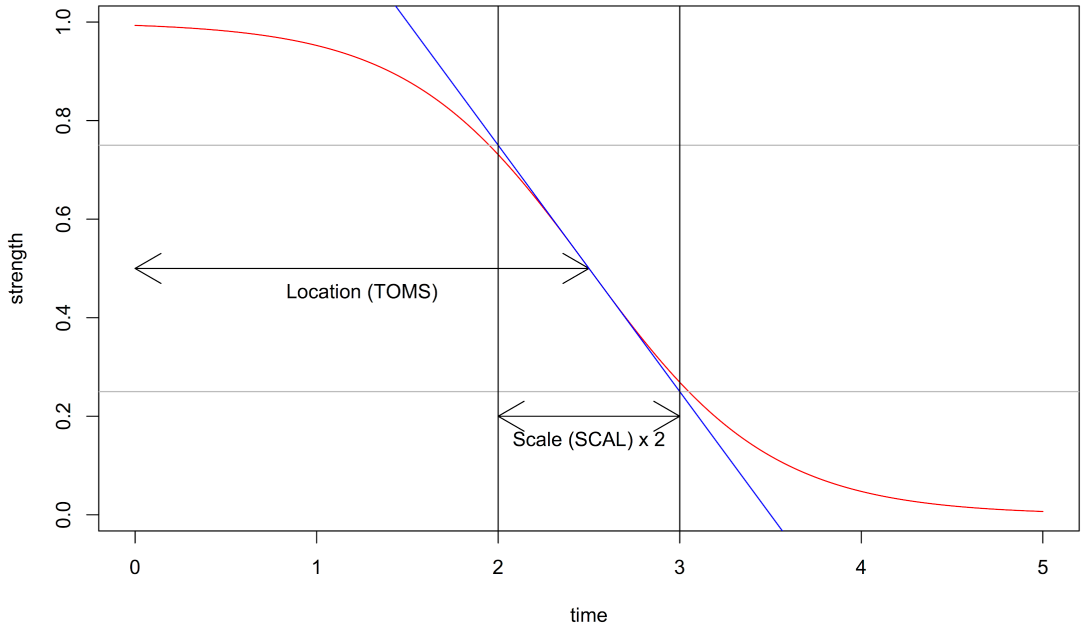
SUPPLEMENTARY FILE

Modeling framework for comparing progressive loss of strength in different muscles

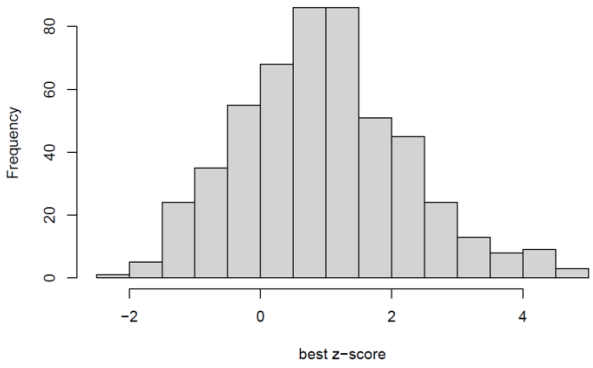
Rationale for the logistic function: The logistic function asymptotic to 1 on the left (normal strength) and 0 on the right (zero strength) has an ogive shape, with a slow initial slope, a rapid intermediate slope, and a slow final slope. It is a fair approximation of trajectories of muscle strength, FVC [1] and functional decline [2] in ALS. This form of the logistic function has two free parameters that can be estimated, namely: (a) the location parameter (that we call TOMS, or *time from onset to midway strength*, the time point where the function crosses the horizontal line at 0.5 strength) and (b) the scale parameter (that we call SCAL, which equals half the time it takes to lose strength from 0.75 strength to 0.25 strength at the maximum rate of decline) (See Supplementary File Figure 1). Given that the primary goal of this study is to find the order of weakness of muscles, the logistic model is mathematically convenient because it permits ready estimation of TOMS.



