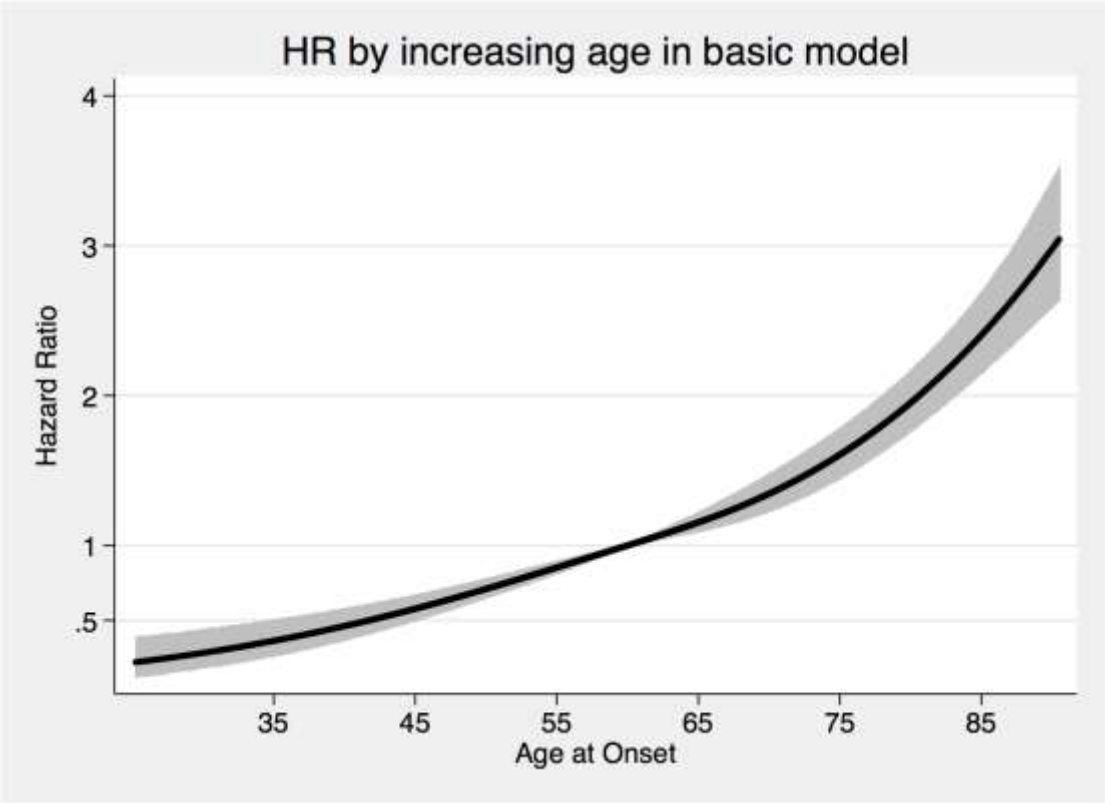
Title: Novel phenotypic subgroups associated with the C9orf72 expansion and survival in European ALS cohorts

