

Title: Novel phenotypic subgroups associated with the C9orf72 expansion and survival in European ALS cohorts

Supplementary Information 2 – Definition of eight level categorical hybrid variable and description of process of model simplification

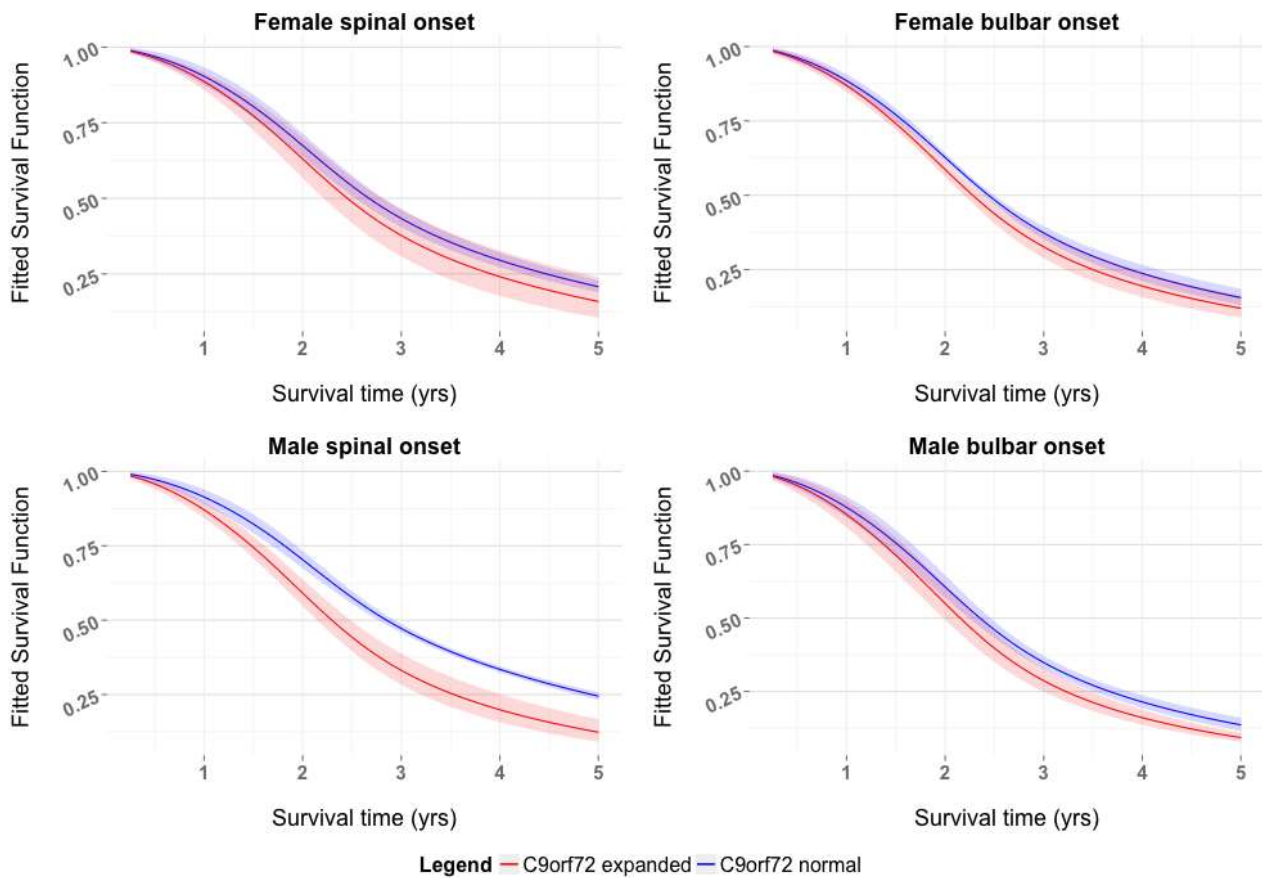
To fully explore survival in subgroups of ALS patients by sex, site of onset and C9orf72 status, a hybrid variable was created allowing for all combinations of the three variables. The level definitions and distribution of levels by country are shown in table S2.Table1.