Supplementary File Figure 1: Schematic of location (TOMS) and scale (SCAL) parameters in a logistic model

Heuristic approximation of (unobserved) normal strength (zero-one scaling): One difficulty with this modeling approach is missing premorbid normal strength data, because the patient is already weak in some or many muscles when first assessed. This normal value is required to assign HHD measures (in force units) a proportional value between 0 (total paralysis) and 1 (normal strength). Normative mean values [3] are an option, but many HHD measures exceed them. To solve this difficulty, we identified the best (highest) z-score $z^{(0)}$ ever for each subject for any muscle (arm or leg) (Supplementary File Figure 2). If $z^{(0)}$ exceeded zero, the assumed normal for each muscle was scaled up from the normative mean of that muscle by that z value ($z^{(0)}$ *SD, where SD is the normative SD for that muscle). If $z^{(0)}$ was between 0 and -2, the assumed normal for each muscle was scaled down from the normative mean of that muscle by half the value (0.5* $z^{(0)}$ *SD). If the best z-score was less than -2, assumed normals were set at the normative means minus the normative SDs. A floor was assigned to address the possibility that all

muscles were weak already at initial evaluation. This approach of pegging unobserved baseline strength of weakened muscles to the strength of the best-preserved (and presumably intact) muscle is supported by significant correlation of strength across different muscles in normal subjects [4,5]. Meaningful variations of this empirical approach (such as setting the floor for normal at the normative mean or some specified distance below the mean in normative SD units, or choosing as the scaling $z^{(0)}$ a value informed by the two best-preserved muscles invoking the conditional probability density under a multivariate normal distribution) in our sensitivity analyses yielded qualitatively similar results. We call this method to adjust for between-subject variation in normal (baseline) strength and standardise strength across muscles and across different subjects 'zero-one scaling' (ZOS).



Supplementary File Figure 2: Histogram of best z-scores. Frequency of values < 0, < -1 and < -2 was 120, 30 and 1 respectively.

Listed below are muscles that contributed the best z-score

RSF	LSF	REF	LEF	REE	LEE	RWE	LWE	RDI	LDI	RHF	LHF	RKE	LKE	RKF	LKF	RAE	LAE
9	14	19	18	45	24	15	19	18	15	9	13	100	115	22	21	15	22

Tethering of modeled trajectory to normal strength before symptom onset: This was accomplished by adding (unobserved but assumed) data points at onset (time 0) and 12 weeks prior to onset that were near normal strength (empirically 0.98 strength) and normal strength (1 strength) respectively. Minor variations in these empirical choices did not significantly alter estimates.

Parameterisation of the logistic function: A typical parameterisation of the 2-parameter logistic function (asymptotic to 1 on the left and 0 on the right) is as follows:

$$\frac{1}{1+exp(\frac{t-TOMS}{SCAL})} \qquad \text{--------(1)}$$

Where *t* is time, *TOMS* is the location parameter and *SCAL* is the scale parameter.

So that we could model proportional differences in TOMS, and also to avoid the possibility of estimation of negative values for TOMS and SCAL, we altered parameterisation so that logarithmic transformations of TOMS and SCAL were estimated instead:

$$\frac{1}{1+exp(\frac{t-exp(a+logTOMS)}{exp(b+c*logSCAL)})} \qquad \text{--------(2)}$$

In this formulation, a, b and c are constants arbitrarily chosen to improve the success of the optimiser to converge at the maximum likelihood estimate in models with a high-dimensional parameter space. Arbitrary values used for a, b and c were 4.2, 3.1 and 5 respectively. It should be emphasised that

reparameterisation by different choices of a, b and c does not alter modeled trajectories or TOMS estimates.

