Supplementary Online Content

Langbehn DR, Stout JC, Gregory S, et al; TRACK-HD and Track-On HD Groups. Association of CAG repeats with long-term progression in Huntington disease. *JAMA Neurol*. Published online August 12, 2019. doi:10.1001/jamaneurol.2019.2368

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods. Supplemental Methods

We also initially considered Beck Depression Inventory (BDI) and Hamilton Depression and Anxiety (HAMD) scores, annulus lengths from the direct and indirect circle tracing tasks, and a larger number of quantitative motor measurements (described in detail in the online supplement of the Tabrizi et al longitudinal analysis of the TRACK-HD study[1]. The BDI and HAMD, while correlated with each other (R = .74), showed little correlation with any other measures under consideration, nor were they predictable by age and CAG repeat length (analysis not shown), so we excluded them from the main analysis. We discarded the circle tracing tasks due to excessive practice effects. A reduction in the final number of highly correlated quantitative motor measures partially balanced the contributions of nominally cognitive versus motor items in the analysis.

An elegant, concise definition of principal component analysis (PCA) requires some linear algebra for which we presume no familiarity among the intended audience. We instead attempt a technically correct description of the key elements that motivated out use of PCA here:

If a set of clinical measures are all affected by a single aspect of HD, then we expect these measures to have some correlation with each other. Assuming this is the case, we would like to summarize these measures with a single number. We naturally consider some sort of mean. Given such a mean, we can examine its correlation with each of the original measures. PCA defines a weighted mean that will have the highest possible mean correlation with the set of original measures. Thus, the first principal component score is a principled solution to our aim of summarizing the measures with a single number. Furthermore, examining the correlations between the principal component score and each of the original measures allows us to assess whether they are similar enough to justify combination in the first place. (Additional principal components are also defined, but we make no use of them here. The second component has the highest mean correlation with the original variables under the constraint that it is uncorrelated with the first component. The third component is defined similarly but must be uncorrelated with both earlier components, and so on.)

Some readers may unknowingly have familiarity with applied PCA. In diffusion tensor analysis of the brain, axonal diffusivity (AD) is the value of the first principal component of three-dimensional water diffusivity. It measures the extent of diffusion in the direction of greatest diffusion. (Medial diffusivity is the mean of the second and third principal components.) In this paper, we use the first principal component to measure severity in the “direction” of greatest common variability among the original measures.

eTable 1: Missing data

eTable 1a. Missing data rates in the combined TRACK-HD and Track-On Studies.

Variable	N Missing	% Missing data
Symbol Digit Modalities Test (number correct)	14	0.68
Stroop word reading	3	0.15
Spot the change 5 sec	48	2.32
UHDRS total motor score	1	0.05
Paced-tapping 3hz SD of inter-tap intervals	39	1.89
Q-Motor Speeded tapping inter-tap interval	24	1.16
Putamen volume (Ratio to ICV)	144	6.97
Caudate volume (Ratio to ICV)	246	11.91
Total Brain volume (Raito to ICV)	234	11.33
Ventricle volume (Ratio to ICV)	234	11.33
White matter volume (Raito to ICV)	392	18.98
Grey matter volume (Ratio to ICV)	392	18.98
Caudate Putamen PC	290	14.04
WM Ventricle PC	408	19.76

Note that total number of observations was 2065.

eTable 1b. Missing data by HD stage in TRACK-HD and Track-On Studies.

