS8	0.8921703
Weight at diagnosis	Weight 2 years prior to illness	0.8821688
Time of tracheostomy (days into illness)	Time of NIPPV (days into illness)	0.8757077
BMI at diagnosis	BMI 2 years prior to illness	0.8456542
Time of PEG (days into illness)	Time of NIPPV (days into illness)	0.8326545
ALSFRS11	ALSFRS10	0.8323421
First ALSFRS-R total	ALSFRS7	0.8261060
ALSFRS5	ALSFRS4	0.8114505
Time of NIPPV (days into illness)	Survival (days)	0.8070177
Weight at diagnosis	BMI at diagnosis	0.8052250
Time of first ALSFRS-R (days into illness)	Delay in diagnosis (days)	0.8007616
ALSFRS6	ALSFRS5	0.7867916

Time of PEG (days into illness)	Survival (days)	0.7821316
Weight 2 years prior to illness	BMI 2 years prior to illness	0.7649205
ALSFRS3	ALSFRS1	0.7642211
First ALSFRS-R total	ALSFRS6	0.7621598
ALSFRS8	ALSFRS7	0.7327403
ALSFRS9	ALSFRS7	0.7269545
ALSFRS6	ALSFRS4	0.7193552
First ALSFRS-R total	ALSFRS9	0.7063663
ALSFRS12	ALSFRS10	0.7031328
First ALSFRS-R total	ALSFRS5	0.7020301
ALSFRS7	ALSFRS6	0.6981110
ALSFRS2	ALSFRS1	0.6961506
First ALSFRS-R total	ALSFRS8	0.6945116
ALSFRS3	ALSFRS2	0.6741214
Weight 2 years prior to illness	BMI at diagnosis	0.6643290
First ALSFRS-R total	ALSFRS4	0.6641170
ALSFRS12	ALSFRS11	0.6452054
Weight at diagnosis	BMI 2 years prior to illness	0.6376970
ALSFRS7	ALSFRS5	0.5916423
First ALSFRS-R total	ALSFRS11	0.5678575
First ALSFRS-R total	ALSFRS10	0.5495165
ALSFRS7	ALSFRS4	0.5433886
Weight 2 years prior to illness	Height	0.5087521
ALSFRS9	ALSFRS6	0.5054623

3.1.2 Correlation Plot of Numerical Variables



4 Target based Analysis

4.1 Grouped Descriptive Statistics

4.1.1 Grouped Numerical Variables

4.1.2 Grouped Categorical Variables

4.2 Grouped Relationship Between Variables

4.2.1 Grouped Correlation Coefficient

4.2.2 Grouped Correlation Plot of Numerical Variables

Exploratory Data Analysis Report on the ERRALS Registry

Report by dlookr package

2021-07-23

1 Introduction

The EDA Report provides exploratory data analysis information on objects that inherit data.frame and data.frame.

1.1 Information of Dataset

The dataset that generated the EDA Report is an **'data.frame'** object. It consists of **1,097 observations** and **66 variables**.

1.2 Information of Variables

The variable information of the data set that generated the EDA Report is shown in the following table.:

Information of Variables

variables	types	missing_count	missing_percent	unique_count	unique_rate

Age at diagnosis	numeric	2	0.1823154	459	0.4184139
Age at symptom onset	numeric	2	0.1823154	467	0.4257065
ALSFRS1	integer	102	9.2980857	6	0.0054695
ALSFRS2	integer	102	9.2980857	6	0.0054695
ALSFRS3	integer	102	9.2980857	6	0.0054695
ALSFRS4	integer	102	9.2980857	6	0.0054695
ALSFRS5	integer	0	0.0000000	5	0.0045579
ALSFRS6	integer	102	9.2980857	6	0.0054695
ALSFRS7	integer	102	9.2980857	6	0.0054695
ALSFRS8	integer	102	9.2980857	6	0.0054695
ALSFRS9	integer	102	9.2980857	6	0.0054695
ALSFRS10	integer	102	9.2980857	6	0.0054695
ALSFRS11	integer	102	9.2980857	6	0.0054695
ALSFRS12	integer	102	9.2980857	6	0.0054695
First ALSFRS-R total	integer	102	9.2980857	45	0.0410210
Anatomical level at onset	character	24	2.1877849	8	0.0072926
Ataxia	character	0	0.0000000	1	0.0009116
BMI 2 years prior to illness	numeric	261	23.7921604	489	0.4457612
BMI at diagnosis	numeric	199	18.1403829	560	0.5104831
C9orf72 status	character	671	61.1668186	3	0.0027347
Cancer	character	0	0.0000000	2	0.0018232
Cancer type	character	976	88.9699180	42	0.0382862
Chorea	character	0	0.0000000	2	0.0018232
Clinical type at one year	character	24	2.1877849	7	0.0063810
Clinical type at onset	logical	1097	100.0000000	1	0.0009116
Cognitive impairment present	character	0	0.0000000	2	0.0018232
Cognitive status 1	character	0	0.0000000	2	0.0018232
Cognitive status 2	logical	1097	100.0000000	1	0.0009116
COPD	character	0	0.0000000	2	0.0018232
Delay in diagnosis (days)	integer	5	0.4557885	170	0.1549681
Diabetes	character	0	0.0000000	2	0.0018232
Education	character	208	18.9608022	7	0.0063810

El Escorial category at diagnosis	character	55	5.0136737	5	0.0045579
El Escorial category at visit 2	logical	1097	100.0000000	1	0.0009116
El Escorial category at visit 3	logical	1097	100.0000000	1	0.0009116
Family history of ALS	character	0	0.0000000	2	0.0018232
FVC percent at diagnosis	numeric	581	52.9626253	171	0.1558797
Height	integer	162	14.7675479	48	0.0437557
Hypertension	character	0	0.0000000	2	0.0018232
Hyperthyroid	character	0	0.0000000	2	0.0018232
Hypothyroid	character	0	0.0000000	2	0.0018232
Initial diagnosis was PLS	character	0	0.0000000	1	0.0009116
Marital status	character	764	69.6444850	6	0.0054695
Mutated gene	character	1051	95.8067457	5	0.0045579
Mutation amino acid change	character	1093	99.6353692	5	0.0045579
Mutation present	character	0	0.0000000	2	0.0018232
NIPPV	character	1	0.0911577	3	0.0027347
Onset side	character	431	39.2889699	4	0.0036463
Parkinsonism	character	0	0.0000000	2	0.0018232
PEG inserted	character	1	0.0911577	3	0.0027347
Place of birth	logical	1097	100.0000000	1	0.0009116
Place of residence	logical	1097	100.0000000	1	0.0009116
Rate of decline ALSFRS-R (per month)	numeric	105	9.5715588	879	0.8012762
Rate of decline BMI (per month)	numeric	272	24.7948952	472	0.4302644
Sex	character	0	0.0000000	2	0.0018232
Site of symptom onset	character	24	2.1877849	4	0.0036463
Smoker	character	924	84.2297174	4	0.0036463
Survival (days)	integer	290	26.4357338	507	0.4621696
Time of first ALSFRS-R (days into illness)	integer	105	9.5715588	605	0.5515041
Time of NIPPV (days into illness)	integer	847	77.2105743	231	0.2105743
Time of PEG (days into illness)	integer	873	79.5806746	207	0.1886964

Time of tracheostomy (days into illness)	integer	985	89.7903373	107	0.0975387
Tracheostomy	character	84	7.6572470	3	0.0027347
Vital status	character	88	8.0218778	3	0.0027347
Weight 2 years prior to illness	numeric	250	22.7894257	87	0.0793072
Weight at diagnosis	numeric	186	16.9553327	119	0.1084777

The target variable of the data is **'Clinical type at one year'**, and the data type of the variable is **character**.

1.3 About EDA Report

EDA reports provide information and visualization results that support the EDA process. In particular, it provides a variety of information to understand the relationship between the target variable and the rest of the variables of interest.

2 Univariate Analysis

2.1 Descriptive Statistics

edaData

66 Variables 1097 Observations

Age at diagnosis

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
1095	2	458	1	67.81	12.67	47.34	52.32	60.88	69.58	75.67	81.17	83.86

lowest : 28.75000 28.83333 31.16667 34.08333 34.25000 , highest: 90.00000 90.91667 91.66667 92.16667 95.08333

Age at symptom onset

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
1095	2	466	1	66.64	12.78	45.78	50.70	59.75	68.50	74.67	80.00	82.69

lowest : 27.33333 27.83333 28.91667 33.25000 33.41667 , highest: 89.58333 90.00000 90.25000 91.08333 91.16667

ALSFRS1

n	missing	distinct	Info	Mean	Gmd
995	102	5	0.863	3.13	1.067

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     43    39   129   319   465
Proportion 0.043 0.039 0.130 0.321 0.467
```

ALSFRS2

n	missing	distinct	Info	Mean	Gmd
995	102	5	0.646	3.534	0.725

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     11    26    84   174   700
Proportion 0.011 0.026 0.084 0.175 0.704
```

ALSFRS3

n	missing	distinct	Info	Mean	Gmd
995	102	5	0.837	3.187	1.045

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     32    27   186   228   522
Proportion 0.032 0.027 0.187 0.229 0.525
```

ALSFRS4

n	missing	distinct	Info	Mean	Gmd
995	102	5	0.838	3.222	0.9834

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     27    56    82   334   496
Proportion 0.027 0.056 0.082 0.336 0.498
```

ALSFRS5

n	missing	distinct	Info	Mean	Gmd
1097	0	5	0.926	2.641	1.499

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency    152    86   157   311   391
Proportion 0.139 0.078 0.143 0.284 0.356
```

ALSFRS6

n	missing	distinct	Info	Mean	Gmd
995	102	5	0.925	2.733	1.298

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     59   117   176   322   321
Proportion 0.059 0.118 0.177 0.324 0.323
```

ALSFRS7

n	missing	distinct	Info	Mean	Gmd
995	102	5	0.802	3.246	1.041

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     39    47   110   233   566
Proportion 0.039 0.047 0.111 0.234 0.569
```

ALSFRS8

n	missing	distinct	Info	Mean	Gmd
995	102	5	0.912	2.862	1.139

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     32    54   279   284   346
Proportion 0.032 0.054 0.280 0.285 0.348
```