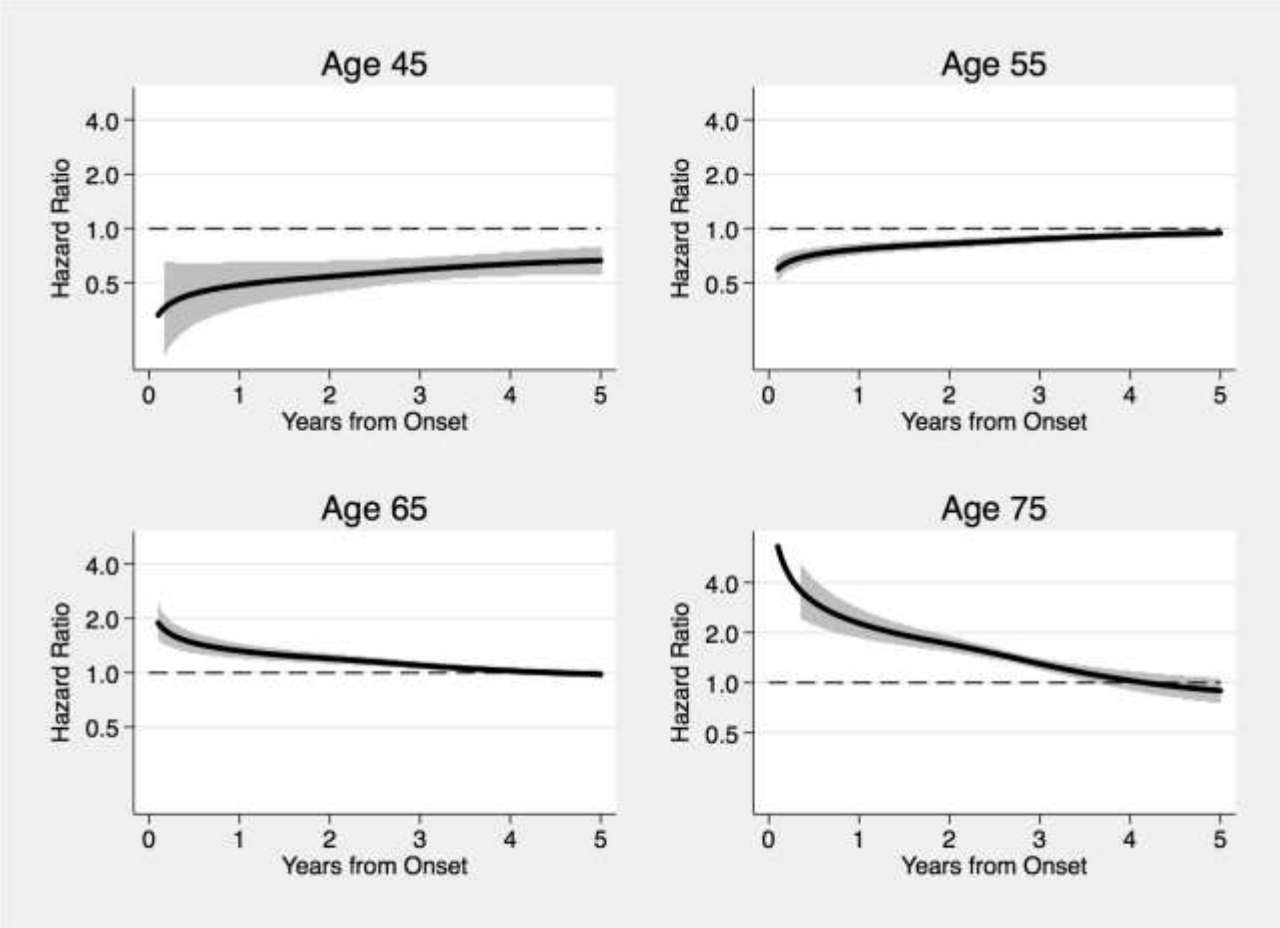
Supplementary Information 1 – Hazard ratios from basic Royston-Parmar flexible parametric model

S1. Figure1: Hazard ratio for age as a continuous variable before addition of time varying component