Non-linear mixed models: Non-linear mixed models employing the described logistic function were constructed using the nlme package in R [6]. Mixed models permit random variability of parameters between subjects (random effects) in addition to fixed effects related to covariates. Our models assigned subject-level random effects to logTOMS as well as logSCAL, thereby permitting variability in the "underlying" TOMS and SCAL between subjects in a lognormal distribution. Our models also assigned fixed effects to logTOMS and logSCAL for each muscle group, thereby permitting a muscle group to have a TOMS that was some proportion of the underlying TOMS, or some proportion of the TOMS of another muscle group. The base model (Model 1, Supplementary File Table 1) therefore had 24 parameters: 10 fixed effects (for each muscle group) assigned to logTOMS, 10 fixed effects (also for each muscle group) assigned to logSCAL, 1 random effects variance term each for logTOMS and logSCAL, 1 random effects covariance term, and 1 residual variance term. More complex models were constructed adding fixed effects terms for the side of the weaker arm and for site of onset. The model presented in the results that had the best fit to the data had a total of 90 fixed effects parameters (described in the main text and also as Model 4, Supplementary File Table 1).

Model selection: Because success of convergence of the optimiser was sensitive to choice of starting values of parameters, simpler non-linear mixed models were initially constructed with only one random effect (logTOMS) to find reasonable starting values of fixed parameters for more complex models where a second random effect (logSCAL) was introduced. Models with single random effects had poor fits. Table 1 below reports Akaike information criteria (AIC) for 4 fitted models with random effects for both logTOMS and logSCAL. One limitation of our analysis is that a second layer of variability (between muscles in one subject) was not modeled for. Such hierarchical mixed models were not identified because of limitations of data.

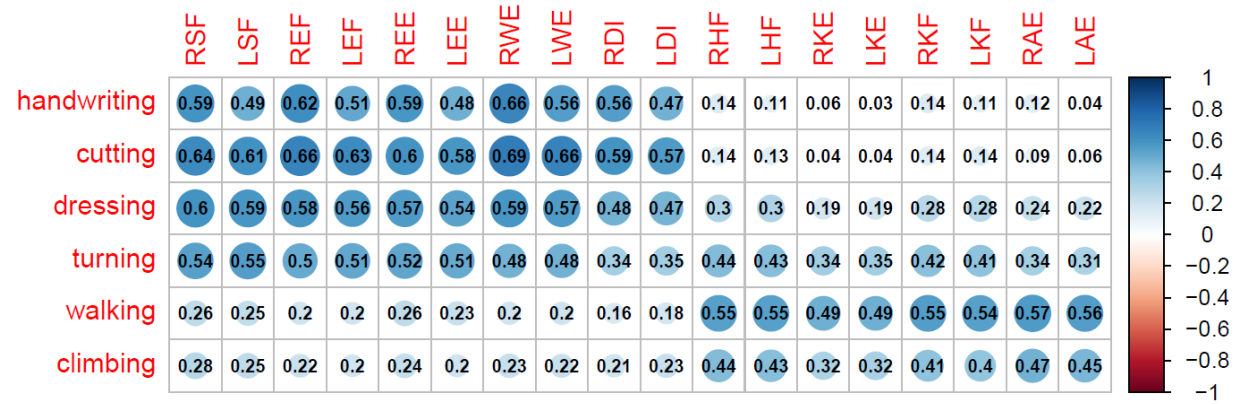
Contrasts: Within this estimation framework, quantities of interest, such as TOMS ratios between muscle groups, and ratios of TOMS ratios between muscle groups by site of onset, are essentially exponents of linear contrasts of estimated parameters (various logTOMS). When a large number contrasts are examined, there is "false discovery" of effects at the Type I error rate (α, usually 0.05). Scheffé's method for multiple comparisons [7] adjusts simultaneously for infinitely many possible linear contrasts of estimated parameters to reduce any false positivity (among all contrasts) to the Type I error rate.