ALSFRS9

n	missing	distinct	Info	Mean	Gmd
995	102	5	0.943	2.28	1.667

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency    156   236    80   219   304
Proportion 0.157 0.237 0.080 0.220 0.306
```

ALSFRS10

n	missing	distinct	Info	Mean	Gmd
995	102	5	0.575	3.572	0.7035

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     14    52    31   152   746
Proportion 0.014 0.052 0.031 0.153 0.750
```

ALSFRS11

n	missing	distinct	Info	Mean	Gmd
995	102	5	0.429	3.727	0.4773

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency      9    12    51    98   825
Proportion 0.009 0.012 0.051 0.098 0.829
```

ALSFRS12

n	missing	distinct	Info	Mean	Gmd
995	102	5	0.207	3.848	0.2878

lowest : 0 1 2 3 4 , highest: 0 1 2 3 4

```
Value          0     1     2     3     4
Frequency     14     6    23    31   921
Proportion 0.014 0.006 0.023 0.031 0.926
```

First ALSFRS-R total

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
995	102	44	0.997	38.25	8.11	23	28	35	40	44	46	46

lowest : 1 2 3 6 8 , highest: 44 45 46 47 48

Anatomical level at onset

n	missing	distinct
1073	24	7

lowest :	bulbar	lower_limbs	lower_limbs_distal	respiratory	upper_limbs
highest:	lower_limbs_distal	respiratory	upper_limbs	upper_limbs_distal	upper_limbs_proximal

```
Value                    bulbar          lower_limbs   lower_limbs_distal
Frequency                   376                   77                  270
Proportion                0.350                0.072                0.252

Value               respiratory          upper_limbs   upper_limbs_distal
```

```
Frequency                    23                  66                 196
Proportion                0.021               0.062               0.183

Value      upper_limbs_proximal
Frequency                    65
Proportion                0.061
```

Ataxia

n	missing	distinct	value
1097	0	1	no

```
Value        no
Frequency  1097
Proportion    1
```

BMI 2 years prior to illness

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
836	261	488	1	25.83	4.161	20.31	21.43	23.34	25.43	27.74	30.47	32.15

lowest : 15.04748 16.38470 16.43655 17.36044 18.35938 , highest: 41.14286 41.52249 42.32413 43.28255 44.08163

BMI at diagnosis

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
898	199	559	1	24.44	4.348	18.38	19.88	21.89	24.22	26.84	29.32	31.05

lowest : 13.77778 13.95776 14.17234 14.66667 14.69238 , highest: 40.56247 40.90066 42.32413 43.28255 44.08163

C9orf72 status

n	missing	distinct
426	671	2

```
Value        EXP    WT
Frequency     22   404
Proportion 0.052 0.948
```

Cancer

n	missing	distinct
1097	0	2

```
Value         no   yes
Frequency    975   122
Proportion 0.889 0.111
```

Cancer type

n	missing	distinct
121	976	41

lowest :	acoustic_neurinoma	adrenal	adrenal;breast	adrenal;ovary;thyroid	bile_duct;melanoma
highest:	skin_cancer	stomach	thymoma	thyroid	uterus

Chorea

n	missing	distinct
1097	0	2

```
Value         no   yes
Frequency   1095     2
Proportion 0.998 0.002
```

Clinical type at one year

n	missing	distinct
1073	24	6

lowest :	bulbar	classical	flailArm	flailLeg	pyramidal
highest:	classical	flailArm	flailLeg	pyramidal	respiratory

```
Value          bulbar   classical    flailArm    flailLeg   pyramidal respiratory
Frequency         376         462          52         109          50          24
Proportion      0.350       0.431       0.048       0.102       0.047       0.022
```

Cognitive impairment present

n	missing	distinct
1097	0	2

```
Value         no   yes
Frequency   1000    97
Proportion 0.912 0.088
```

Cognitive status 1

n	missing	distinct

1097	0	2

```
Value         FTD normal
Frequency      97   1000
Proportion  0.088  0.912
```

COPD

n	missing	distinct
1097	0	2

```
Value         no   yes
Frequency   1000    97
Proportion 0.912 0.088
```

Delay in diagnosis (days)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
1092	5	169	0.999	423.5	371.7	90.0	120.0	182.0	305.0	486.2	820.9	1201.5

lowest : 0 28 30 31 59 , highest: 2739 2891 3652 4018 4171

Diabetes

n	missing	distinct
1097	0	2

```
Value         no   yes
Frequency   1012    85
Proportion 0.923 0.077
```

Education

n	missing	distinct
889	208	6

lowest :	11years	13years	18years	5years	8years
highest:	13years	18years	5years	8years	lessthan5years

```
Value            11years       13years       18years        5years
Frequency             28           223            51           280
Proportion         0.031         0.251         0.057         0.315

Value             8years lessthan5years
Frequency            260            47
Proportion         0.292         0.053
```

El Escorial category at diagnosis

n	missing	distinct
1042	55	4

```
Value                   definite   possible   probable
Frequency                    336        222        345
Proportion                 0.322      0.213      0.331

Value      probable_labSupported
Frequency                    139
Proportion                 0.133
```

Family history of ALS

n	missing	distinct
1097	0	2

```
Value         no   yes
Frequency   1046    51
Proportion 0.954 0.046
```

FVC percent at diagnosis

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
516	581	170	1	84.51	27.97	40.00	48.00	69.00	89.05	100.25	113.00	120.25

lowest : 13.0 15.0 19.0 20.0 21.0 , highest: 139.9 143.0 146.0 150.0 151.0

Height

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
935	162	47	0.995	167.9	9.869	152	157	162	168	174	179	182

lowest : 143 144 145 146 148 , highest: 186 187 188 190 195

Hypertension

n	missing	distinct
1097	0	2

```
Value         no   yes
Frequency    624   473
Proportion 0.569 0.431
```

Hyperthyroid

n	missing	distinct
1097	0	2

```
Value         no   yes
Frequency   1078    19
Proportion 0.983 0.017
```

Hypothyroid

n	missing	distinct
1097	0	2

```
Value         no   yes
Frequency   1041    56
Proportion 0.949 0.051
```

Initial diagnosis was PLS

n	missing	distinct	value
1097	0	1	no

```
Value        no
Frequency  1097
Proportion    1
```

Marital status

n	missing	distinct
333	764	5

lowest :	divorced	married	neverMarried	separated	widowed
highest:	divorced	married	neverMarried	separated	widowed

```
Value          divorced      married neverMarried    separated      widowed
Frequency            10          246           25            3           49
Proportion        0.030        0.739        0.075        0.009        0.147
```

Mutated gene

n	missing	distinct
46	1051	4

```
Value      C9ORF72     FUS    SOD1  TARDBP
Frequency       22       7      14       3
Proportion   0.478   0.152   0.304   0.065
```

Mutation amino acid change

n	missing	distinct
4	1093	4

```
Value      G357D  G94D  N66T S134N
Frequency      1     1     1     1
Proportion  0.25  0.25  0.25  0.25
```

Mutation present

n	missing	distinct
1097	0	2

```
Value         no   yes
Frequency   1051    46
Proportion 0.958 0.042
```

NIPPV

n	missing	distinct
1096	1	2

```
Value        no  yes
Frequency   833  263
Proportion 0.76 0.24
```

Onset side

n	missing	distinct
666	431	3

```
Value       both  left right
Frequency     52   407   207
Proportion 0.078 0.611 0.311
```

Parkinsonism

n	missing	distinct
1097	0	2

```
Value         no   yes
Frequency   1067    30
Proportion 0.973 0.027
```

PEG inserted

n	missing	distinct
1096	1	2

```
Value         no   yes
Frequency    863   233
Proportion 0.787 0.213
```

Rate of decline ALSFRS-R (per month)

```
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25
     992      105      878        1   0.9751   0.9964  0.08816  0.15115  0.32014
     .50      .75      .90      .95
 0.60411  1.20874  2.04522  2.71668
```

lowest :	0.00000000	0.02934103	0.03588235	0.03836478	0.03880407
highest:	6.25641026	6.82236842	10.16666667	17.42857143	30.50000000

```
Value        0.0   0.5   1.0   1.5   2.0   2.5   3.0   3.5   4.0   4.5   5.0   5.5
Frequency    200   365   195    90    59    36    12     6    10     7     6     1
Proportion 0.202 0.368 0.197 0.091 0.059 0.036 0.012 0.006 0.010 0.007 0.006 0.001

Value        6.5   7.0  10.0  17.5  30.5
Frequency      1     1     1     1     1
Proportion 0.001 0.001 0.001 0.001 0.001
```

For the frequency table, variable is rounded to the nearest 0.5

Rate of decline BMI (per month)

```
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25
     825      272      471    0.926  0.03973  0.06109  0.00000  0.00000  0.00000
     .50      .75      .90      .95
 0.01354  0.06510  0.12226  0.16470
```

lowest :	-0.2025815	-0.1136962	-0.1078196	-0.1004388	-0.0984190
highest:	0.2797582	0.3063702	0.3258176	0.3364091	0.3671486

Sex

n	missing	distinct
1097	0	2

```
Value          F     M
Frequency    485   612
Proportion 0.442 0.558
```

Site of symptom onset 

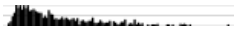
n	missing	distinct
1073	24	3

```
Value          bulbar respiratory      spinal
Frequency         376          23         674
Proportion      0.350       0.021       0.628
```

Smoker 

n	missing	distinct
173	924	3