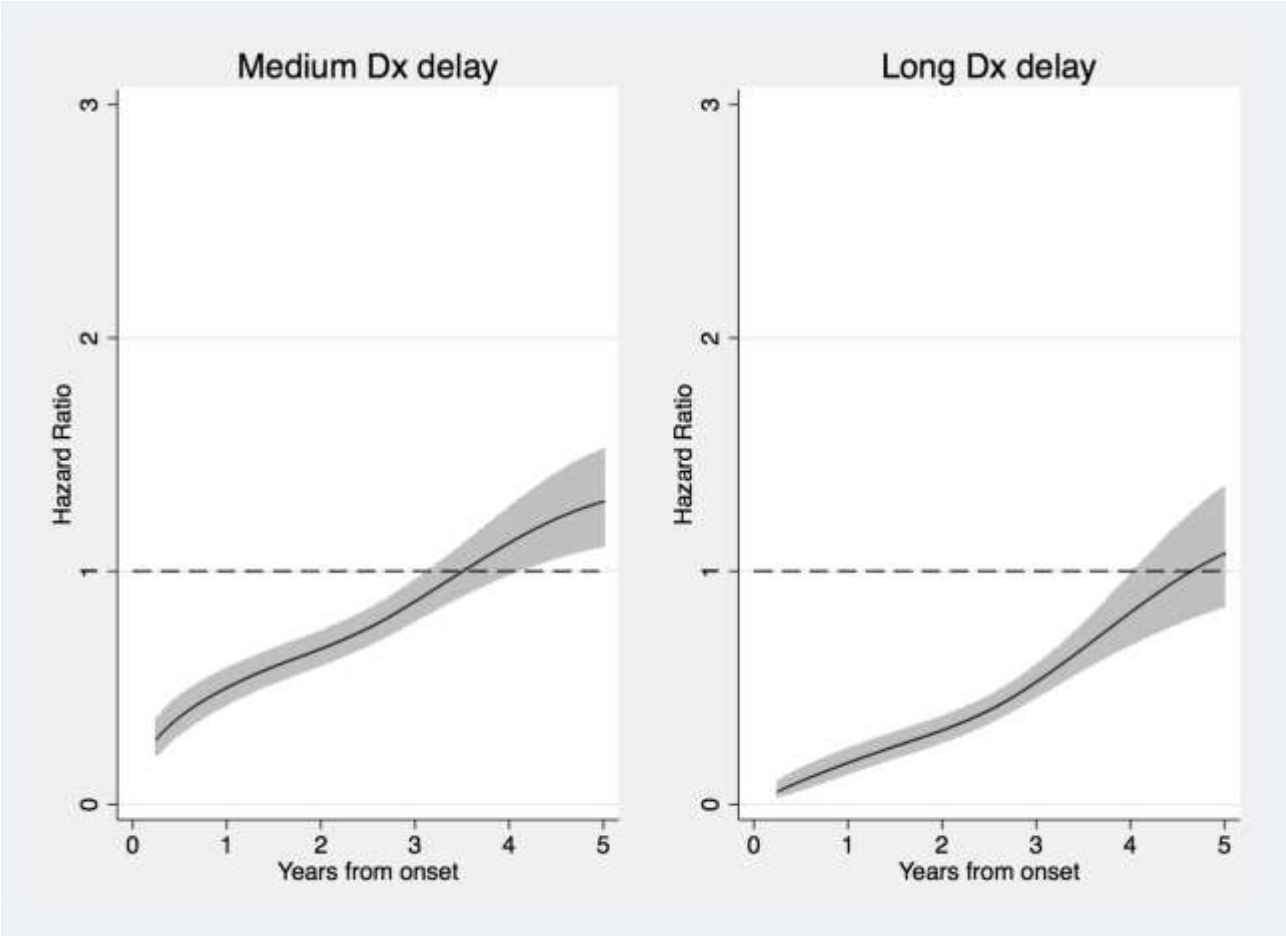
Legend: Hazard ratio for age as a continuous variable from a Royston-Parmar model on the hazard scale with 3.d.f. before the addition of time varying components. Model covariates included: site of onset, C9orf72 status, grouped diagnostic delay, country and robust variance-covariance matrix clustered by country was used.

S1. Figure 2: Hazard ratio for age as a continuous variable after addition of a 1 d.f. time varying component to the age variable