Supplementary File Table 1:

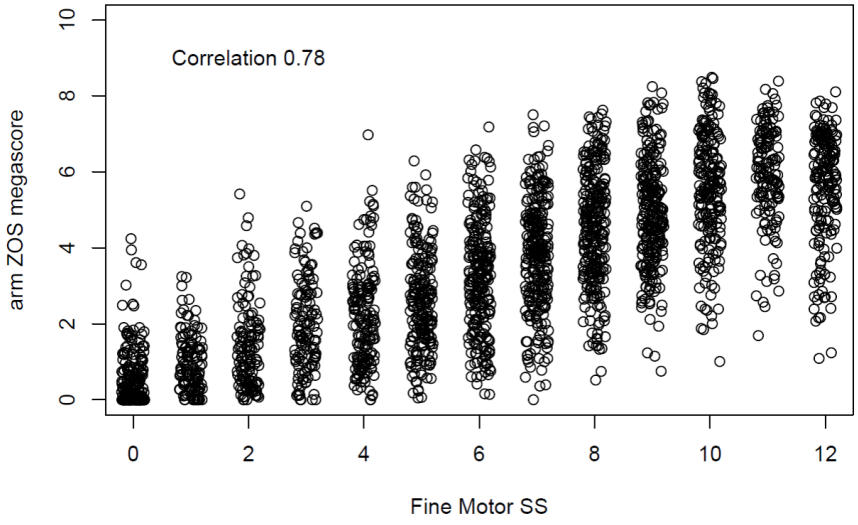
Model	Fixed effects parameters	Variance components (parameters)	df	AIC
1 (base)	logTOMS: muscle group (10) logSCAL: muscle group (10)	logTOMS, logSCAL, covariance, residual	24	-40052
2	logTOMS: muscle group (10) x onset_site (3) logSCAL: muscle group (10) x onset_site (3)	logTOMS, logSCAL, covariance, residual	64	-40344
3	logTOMS: muscle group (10) x weaker_arm (2) logSCAL: muscle group (10)	logTOMS, logSCAL, covariance, residual	34	-42955
4 (final)	logTOMS: muscle group (10) x onset_site (3) x weaker_arm (2) logSCAL: muscle group (10) x onset_site (3)	logTOMS, logSCAL, covariance, residual	94	-43538

Associations of muscle strength and TOMS with ALSFRS-R responses

As would be expected, across all measures, strength was correlated with ALSFRS-R motor responses. Arm muscle strength correlated with fine motor responses (ALSFRS-R questions 4-6, relating to handwriting, cutting food, and dressing) as well as ALSFRS-R question 7 relating to turning in bed. Leg muscle strength correlated with gross motor responses (especially ALSFRS-R questions 8 and 9 relating to walking and climbing, less with question 7). A color-coded correlation matrix of ZOS HHD strength and ALSFRS-R responses is presented in Supplementary File Figure 3. Summated arm muscle ZOS strength ("arm ZOS-megascore") was closely correlated with the fine motor subscore (r = 0.78, Supplementary File Figure 4). All correlations were adjusted for the grouped structure of the data (within-subject correlation).