```
Value      current  former   never
Frequency       92      20      61
Proportion   0.532   0.116   0.353
```

Survival (days) 

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
807	290	506	1	1272	1022	273.0	332.6	548.0	958.0	1742.5	2721.4	3246.6

lowest : 121 152 165 178 183 , highest: 4688 4699 4747 4839 5843

Time of first ALSFRS-R (days into illness) 

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
992	105	604	1	484.1	396.1	108.6	146.0	220.0	366.0	590.5	944.0	1271.5

lowest : 8 18 25 31 35 , highest: 2706 2869 3174 3606 3846

Time of NIPPV (days into illness) 

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
250	847	230	1	733	506.9	188.4	270.7	404.5	608.0	878.2	1429.5	1793.0

lowest : 44 90 102 104 143 , highest: 2234 2272 2571 2641 2814

Time of PEG (days into illness)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
224	873	206	1	738.3	469.8	229.6	299.3	434.5	636.0	957.8	1234.5	1628.1

lowest : -76 45 161 193 195 , highest: 2217 2234 2277 2560 2571

Time of tracheostomy (days into illness)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
112	985	106	1	764.1	507.5	250.9	273.5	402.5	659.0	1015.5	1329.7	1670.2

lowest : 165 196 203 211 213 , highest: 1859 2116 2239 2321 2593

Tracheostomy

n	missing	distinct
1013	84	2

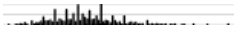
```
Value         no   yes
Frequency    714   299
Proportion 0.705 0.295
```

Vital status

n	missing	distinct
1009	88	2

```
Value      alive  dead
Frequency    346   663
Proportion 0.343 0.657
```

Weight 2 years prior to illness



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
847	250	86	0.999	73.03	14.84	54.0	56.0	64.0	72.0	80.0	90.0	96.7

lowest : 40 43 44 45 46 , highest: 115 116 120 126 135

Weight at diagnosis

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
911	186	118	0.999	69.11	14.97	48.0	53.0	60.0	69.0	78.0	85.0	91.5

lowest : 30 31 33 35 38 , highest: 107 111 112 120 135

Variables with all observations missing:

`Clinical type at onset`, `Cognitive status 2`, `El Escorial category at visit 2`, `El Escorial category at visit 3`, `Place of birth`, `Place of residence`

2.2 Normality Test of Numerical Variables

2.2.1 Statistics and Visualization of (Sample) Data

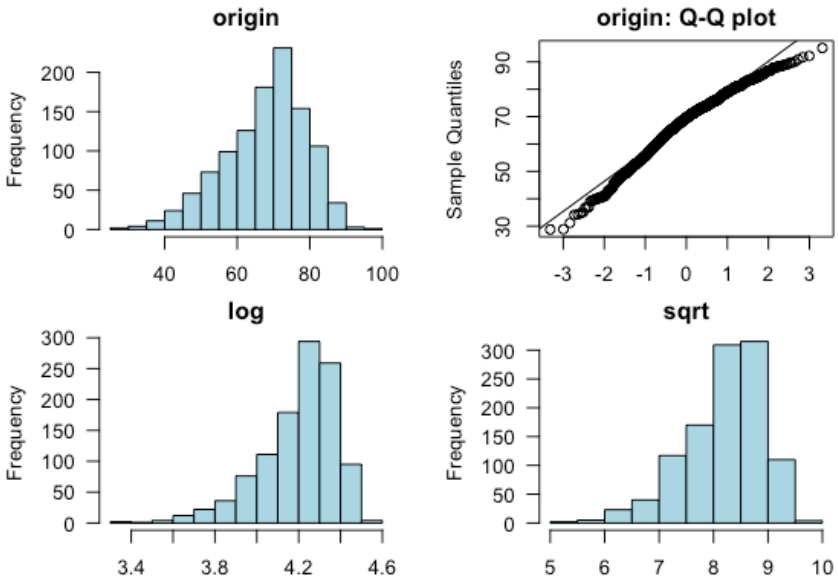
[Age at diagnosis]

normality test : Shapiro-Wilk normality test

```
statistic : 0.97687,  p-value : 3.27035E-12
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.5686429	3.061584
log transformation	-1.1154819	4.549626
sqrt transformation	-0.8234084	3.627880



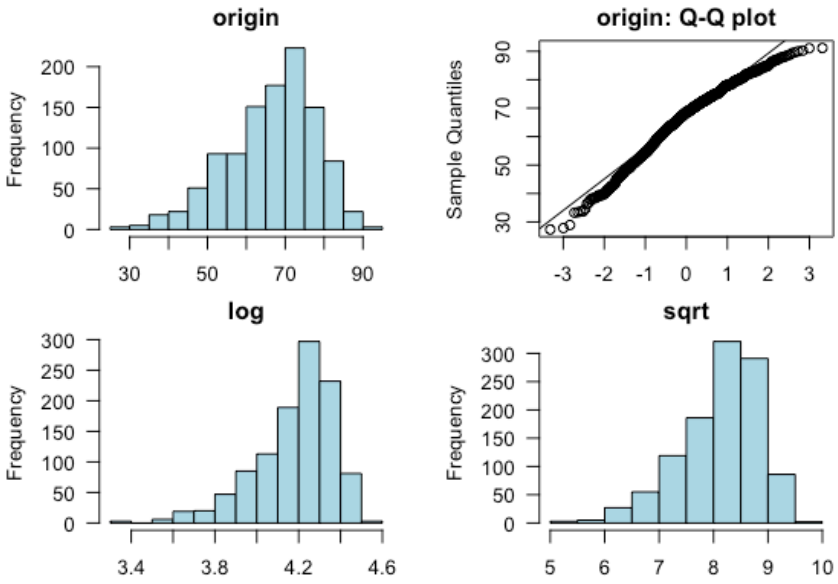
[Age at symptom onset]

normality test : Shapiro-Wilk normality test

```
statistic : 0.97588,  p-value : 1.5656E-12
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.5716038	3.023867
log transformation	-1.1247779	4.547863
sqrt transformation	-0.8283748	3.599474



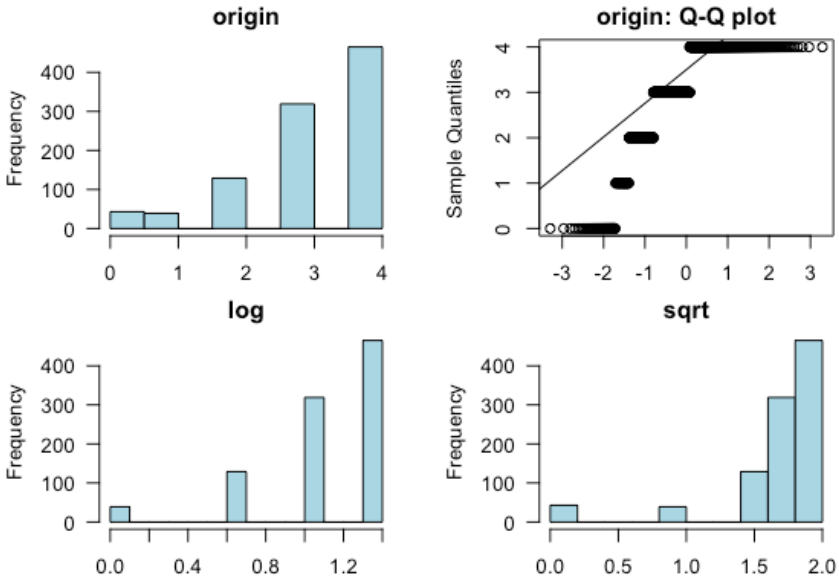
[ALSFRS1]

normality test : Shapiro-Wilk normality test

```
statistic : 0.77009,  p-value : 3.96979E-35
```

skewness and kurtosis

type	skewness	kurtosis
original	-1.324877	4.284758
log transformation	NaN	NaN
sqrt transformation	-2.558313	9.954455



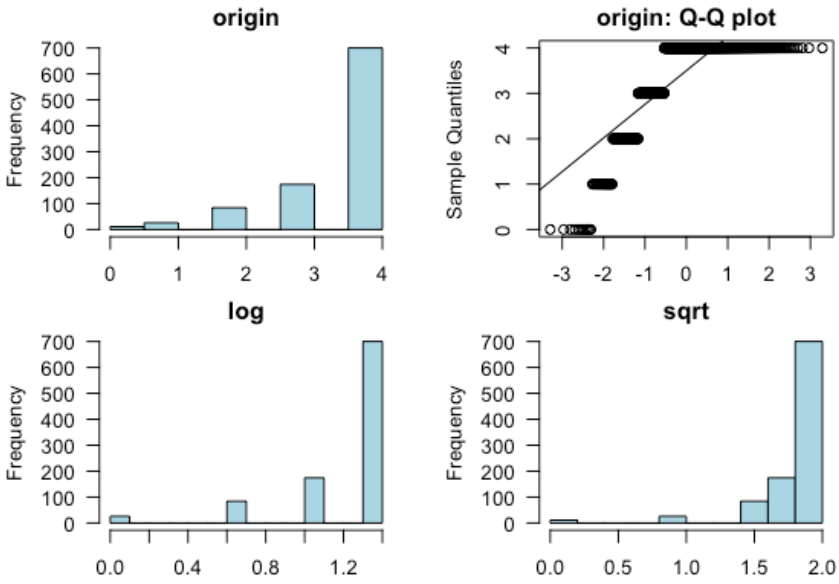
[ALSFRS2]

normality test : Shapiro-Wilk normality test

```
statistic : 0.61226,  p-value : 1.87682E-42
```

skewness and kurtosis

type	skewness	kurtosis
original	-1.971849	6.629249
log transformation	NaN	NaN
sqrt transformation	-3.368184	17.948477



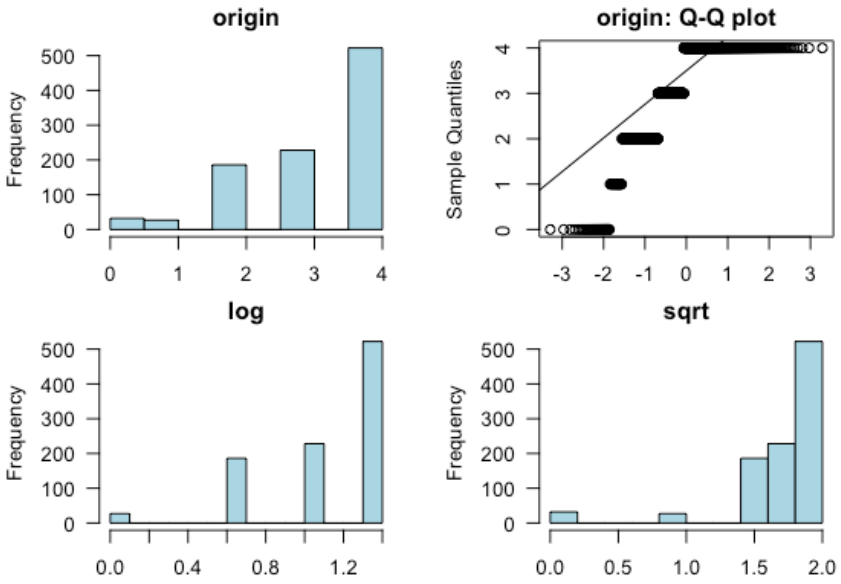
[ALSFRS3]

normality test : Shapiro-Wilk normality test