Ratio	Controls		PreHD A		preHD B		HD Stage 1		HD Stage2		HD Stage 3+		Chi Sq	p
	Miss	%	Miss	%	Miss	%	Miss	%	Miss	%	Miss	%	(5 df)	
(Total N)	***764***		***313***		***408***		***345***		***198***		***37***			
Brain/ICV	60	*7.9%*	35	*11.2%*	37	*9.1%*	55	*15.9%*	35	*17.7%*	12	*32.4%*	42.91	<.0001
Putamen/ICV	38	*5.0%*	10	*3.2%*	14	*3.4%*	37	*10.7%*	32	*16.2%*	13	*35.1%*	97.97	<.0001
Caudate/ICV	75	*9.8%*	33	*10.5%*	35	*8.6%*	52	*15.1%*	39	*19.7%*	13	*35.1%*	42.22	<.0001
Ventricles/ICV	60	*7.9%*	34	*10.9%*	37	*9.1%*	55	*15.9%*	36	*18.2%*	12	*32.4%*	44.29	< .0001
White matter/ICV	132	*17.3%*	57	*18.2%*	69	*16.9%*	74	*21.4%*	46	*23.2%*	14	*37.8%*	14.94	0.0106
Grey matter/ICV	132	*17.3%*	57	*18.2%*	69	*16.9%*	74	*21.4%*	46	*23.2%*	14	*37.8%*	14.94	0.0106
Caudate Putamen PC	86	*11.3%*	36	*11.5%*	42	*10.3%*	61	*17.7%*	50	*25.3%*	15	*40.5%*	57.25	< .0001
WM Ventricle PC	138	*18.1%*	57	*18.2%*	71	*17.4%*	79	*22.9%*	48	*24.2%*	15	*40.5%*	18.02	0.0029

Miss = Number missing measures.

PC = principal component score.

Chi-Square tests are approximate, due to repeated within-subject measures.

eTable 2A: Regression coefficients for principal component score models in cases

	Motor-Cognitive			UHDRS Motor-Cog		
Parameter	**Coef**	**SE**	**p**	**Coef**	**SE**	**p**
Intercept	0.02304	0.05455	0.6731	0.01982	0.05437	0.7157
age50	0.09521	0.00482	<.0001	0.09504	0.00484	<.0001
cag42	0.4571	0.03108	<.0001	0.4093	0.03097	<.0001
(age50)*(cag42)	0.03596	0.00326	<.0001	0.03403	0.00325	<.0001
$(age50)^2$	0.002947	0.00037	<.0001	0.00308	0.00037	<.0001
$(age50)^2$ *(cag42)	0.000545	0.00012	<.0001	0.00047	0.00013	0.0002
$(cag42)^2$	0.045603	0.00834	<.0001	0.04342	0.00849	<.0001
(age50)*$(cag42)^2$	0.001207	0.00037	0.001	0.00105	0.00037	0.0051

	Caudate-putamen			White Matter - Ventricle			Grey Matter Std		
Parameter	**Coef**	**SE**	**p**	**Coef**	**SE**	**p**	**Coef**	**SE**	p
Intercept	0.2516	0.05397	<.0001	0.0364	0.06628	0.5834	0.2069	0.06061	0.0007
age50	0.1011	0.00231	<.0001	0.1398	0.004217	<.0001	0.0449	0.00240	<.0001
cag42	0.4982	0.02922	<.0001	0.4574	0.03786	<.0001	0.02184	0.03356	0.5156
(age50)*(cag42)	0.00935	0.00151	<.0001	0.02665	0.002685	<.0001	0.00778	0.00157	<.0001
$(age50)^2$	-0.00031	0.00016	0.047	0.003149	0.000266	<.0001	0.00085	0.00015	<.0001
$(age50)^2$*(cag42)	-0.00018	0.00006	0.0016	0.0002668	0.00010	0.0081	0.00011	0.000060	0.0787
$(cag42)^2$	-0.01636	0.00469	0.0005	0.0158381	0.007267	0.0297	0.00763	0.00524	0.1462
(age50)*$(cag42)^2$	-0.00104	0.00020	<.0001	0.0001663	0.000302	0.5817	0.000017	0.00020	0.9332