S2.Table1. Definition and distribution of hybrid variable

Level	Belgium	Ireland	Italy	The Netherlands	United Kingdom	Total
Female, spinal onset C9orf72 normal	77 (16.1%)	138 (21.6%)	218 (25.1%)	421 (20.7%)	191 (21.1%)	1045 (21.2%)
Female, spinal onset C9orf72 expanded	22 (4.6%)	17 (2.7%)	13 (1.5%)	51 (2.5%)	12 (1.3%)	115 (2.3%)
Female, bulbar C9orf72 normal	68 (14.3%)	98 (15.3%)	164 (18.9%)	327 (16.1%)	138 (15.3%)	795 (16.1%)
Female, bulbar C9orf72 expanded	13 (2.7%)	13 (2.0%)	16 (1.8%)	45 (2.2%)	11 (1.2%)	98 (2.0%)
Male, spinal C9orf72 normal	193 (40.5%)	260 (40.6%)	331 (38.2%)	804 (39.5%)	382 (42.3%)	1970 (40.0%)
Male, spinal C9orf72 expanded	34 (7.1%)	22 (3.4%)	21 (2.4%)	57 (2.8%)	25 (2.8%)	159 (3.2%)
Male, bulbar C9orf72 normal	54 (11.3%)	82 (12.8%)	92 (10.6%)	309 (15.2%)	130 (14.4%)	667 (13.5%)
Male, bulbar C9orf72 expanded	16 (3.4%)	10 (1.6%)	12 (1.4%)	23 (1.1%)	15 (1.7%)	76 (1.5%)
Total	477	640	867	2037	904	4925

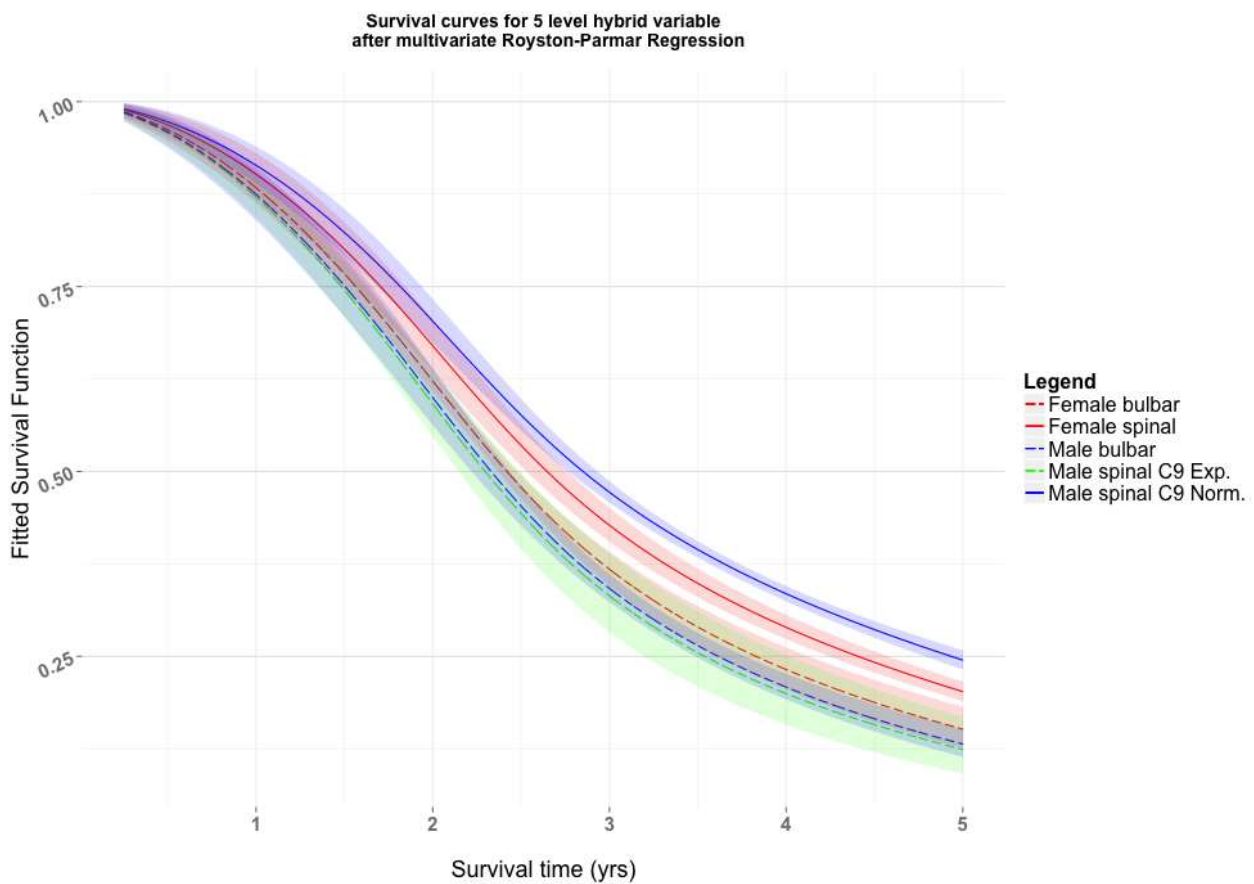
Variables for sex, site of onset and C9orf72 status were removed from the Royston-Parmar model and the above variable inserted. The modified model retained the hazard scale and 3 d.f, included age and grouped diagnostic delay as time varying covariates, including country and used a robust variance-covariance matrix clustered across country. The survival curves for this model are shown in S2Fig1.