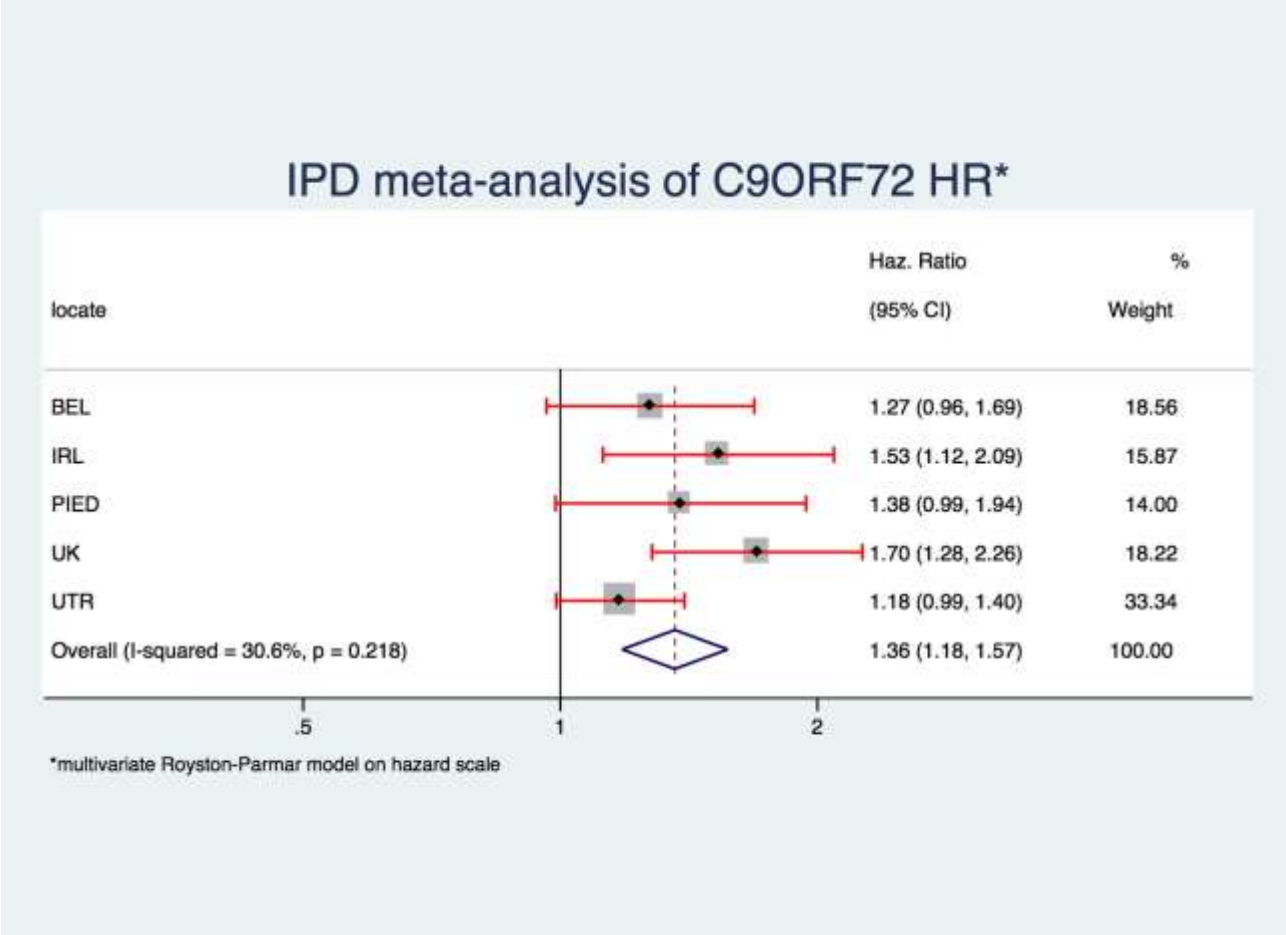
Legend: The graph shows time dependent hazard ratios for different ages of onset in a multivariate Royston Parmar flexible parametric model on the hazard scale with 3 degrees of freedom and variance-covariance matrix clustered by country. These graphs imply that the relative survival advantage or disadvantage for a given age dissipate over time – with only the 45 year old graph still showing a survival advantage after 5 years. Model covariates included site of onset, grouped diagnostic delay, country and C9orf72 status, and robust variance-covariance matrix clustered by country was used.

S1. Figure 3: Hazard ratio for grouped diagnostic delay after addition of a 1 d.f. time varying component to the diagnostic delay variable



Legend: In addition to time varying age, the basic model included time varying diagnostic delay categories shown above. Three categories of diagnostic delay were defined: short, medium and long diagnostic delay. For each country the cohort was split into three groups defined by the 33rd and 67th centiles of diagnostic delay in that country, thus allowing for inter-country differences in diagnostic delay. Short diagnostic delay is the reference group to which the above graphs are compared, therefore they indicate that for medium delay patients the survival advantage they experience has dissipated after approx. 3.5 years, whilst for long delay patients their survival advantage has dissipated after approx. 4.5 years.

S1. Figure 4: IPD meta-analysis of the hazard ratio for the C9orf72 expansion in the basic model



Legend: Individual patient data meta-analysis of hazard ratio for C9orf72 expanded compared to C9orf72 normal patients after multivariate Royston-Parmar regression. The model included, age and grouped diagnostic delay as time-varying covariates, site of onset, country and C9 status. The Stata *metan* command implements IPD meta-analysis as a two step process and we used a random effects model to pool the estimate across countries.

S1. Table 1: Hazard ratios for time independent covariates included in the basic model.

Variable	Hazard Ratio (95% CI)	Wald Test
Spinal onset	1	-
Bulbar onset	1.35 (1.25 – 1.45)	<0.001
Country		
Belgium	1	-
Ireland	1.01 (0.99 – 1.03)	0.58
Italy	0.66 (0.64 – 0.69)	<0.001
The Netherlands	0.95 (0.94 – 0.96)	<0.001
United Kingdom	0.70 (0.68 – 0.71)	<0.001
C9orf72 Normal	1	
C9orf72 Expanded	1.32 (1.14 – 1.54)	<0.001

The above hazard ratios are for the covariates included in the completed basic model alongside time dependent age at onset and diagnostic delay variables (shown in S1Fig.2 & S1Fig.3). Again robust errors were used. Note that the hazard ratio for C9orf72 generated using robust errors, 1.32 (95% CI: 1.14 – 1.54), is very similar to the estimate obtained via the alternate method IPD meta-analysis: 1.36 (95% CI: 1.18 – 1.57) (S1Fig.4).