```
statistic : 0.75951,  p-value : 9.92163E-36
```

skewness and kurtosis

type	skewness	kurtosis
original	-1.218091	3.930507
log transformation	NaN	NaN
sqrt transformation	-2.573329	10.892644



[ALSFRS4]

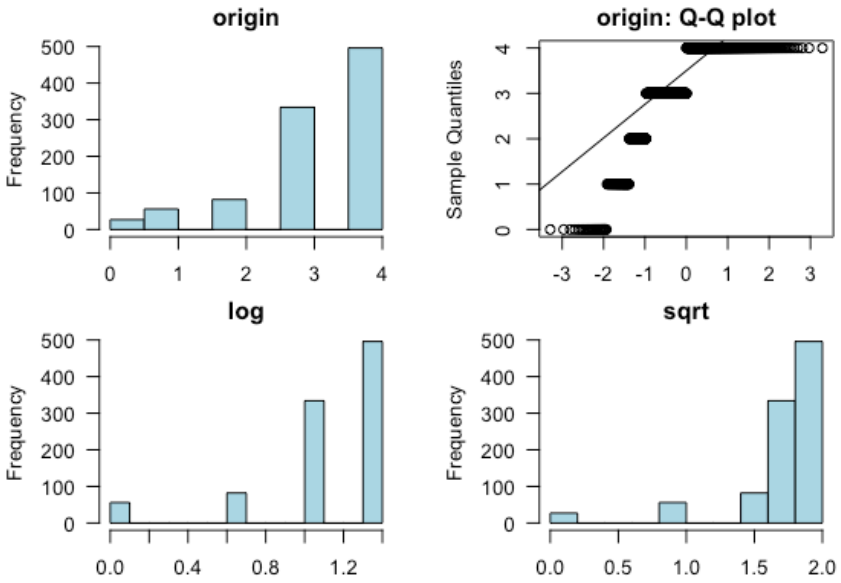
normality test : Shapiro-Wilk normality test

```
statistic : 0.74568,  p-value : 1.74049E-36
```

skewness and kurtosis

type	skewness	kurtosis
original	-1.442748	4.655703
log transformation	NaN	NaN

sqrt transformation	-2.696915	11.446089



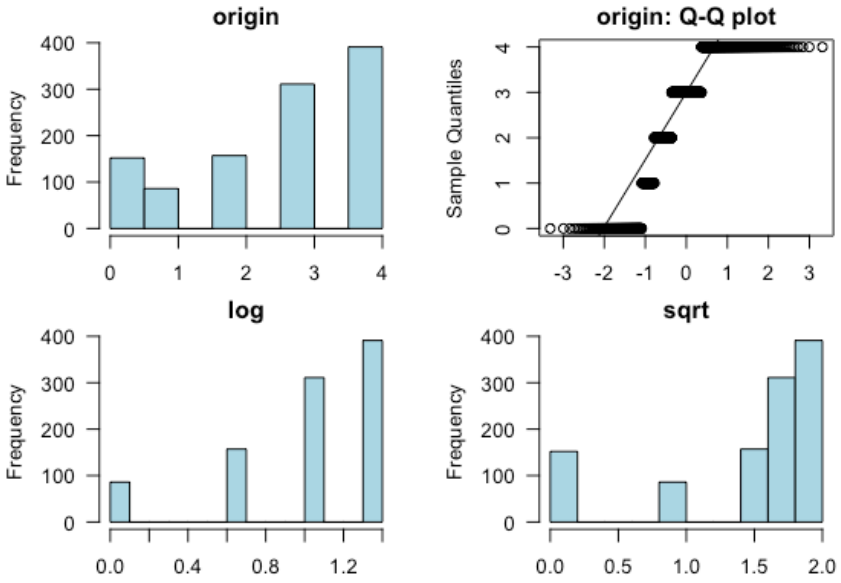
[ALSFRS5]

normality test : Shapiro-Wilk normality test

```
statistic : 0.82422,  p-value : 5.0552E-33
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.7556191	2.293288
log transformation	NaN	NaN
sqrt transformation	-1.4192800	3.692880



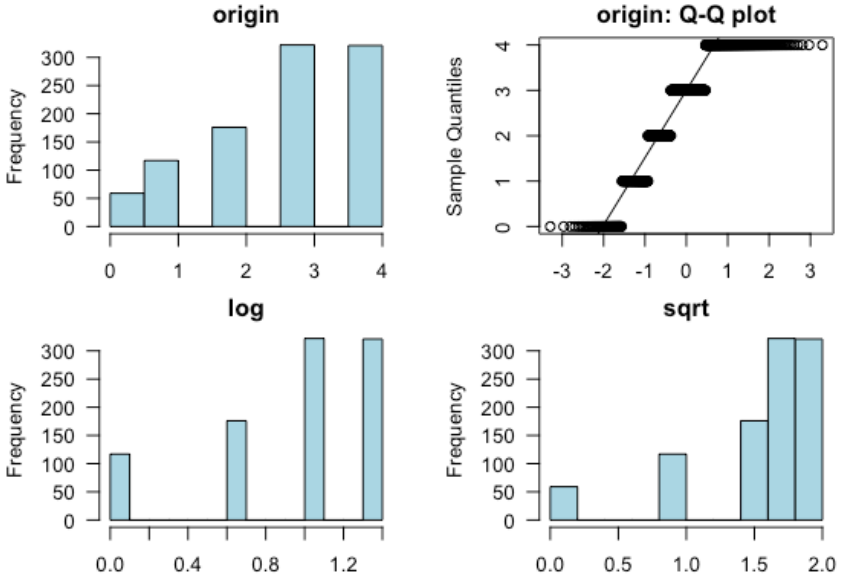
[ALSFRS6]

normality test : Shapiro-Wilk normality test

```
statistic : 0.85748,  p-value : 4.64907E-29
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.7167829	2.560603
log transformation	NaN	NaN
sqrt transformation	-1.7488419	5.871652



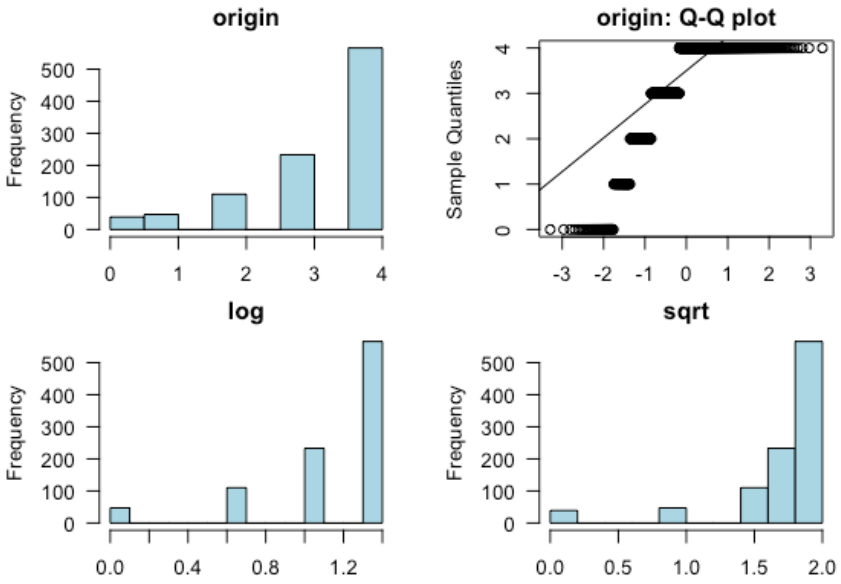
[ALSFRS7]

normality test : Shapiro-Wilk normality test

```
statistic : 0.71773,  p-value : 6.44697E-38
```

skewness and kurtosis

type	skewness	kurtosis
original	-1.479893	4.460061
log transformation	NaN	NaN
sqrt transformation	-2.601253	10.114175



[ALSFRS8]

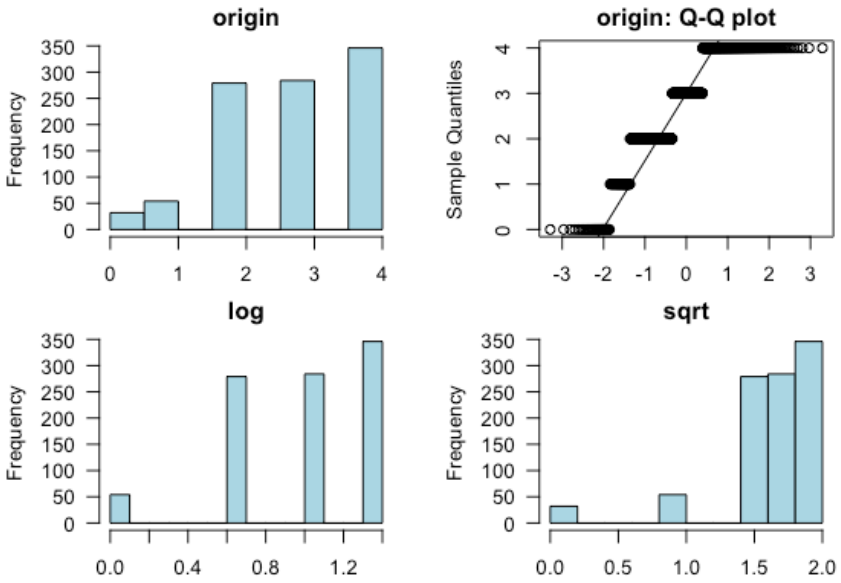
normality test : Shapiro-Wilk normality test