S2. Figure 1 – Survival curves of eight level sex, site and C9 variable



These 4 plots are all generated from the same multivariate Royston-Parmar model, but printed separately for clarity. They show the effect of C9orf72 status in each combination of possible gender and site of onset subgroup. Note that the shaded areas represent 95% confidence intervals. Therefore, it can be seen that C9orf72 expanded cases have poorer survival over a wide time period only in male spinal onset patients. Therefore the model can be simplified by combining C9orf72 normal and expanded groups for female spinal onset, female bulbar onset and male bulbar onset. Thus a new five level categorical variable was created having levels: female spinal onset, female bulbar onset, male spinal onset C9orf72 normal, male spinal onset C9orf72 expanded and male bulbar onset. The eight level variable was replaced in the model with this new 5 level variable and the adjusted survival curves are shown in S2.Fig2.

S2. Figure 2 – Survival curves of five level sex, site and C9 variable



This plot was generated from a multivariate Royston-Parmar model after the 8 level variable was simplified to a 5 level variable as shown above, with otherwise the same model specifications. Again, shaded areas represent 95% confidence intervals. Male bulbar onset and female bulbar onset patients are very closely matching with overlapping confidence intervals throughout. Female spinal onset and C9orf72 normal male spinal onset curves are also in close agreement and minor differences appear only after three years. The male spinal onset C9 expanded curve (green) is however quite distinct from the other spinal onset curves and is indistinguishable from the bulbar onset curves. Therefore the hybrid variable can be further simplified into a three level variable: all bulbar onset patients, all spinal onset patients excluding male C9orf72 expanded patients, and finally male spinal onset C9orf72 expanded patients only.