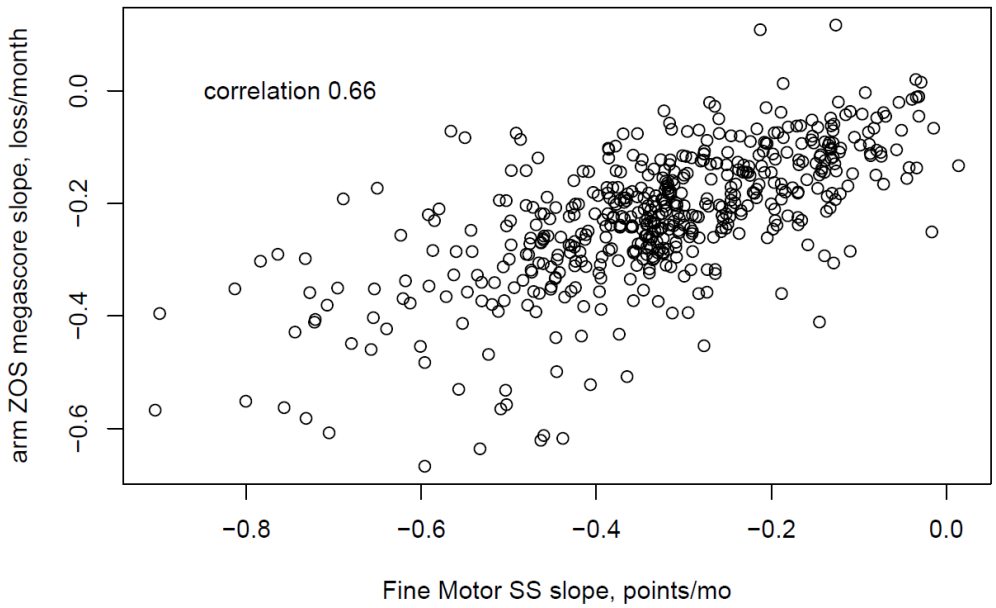


Supplementary File Figure 3: Correlation coefficients of ZOS strength for individual muscles (SF=shoulder flexion, EF=elbow flexion, EE=elbow extension, WE=wrist extension, DI=first dorsal interosseous, HF=hip flexion, KE=knee extension, KF=knee flexion, AE=ankle extension (dorsiflexion). R/L indicate sides. All achieved statistical significance at the threshold of $p < 0.05$ except for associations of LKE, RKE, and LAE with questions 4 and 5 (handwriting and cutting food).



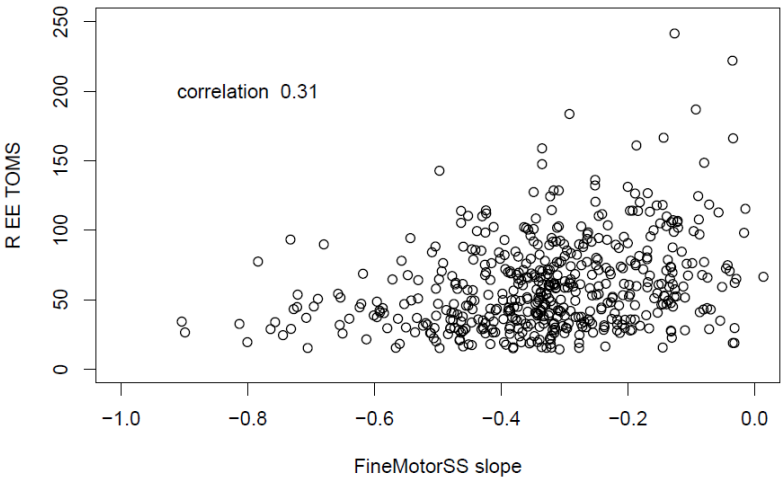
Supplementary File Figure 4: Correlation of the fine motor subscore of ALSFRS-R and arm ZOS megascore.

Slope of decline of arm ZOS megascore was also correlated with slope of decline of the fine motor subscore of ALSFRS-R (Supplementary File Figure 5).



Supplementary File Figure 5: Fine motor subscore slope – arm ZOS megascore slope scatter plot and correlation. Note that linear mixed effects models with random intercept and slope were employed to extract individual slopes (best linear unbiased predictions), rather than the logistic model described in the manuscript.