```
statistic : 0.85307,  p-value : 1.99148E-29
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.6548167	2.851189
log transformation	NaN	NaN

sqrt transformation	-2.0340161	8.555226



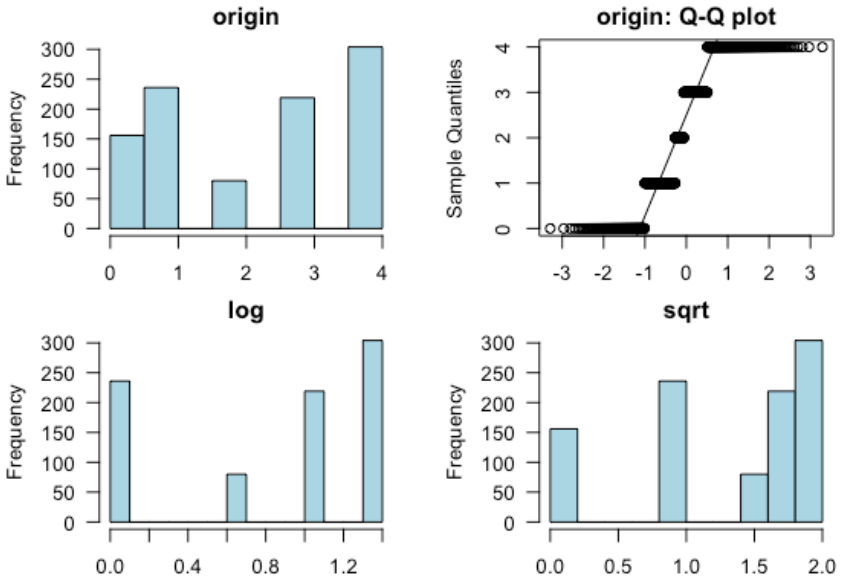
[ALSFRS9]

normality test : Shapiro-Wilk normality test

```
statistic : 0.84671,  p-value : 6.05301E-30
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.2175102	1.532836
log transformation	NaN	NaN
sqrt transformation	-0.8821221	2.536863



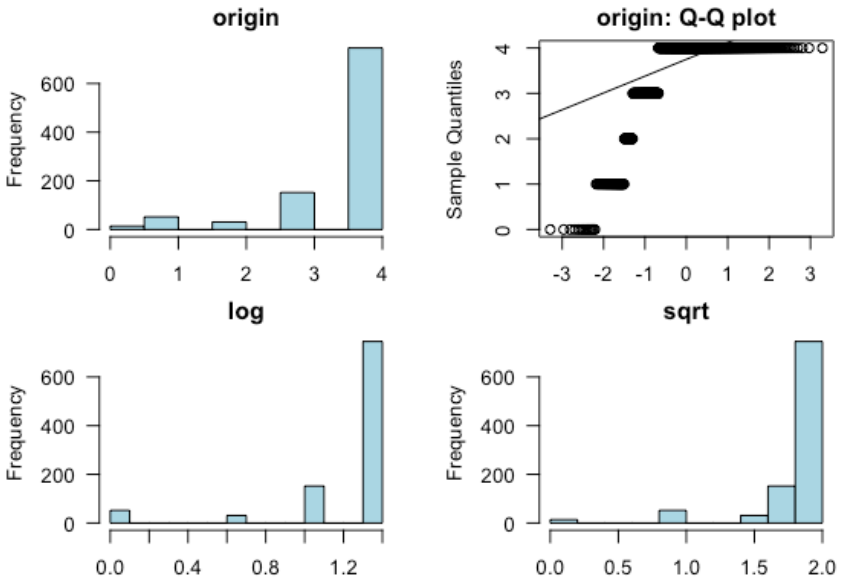
[ALSFRS10]

normality test : Shapiro-Wilk normality test

```
statistic : 0.54302,  p-value : 6.62524E-45
```

skewness and kurtosis

type	skewness	kurtosis
original	-2.310440	7.713219
log transformation	NaN	NaN
sqrt transformation	-3.391445	16.214186



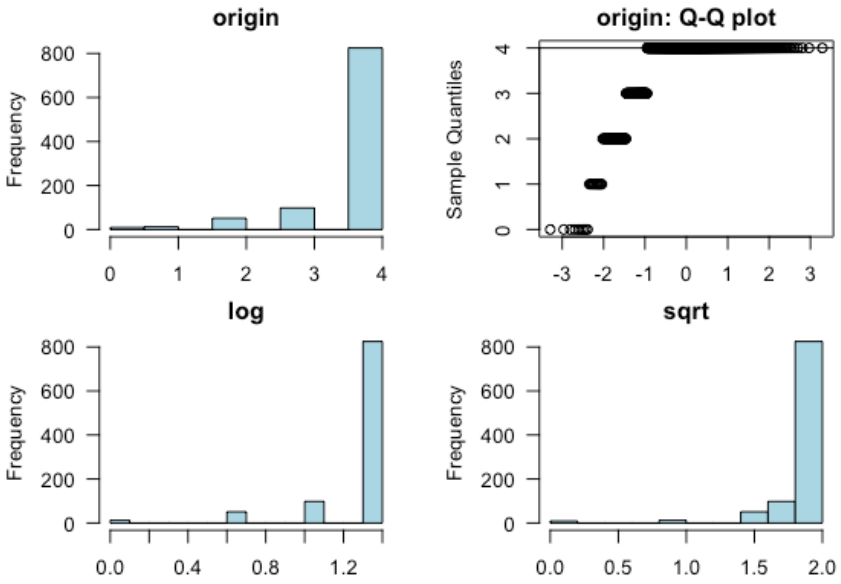
[ALSFRS11]

normality test : Shapiro-Wilk normality test

```
statistic : 0.45176,  p-value : 1.04845E-47
```

skewness and kurtosis

type	skewness	kurtosis
original	-2.979773	12.47462
log transformation	NaN	NaN
sqrt transformation	-4.721782	31.08390



[ALSFRS12]

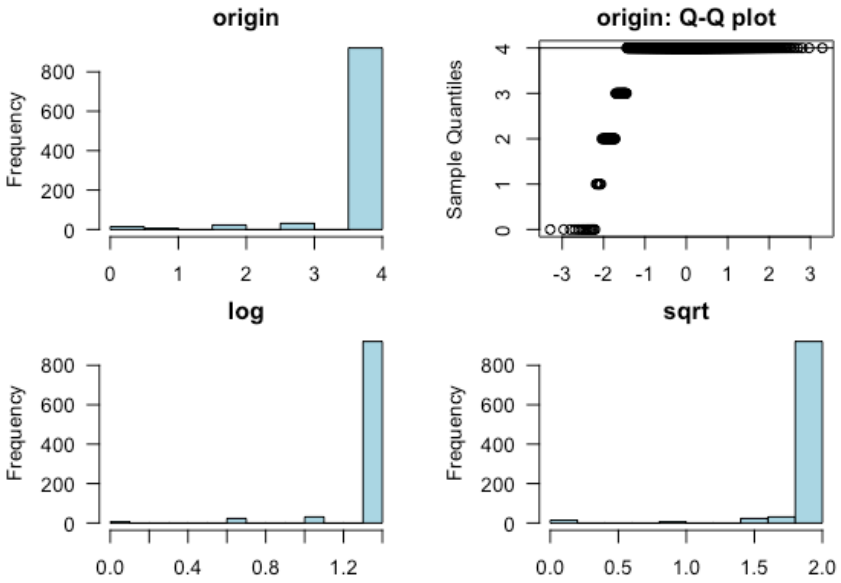
normality test : Shapiro-Wilk normality test

```
statistic : 0.26417,  p-value : 2.03035E-52
```

skewness and kurtosis

type	skewness	kurtosis
original	-4.708571	26.10269
log transformation	NaN	NaN

sqrt transformation	-6.134356	43.18172



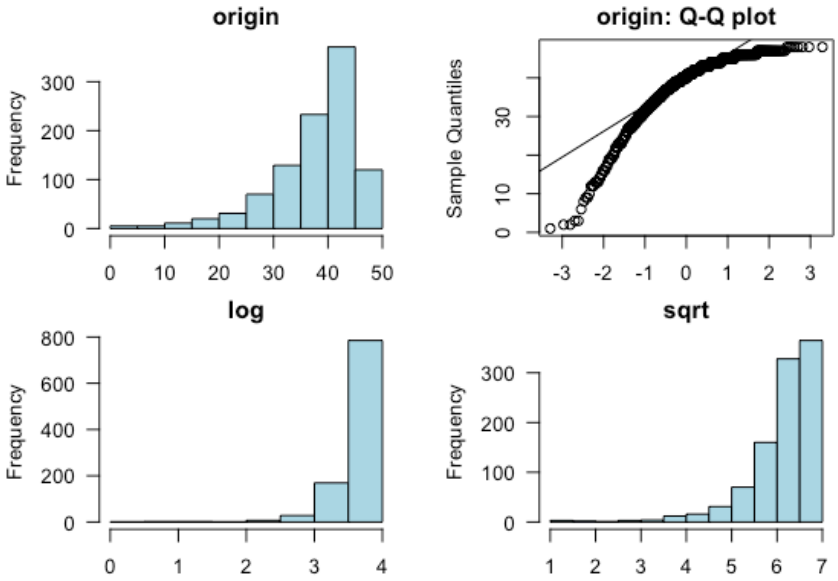
[First ALSFRS-R total]

normality test : Shapiro-Wilk normality test

```
statistic : 0.85717,  p-value : 4.38196E-29
```

skewness and kurtosis

type	skewness	kurtosis
original	-1.618671	6.199971
log transformation	-4.860760	39.513996
sqrt transformation	-2.539912	12.644283



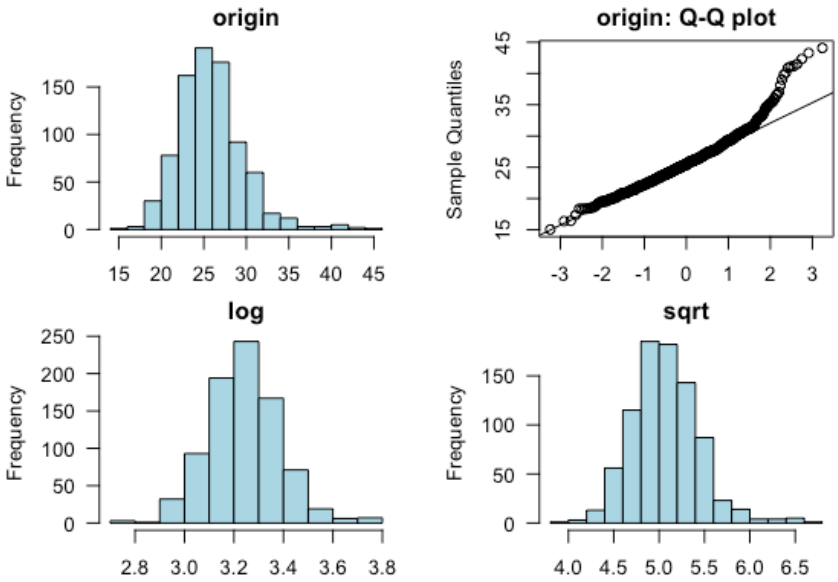
[BMI 2 years prior to illness]

normality test : Shapiro-Wilk normality test