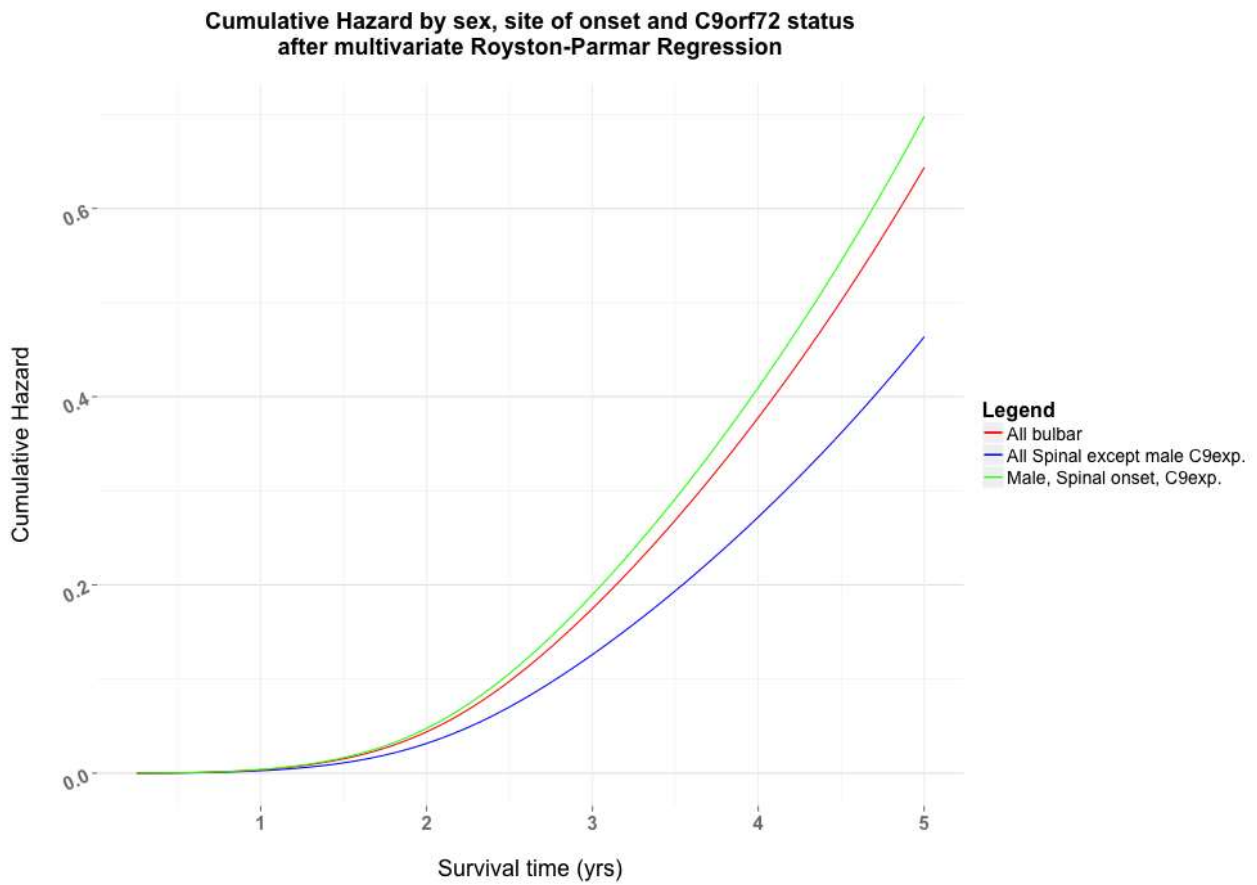
The five level variable was replaced in the model with this new 3 level variable and the details of this new variable are shown in S2 Table 2.

S2.Table2. Definition and distribution of hybrid variable

Level	Belgium	Ireland	Italy	The Netherlands	United Kingdom	Total
Spinal onset excluding C9orf72 expanded males	292 (61.2%)	415 (64.8%)	562 (64.8%)	1276 (62.6%)	585 (64.7%)	3130 (63.6%)
All bulbar onset	151 (31.7%)	203 (31.7%)	284 (32.8%)	704 (34.6%)	294 (32.5%)	1636 (33.2%)
Male spinal onset C9orf72 only	34 (7.1%)	22 (3.4%)	21 (2.4%)	57 (2.8%)	25 (2.8%)	159 (3.2%)
Total	477	640	867	2037	904	4925

The multivariate Royston-Parmar model was re-calculated using the above 3 level categorical variable, with otherwise the same model specifications as before. The survival curves for these levels from the recalculated model are shown in the main manuscript, Figure 3, whilst cumulative hazard curves are shown in S2. Fig 3.

S2. Figure 3 – Cumulative hazard curves of final three level sex, site and C9 variable



Cumulative hazard curves for the novel sex, site & C9orf72 groupings generated from the multivariate Royston-Parmar model 3 level categorical variable, with otherwise the same model specifications as before.