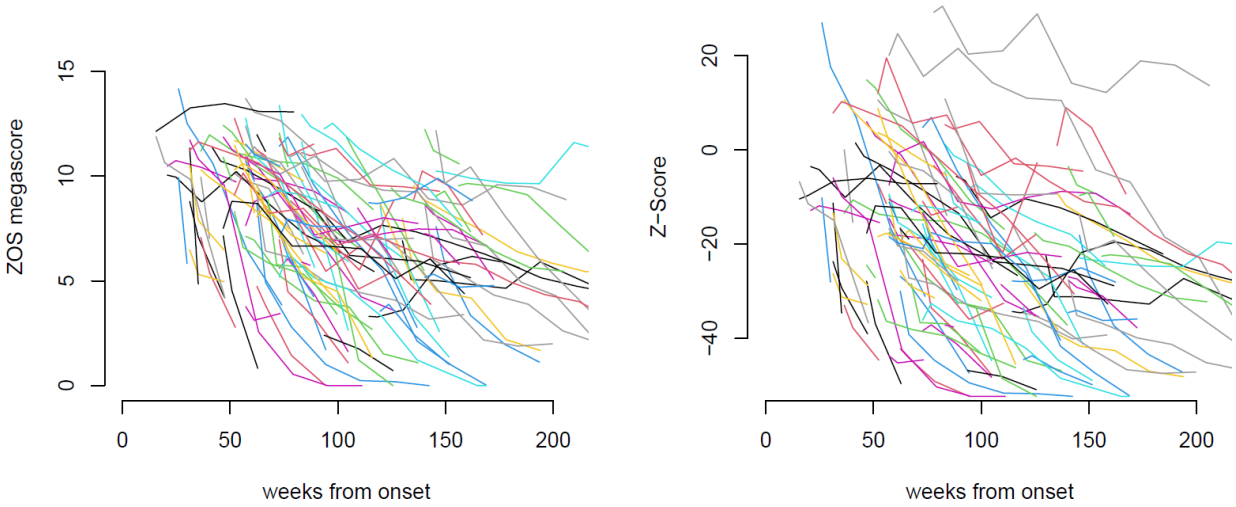
In contrast, the correlation between predicted fine motor subscore slopes and individual muscle TOMS estimates was weak (Supplementary File Figure 6)



Supplementary File Figure 6: Fine motor subscore slope and Right EE TOMS scatterplot and correlation.

Performance of the ZOS megascore relative to z-megascore and ALSFRS-R, examining linear decline using linear longitudinal mixed effects models

Inspection of conventional (normalised or z-scaled) megascore and ZOS megascores trajectory plots suggests reduced between-patient variability with the latter.



Supplementary File Figure 7: Spaghetti plots of ZOS megascore and z-megascore (total of 18 muscle groups, arms and legs) trajectories for 100 patients.

For sample size calculation and determining the optimal number of repeated measures required, the quantities of interest are (a) the ratio of the SD of slope to the value of clinically meaningful difference (often set at 0.25 of mean rate of decline) - therefore the coefficient of variation (CoV = SD/Mean), and (b) the ratio of the residual SD to the SD of slope respectively [8]. Low values of both quantities improve efficiency. They are displayed in rows 4 and 5 of Supplementary File Table 2. Although coefficient of variation of HHD measures remains higher than that of the ALSFRS-R total score, the ratio of residual SD to slope SD is lower for HHD measures. ZOS megascore improves efficiency relative to z-scaled megascore.

Supplementary File Table 2: Comparison of longitudinal outcome measures. Note that linear mixed effects models with random intercept and slope were employed to extract a fixed effect (mean slope) as well as variance components (variance of slope and residual variance). For megascores, arm as well as leg muscle groups were included (total of 18 muscle groups).