```
statistic : 0.95483,  p-value : 2.38269E-15
```

skewness and kurtosis

type	skewness	kurtosis
original	0.9609847	5.468668
log transformation	0.2925616	3.984020
sqrt transformation	0.6151608	4.520224



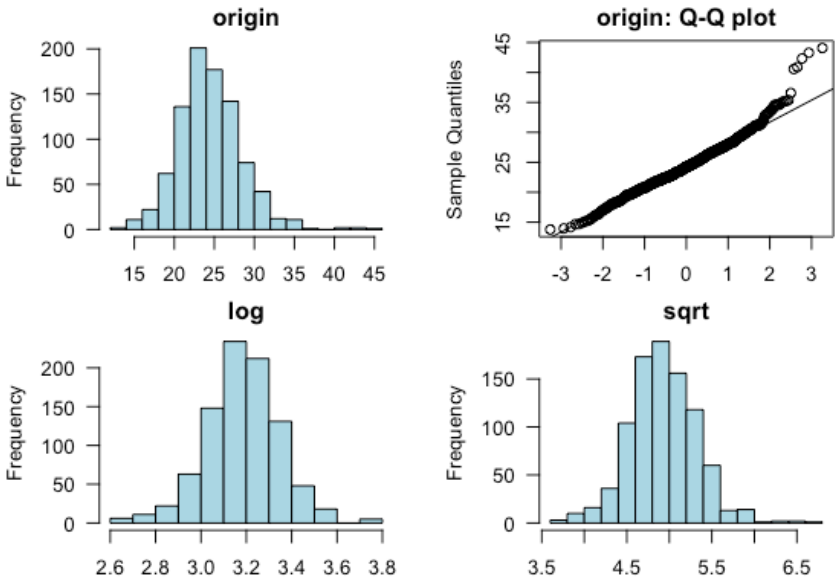
[BMI at diagnosis]

normality test : Shapiro-Wilk normality test

```
statistic : 0.97579,  p-value : 4.61281E-11
```

skewness and kurtosis

type	skewness	kurtosis
original	0.6269147	4.892236
log transformation	-0.1049349	3.899200
sqrt transformation	0.2503565	4.127763



[Delay in diagnosis (days)]

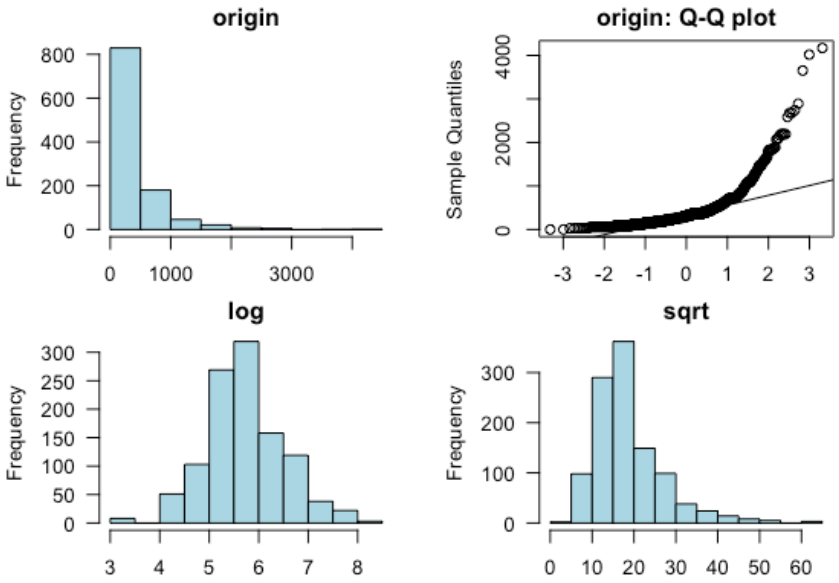
normality test : Shapiro-Wilk normality test

```
statistic : 0.67213,  p-value : 1.54091E-41
```

skewness and kurtosis

type	skewness	kurtosis
original	3.455771	20.711731
log transformation	NaN	NaN

sqrt transformation	1.540590	6.823532



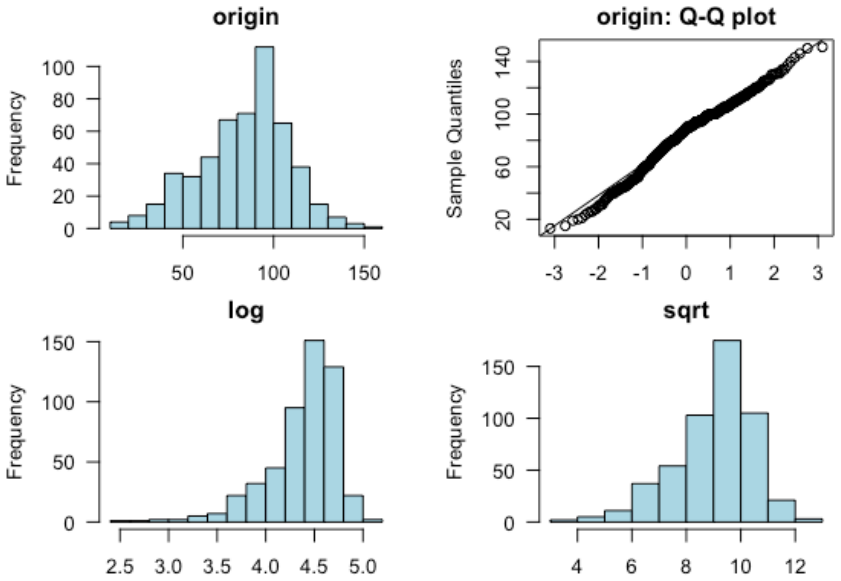
[FVC percent at diagnosis]

normality test : Shapiro-Wilk normality test

```
statistic : 0.98397,  p-value : 1.85344E-05
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.3649251	2.899146
log transformation	-1.4987068	6.072695
sqrt transformation	-0.8540960	3.715406



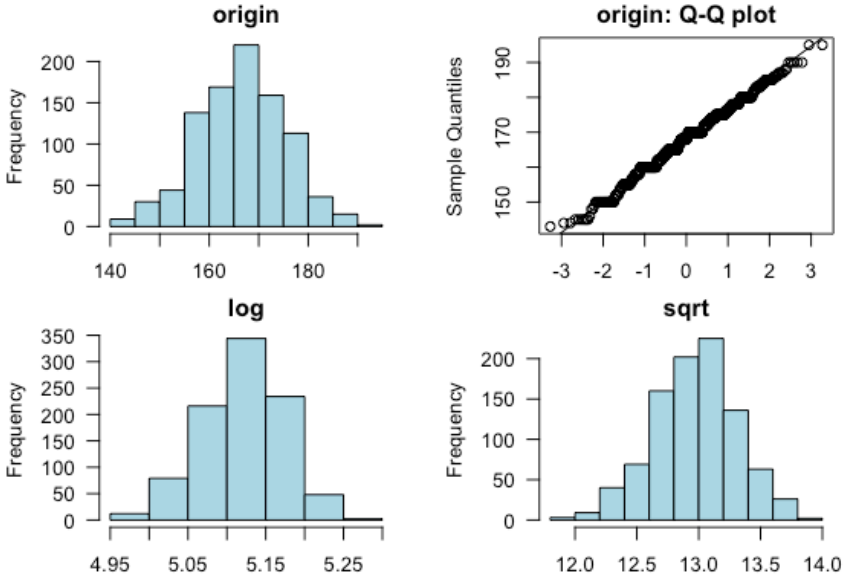
[Height]

normality test : Shapiro-Wilk normality test

```
statistic : 0.99259,  p-value : 0.000127709
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.0975817	2.933227
log transformation	-0.2490665	2.993306
sqrt transformation	-0.1732104	2.954348



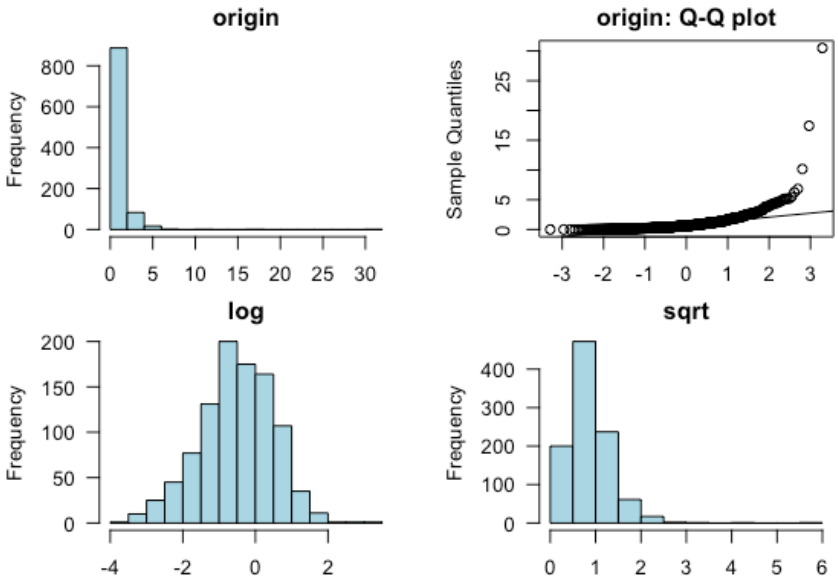
[Rate of decline ALSFRS-R (per month)]

normality test : Shapiro-Wilk normality test