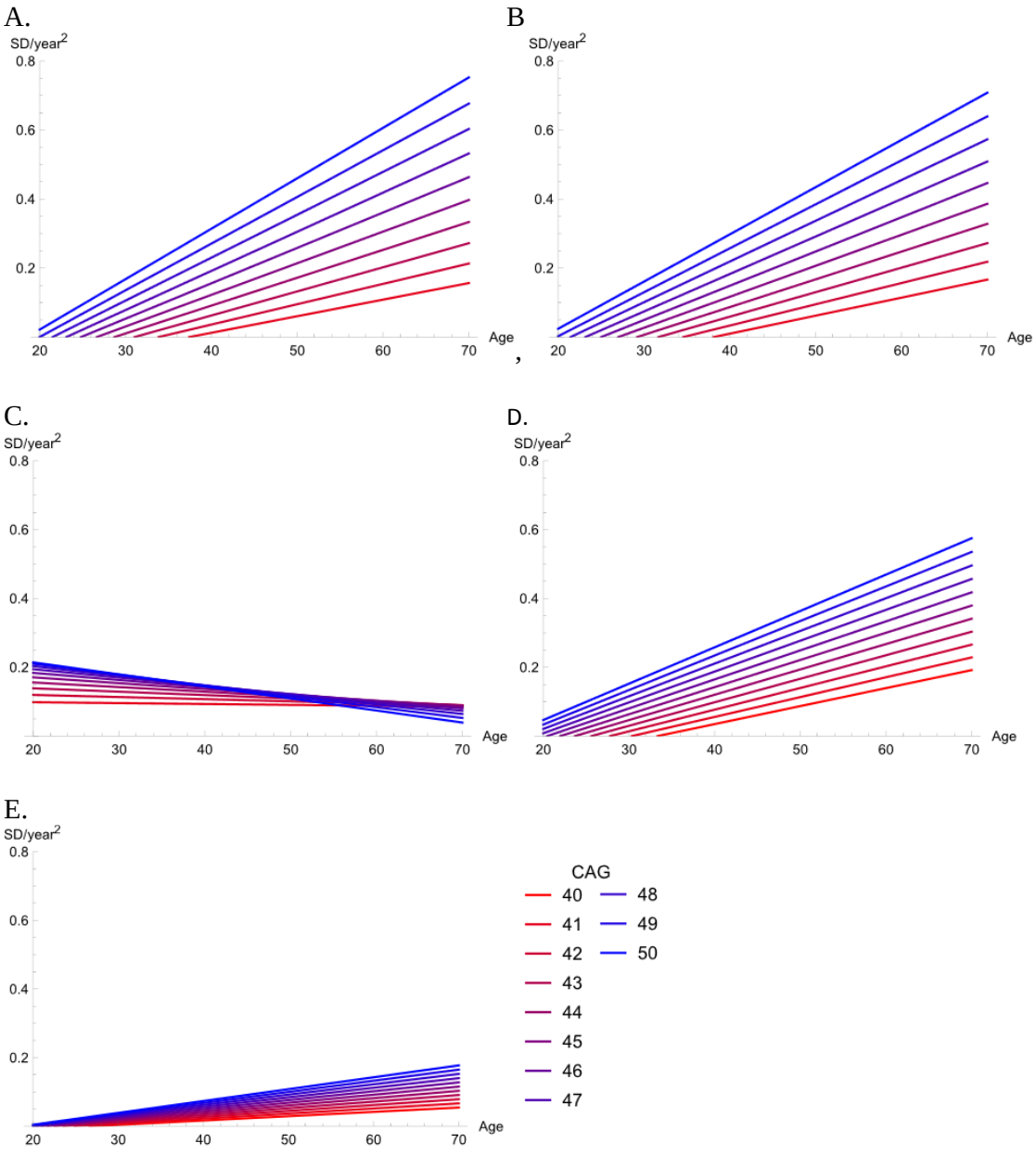
eTable 2B.Regression coefficients for principal component score models in controls