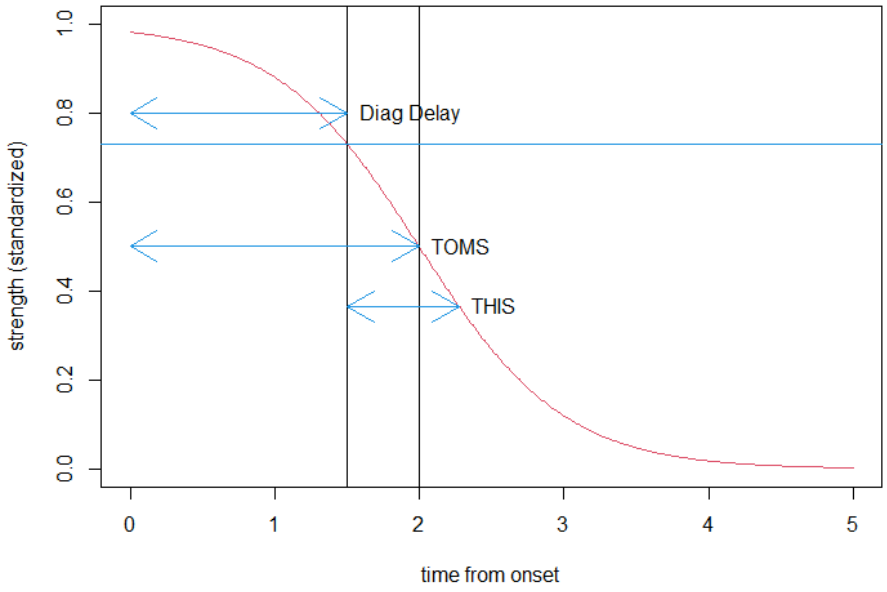
	ZOS-megascore, HHD (normalised 0-1)	z-megascore, HHD (standardised)	ALSFRS-R total score
Residual SD	0.900	3.767	2.808
Slope SD	0.252	0.986	0.666
Mean slope (per month)	-0.398	-1.511	-1.147
CoV (Slope SD/Mean slope)	0.633	0.653	0.580
ResidSD/SlopeSD	3.571	3.820	4.219

An alternative construct: Time to half (initial) strength (THIS)

One valid criticism of our presented analytic approach is the possibility of inaccurate prediction of unknown baseline strength, which can confound all estimates. An alternative method is to disregard baseline strength, and instead focus on the time it takes from initial visit to reach 50% of initially observed strength. We call this construct 'time to half (initial) strength' (THIS). Clearly, TOMS, diagnostic delay (DD) and TOMS are intimately related. If an unrealistic linear trajectory of strength decline from onset is presumed, the relationship is simply

$$THIS = TOMS - \frac{DD}{2} \qquad \text{----------(3)}$$

If a logistic trajectory of strength decline is presumed, as depicted in the schematic below,



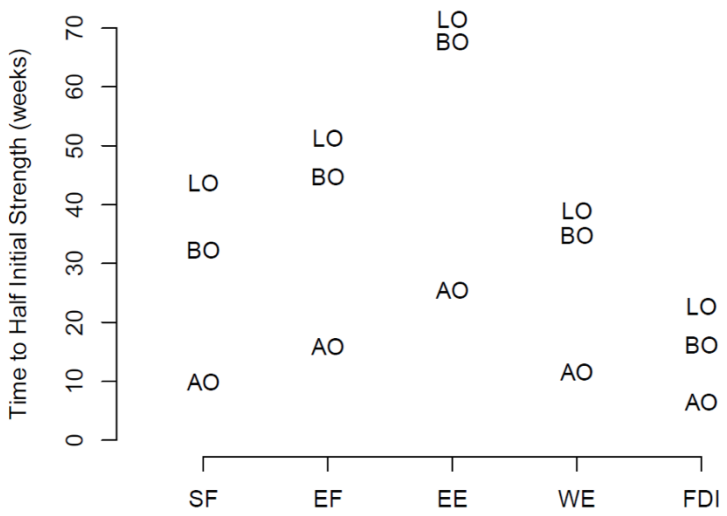
Supplementary File Figure 8: Relationship of time from onset to midway strength (TOMS), time to half initial strength (THIS) and diagnostic delay, depicted on a logistic trajectory.

the relationship is more complex (Equation 4). The relationship between THIS and TOMS is sensitive to the functional form of the strength decline trajectory.