```
statistic : 0.45084,  p-value : 1.11835E-47
```

skewness and kurtosis

type	skewness	kurtosis
original	10.855001	194.98672
log transformation	NaN	NaN
sqrt transformation	2.117055	15.95258



[Rate of decline BMI (per month)]

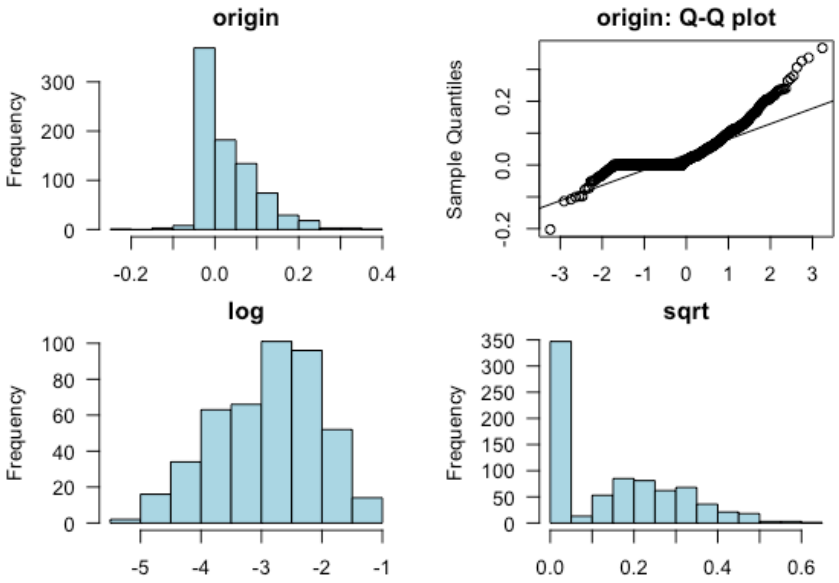
normality test : Shapiro-Wilk normality test

```
statistic : 0.81187,  p-value : 6.27561E-30
```

skewness and kurtosis

type	skewness	kurtosis
original	1.5277608	6.574505
log transformation	NaN	NaN

sqrt transformation	0.5701026	2.190665



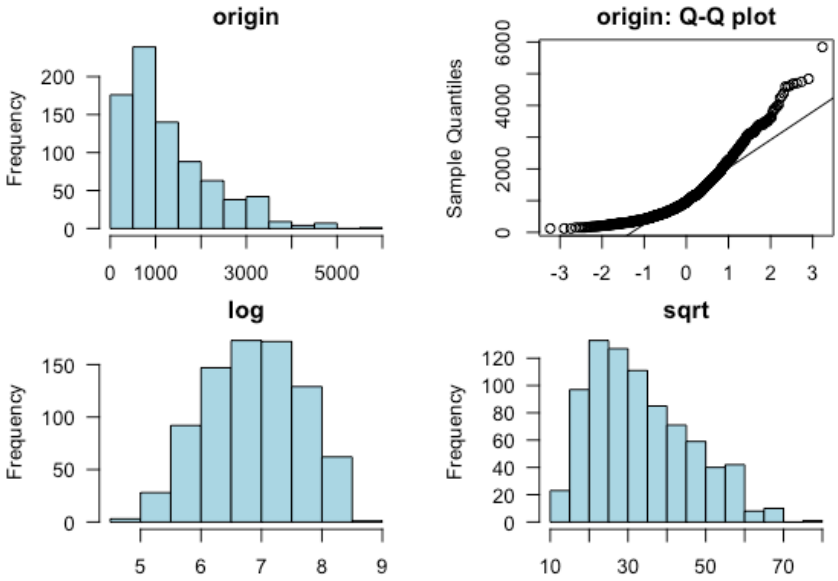
[Survival (days)]

normality test : Shapiro-Wilk normality test

```
statistic : 0.87498,  p-value : 6.47691E-25
```

skewness and kurtosis

type	skewness	kurtosis
original	1.3023905	4.483511
log transformation	-0.1239991	2.335726
sqrt transformation	0.6027388	2.680037



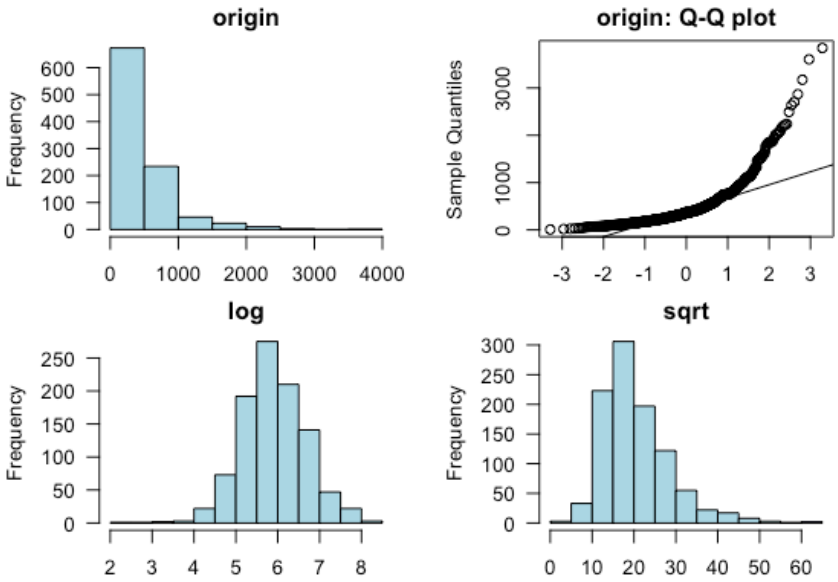
[Time of first ALSFRS-R (days into illness)]

normality test : Shapiro-Wilk normality test

```
statistic : 0.73542,  p-value : 5.58944E-37
```

skewness and kurtosis

type	skewness	kurtosis
original	2.8523601	15.042545
log transformation	-0.0975394	3.872581
sqrt transformation	1.3013225	5.624284



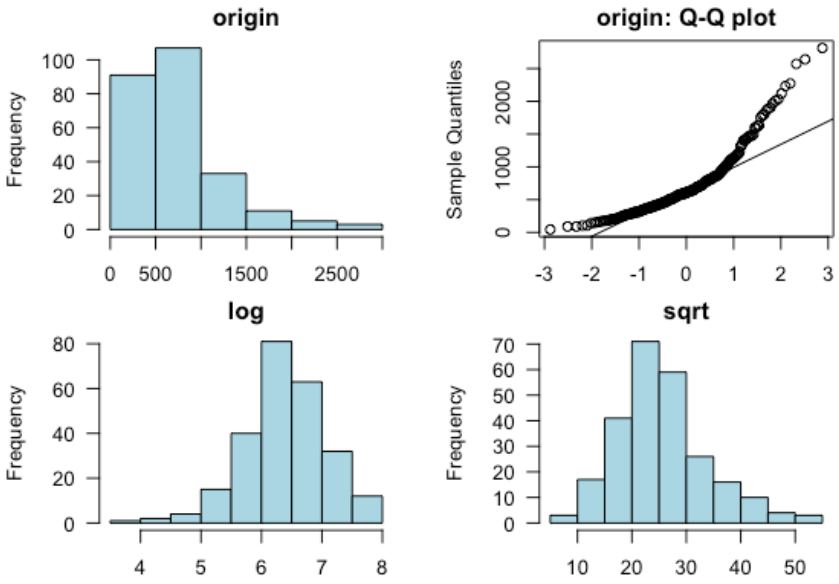
[Time of NIPPV (days into illness)]

normality test : Shapiro-Wilk normality test

```
statistic : 0.8642,  p-value : 4.51671E-14
```

skewness and kurtosis

type	skewness	kurtosis
original	1.5806302	5.807980
log transformation	-0.3938839	3.734917
sqrt transformation	0.6964814	3.525247



[Time of PEG (days into illness)]

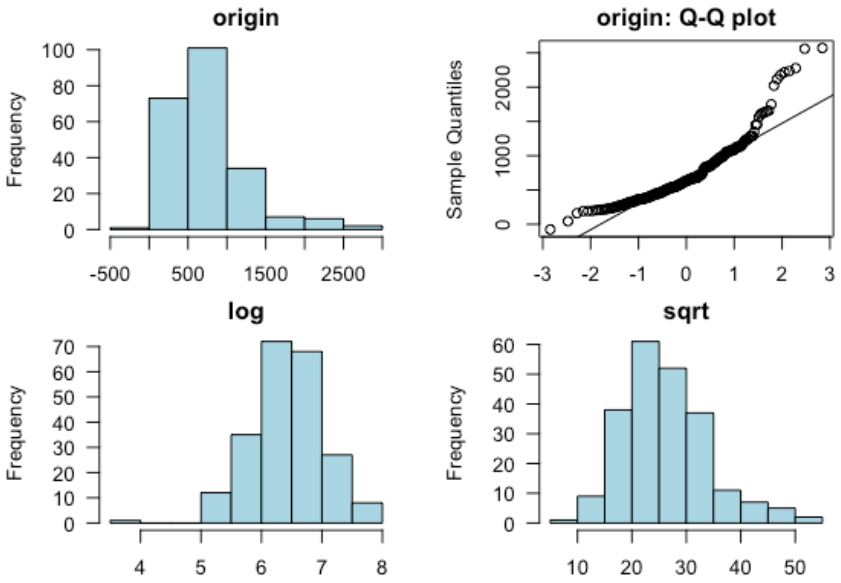
normality test : Shapiro-Wilk normality test