	Motor-Cognitive			UHDRS Motor-Cog		
Parameter	**Coef**	**SE**	**p**	**Coef**	**SE**	**p**
Intercept	-0.8897	0.0337	<.0001	-0.809	0.04449	<.0001
age50	0.006714	0.00244	0.0062	0.00469	0.0033	0.1557
$(age50)^2$	0.000184	0.00016	0.2399	0.00047	0.0002	0.0204

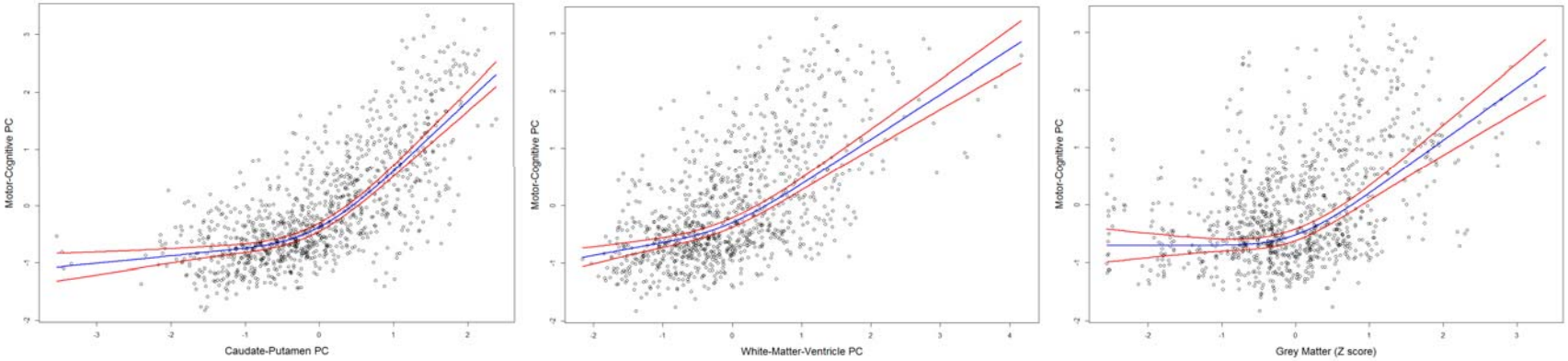
	Basal Ganglia			White Matter - Ventricle			Grey Matter Std		
Parameter	**Coef**	**SE**	**p**	**Coef**	**SE**	**p**	**Coef**	**SE**	**p**
Intercept	-1.5497	0.04487	<.0001	-0.6761	0.06789	<.0001	-0.3288	0.06513	<.0001
age50	0.03327	0.00188	<.0001	0.07469	0.003904	<.0001	0.03364	0.00279	<.0001
$(age50)^2$	0.000043	0.000096	0.657	0.001073	0.000204	<.0001	0.00019	0.00014	0.173

Age and cag expansion lengths are centered for interpretability and estimation stability. Age50 = age (yrs) – 50. Cag42 = cag expansion length – 42. Coef = Regression coefficient. SE = Standard Error. Note that age and CAG are centered around typical values in the data to assist interpretability of the model intercepts.