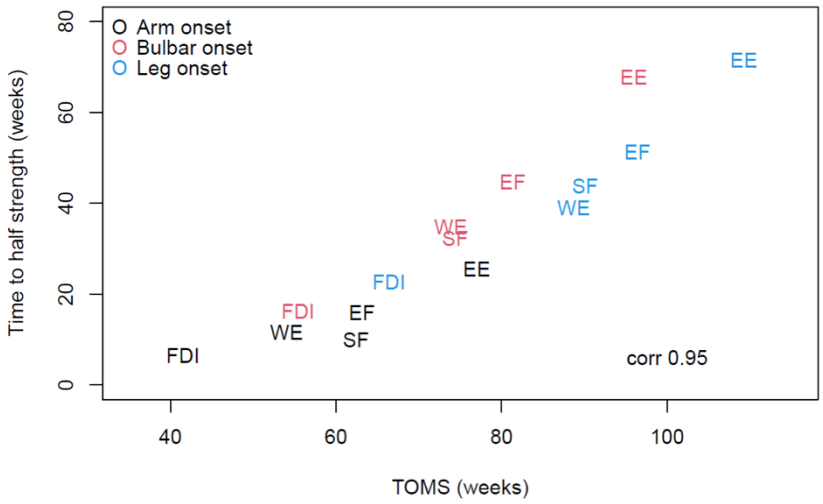
$$THIS = SCAL * log\,(2 + exp\,(\frac{TOMS-DD}{SCAL})) \qquad \text{----------(4)}$$

We estimated THIS from our data using a non-linear mixed effects model employing a Weibull functional form, with a shared shape parameter across muscles, and different scale parameters. Not unexpectedly, the order of THIS estimates closely paralleled the order of TOMS estimates (see Supplementary File Figures 9 and 10).

Although THIS requires fewer assumptions than TOMS, it is confounded by diagnostic delay and is clearly less meaningful than TOMS from a biological or disease-description perspective.



Supplementary File Figure 9: Estimates of time to half (initial) strength (THIS) for each muscle (geometric mean of two sides) by onset site. AO, BO and LO indicate arm, bulbar and leg onset.



Supplementary File Figure 10: Time to half (initial) strength (THIS) estimates plotted against corresponding TOMS estimates for muscles by site of onset. Note that they are highly correlated.

References

1 Panchabhai TS, Mireles Cabodevila E, Pioro EP, *et al.* Pattern of lung function decline in patients with amyotrophic lateral sclerosis: implications for timing of noninvasive ventilation. *ERJ Open Res* 2019;**5**. doi:10.1183/23120541.00044-2019

2 Gaur N, Stubendorff B, Grehl T, *et al.* Modelling individual amyotrophic lateral sclerosis disease courses in different centers using the D50 progression model. 2018. doi:10.13140/RG.2.2.15604.86402

3 Shefner JM, Liu D, Leitner ML, *et al.* Quantitative strength testing in ALS clinical trials. *Neurology* 2016;**87**:617–24. doi:10.1212/WNL.0000000000002941

4 Bohannon RW. Is it legitimate to characterize muscle strength using a limited number of measures? *J Strength Cond Res* 2008;**22**:166–73. doi:10.1519/JSC.0b013e31815f993d

5 Bohannon RW, Magasi SR, Bubela DJ, *et al.* Grip and knee extension muscle strength reflect a common construct among adults. *Muscle Nerve* 2012;**46**:555–8. doi:10.1002/mus.23350

6 Pinheiro J, Bates D, DebRoy S, *et al.* nlme: Linear and Nonlinear Mixed Effects Models. 2019.https://cran.r-project.org/package=nlme (accessed 21 Oct 2019).

7 Seber GA, Lee AJ. *Linear Regression Analysis*. 2nd ed. John Wiley & Sons, Inc., Hoboken, New Jersey 2003.

8. van Eijk RPA. Frequent self-assessments in ALS Clinical Trials: worthwhile or an unnecessary burden for patients? *Annals of Clinical and Translational Neurology* 2020;**7**:2074-2075. doi: 10.1002/acn3.51178