```
statistic : 0.87414,  p-value : 1.14951E-12
```

skewness and kurtosis

type	skewness	kurtosis
original	1.5716974	6.124369
log transformation	-0.3066038	4.086179

sqrt transformation	0.7342039	3.718329



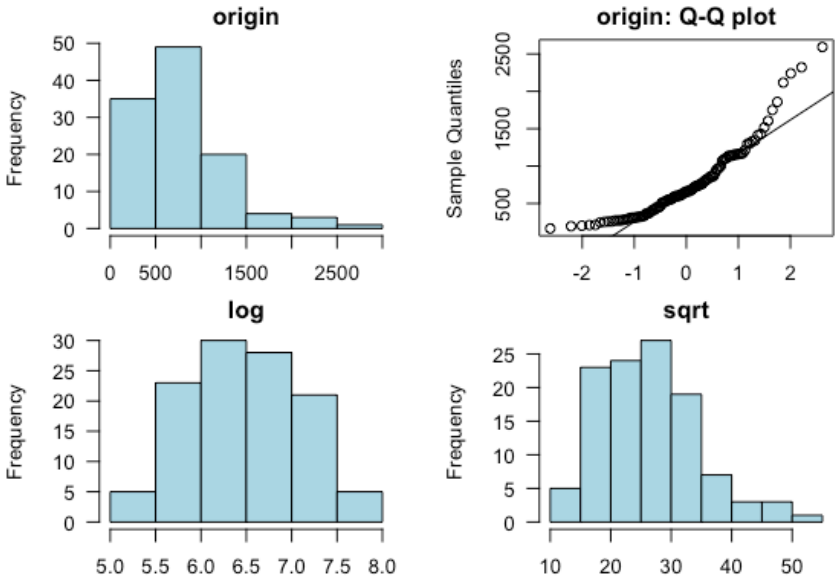
[Time of tracheostomy (days into illness)]

normality test : Shapiro-Wilk normality test

```
statistic : 0.88057,  p-value : 5.18023E-08
```

skewness and kurtosis

type	skewness	kurtosis
original	1.4169197	5.313599
log transformation	-0.0258596	2.386614
sqrt transformation	0.6532936	3.170014



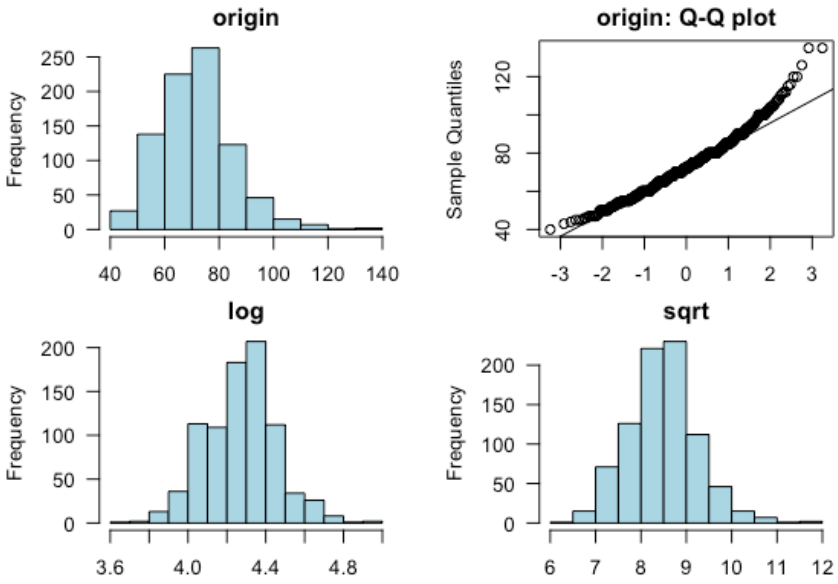
[Weight 2 years prior to illness]

normality test : Shapiro-Wilk normality test

```
statistic : 0.97546,  p-value : 9.70346E-11
```

skewness and kurtosis

type	skewness	kurtosis
original	0.6582430	4.267136
log transformation	0.0025790	3.242574
sqrt transformation	0.3177279	3.551674



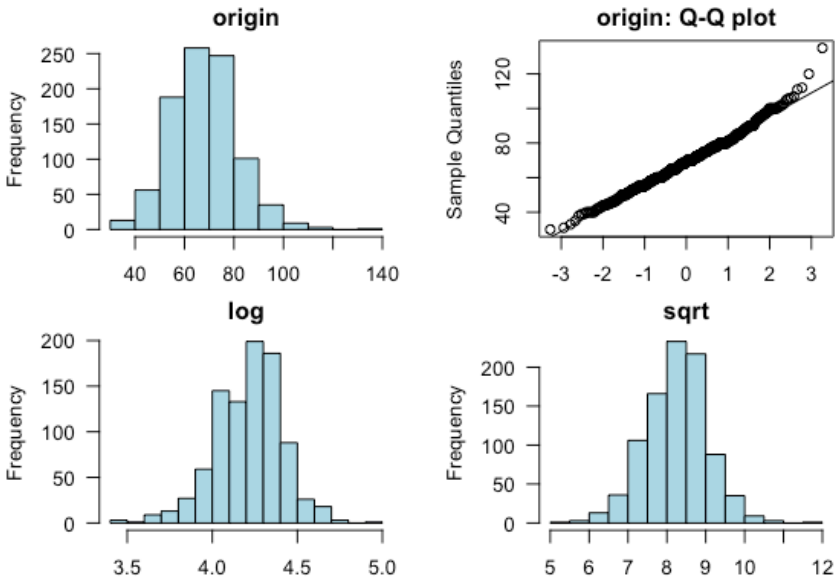
[Weight at diagnosis]

normality test : Shapiro-Wilk normality test

```
statistic : 0.99126,  p-value : 3.16184E-05
```

skewness and kurtosis

type	skewness	kurtosis
original	0.3241353	3.734421
log transformation	-0.4038662	3.679122
sqrt transformation	-0.0377384	3.441667



3 Relationship Between Variables

3.1 Correlation Coefficient

3.1.1 Correlation Coefficient by Variable Combination

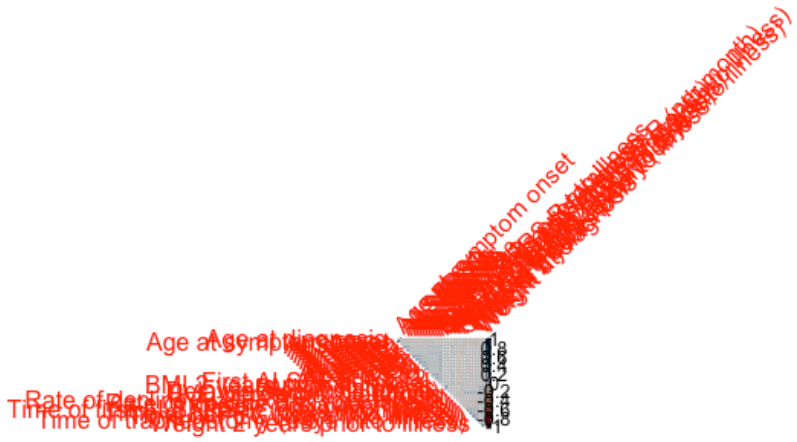
Table of correlation coefficients (0.5 or more)

Variable1	Variable2	Correlation Coefficient
Time of tracheostomy (days into illness)	Survival (days)	1.0000000
Age at symptom onset	Age at diagnosis	0.9945026
Time of first ALSFRS-R (days into illness)	Delay in diagnosis (days)	0.9208946

Time of PEG (days into illness)	Time of NIPPV (days into illness)	0.9051451
Weight at diagnosis	Weight 2 years prior to illness	0.8948297
ALSFRS9	ALSFRS8	0.8640204
Time of tracheostomy (days into illness)	Time of PEG (days into illness)	0.8461230
BMI at diagnosis	BMI 2 years prior to illness	0.8419137
Weight at diagnosis	BMI at diagnosis	0.8395588
Time of NIPPV (days into illness)	Survival (days)	0.8236884
Time of tracheostomy (days into illness)	Time of NIPPV (days into illness)	0.8236884
Weight 2 years prior to illness	BMI 2 years prior to illness	0.8168989
First ALSFRS-R total	ALSFRS7	0.8057479
First ALSFRS-R total	ALSFRS6	0.7971927
ALSFRS7	ALSFRS6	0.7924763
ALSFRS5	ALSFRS4	0.7899101
ALSFRS12	ALSFRS10	0.7749520
Time of NIPPV (days into illness)	Time of first ALSFRS-R (days into illness)	0.7506319
Time of NIPPV (days into illness)	Delay in diagnosis (days)	0.7492499
ALSFRS3	ALSFRS1	0.7436443
ALSFRS6	ALSFRS5	0.7314105
Weight at diagnosis	BMI 2 years prior to illness	0.7048297
First ALSFRS-R total	ALSFRS8	0.7008749
First ALSFRS-R total	ALSFRS5	0.6995100
ALSFRS11	ALSFRS10	0.6981255
Weight 2 years prior to illness	BMI at diagnosis	0.6916708
First ALSFRS-R total	ALSFRS9	0.6851391
First ALSFRS-R total	ALSFRS4	0.6791736
Time of PEG (days into illness)	Time of first ALSFRS-R (days into illness)	0.6779765
ALSFRS8	ALSFRS7	0.6639665
Time of PEG (days into illness)	Delay in diagnosis (days)	0.6629565
ALSFRS7	ALSFRS5	0.6507141
ALSFRS12	ALSFRS11	0.6451175
ALSFRS2	ALSFRS1	0.6421880
ALSFRS6	ALSFRS4	0.6388955
Time of PEG (days into illness)	Survival (days)	0.6266256

ALSFRS3	ALSFRS2	0.6257266
ALSFRS7	ALSFRS4	0.6169414
ALSFRS8	ALSFRS6	0.6149098
First ALSFRS-R total	ALSFRS11	0.6137519
ALSFRS9	ALSFRS6	0.6136345
ALSFRS9	ALSFRS7	0.6108561
Weight 2 years prior to illness	Height	0.5931270
First ALSFRS-R total	ALSFRS10	0.5921760
First ALSFRS-R total	ALSFRS3	0.5911316
Survival (days)	Delay in diagnosis (days)	0.5748130
Time of first ALSFRS-R (days into illness)	Survival (days)	0.5744507
Weight at diagnosis	Height	0.5628746
Time of tracheostomy (days into illness)	Rate of decline ALSFRS-R (per month)	-0.5310273
First ALSFRS-R total	ALSFRS12	0.5236304
First ALSFRS-R total	ALSFRS1	0.5006015

3.1.2 Correlation Plot of Numerical Variables



4 Target based Analysis

4.1 Grouped Descriptive Statistics

4.1.1 Grouped Numerical Variables

4.1.2 Grouped Categorical Variables

4.2 Grouped Relationship Between Variables

4.2.1 Grouped Correlation Coefficient

4.2.2 Grouped Correlation Plot of Numerical Variables