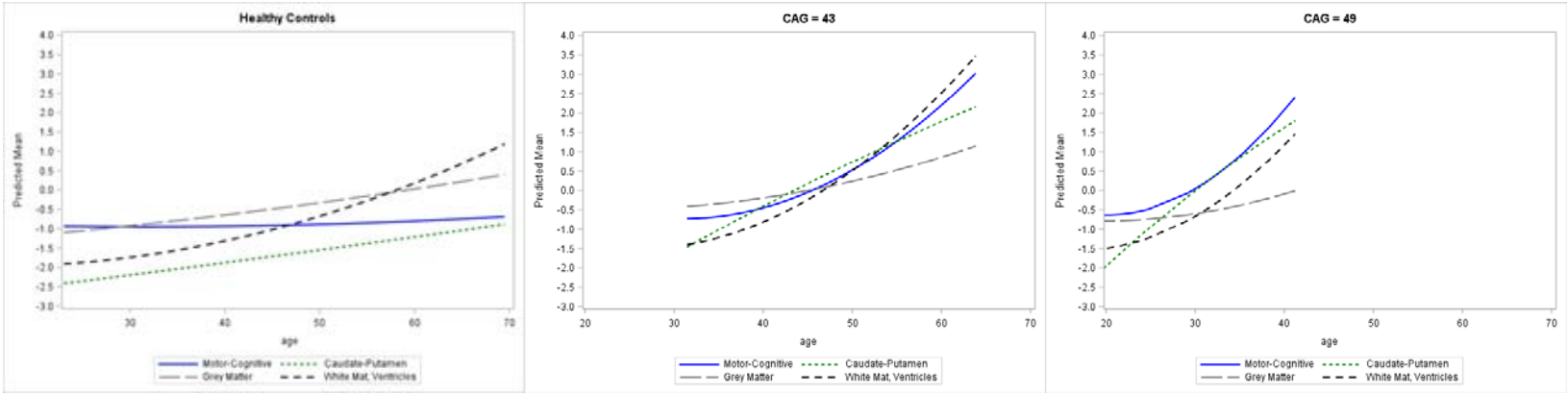
eFigure 1. Estimated Acceleration Rates by CAG Repeat Length: The models linking severity progression to age and CAG length are quadratic. (See Supplementary table 2A and Figure 2 of the main paper.) This implies a linear increase with ageing for the acceleration of progression. Here we plot those concomitant predicted rates of principal component score acceleration. Units of measure are principal component score standard deviations per year2. Horizontal scaling is constant across the plots A) Motor-Cognitive PC; B) UHDRS Motor-Cog PC; C) Caudate-Putamen PC; D) White-Mat-Vent PC; E) Standardized matter score.



eFigure 2: **Non-linear relationship between the motor-cognitive score and brain imaging measures:** Mixed-effect restricted cubic spline modelling shows the non-linear relationship between the motor-cognitive score and brain imaging measures. These models used only a random intercept (per participant) and no random slope, as we do not explicitly model those relationships as a function of age Slopes increase notably when the volume scores reach 0 or higher. There is no significant relationship between grey matter scores < 0 and the motor-cognitive scores. Patterns (not shown) are similar if the UHDRS score is substituted for the motor-cognitive score. The blue curves are the expected values and the red are the 95% CI boundaries.



eFigure 3: **Comparison of age-dependent Motor-Cognitive and brain volume PC scores in three groups**:: Healthy controls, CAG repeat length 43 (a fairly typical HD repeat length), and CAG repeat length 49 (a relatively long repeat length). Axes are scaled identically for all plots. Predictions are only depicted for the age ranges within our data.



eResults. Supplemental Results

The curves in figure 2 form portions of CAG-specific parabolas. A complete parabola is U-shaped, and we of course do not mean to extrapolate and imply that at earlier, unobserved ages these scores were worse, then improved, and then began to increase again. Instead, it is an empirical observation that the "flat" parts of the parabolas approximately coincide with the least severe mean scores, seen in the youngest subjects for a given CAG length.

In supplementary figure 2 we simultaneously plot the mean relationship with age of the motor-cognitive score and the three brain scores, displaying these for healthy controls, for a typical CAG repeat length of 43, and for the rarer, longer repeat length, 49. Extrapolation is avoided by plotting the patterns only within the observed age ranges for each group. In controls, motor-cognitive function is maintained despite insidious age-related volume loss. The shapes of the mean loss patterns are similar for CAG repeat lengths 43 and 49. However, compared to CAG length 43, the compressed time-scale and earlier age for CAG length 49 are clearly evident.

The TFC score from the UHDRS Motor-Cog was excluded from the PCA because of the substantial ceiling effect. The ceiling effect also causes notable differences between the predicted mean and median TFC scores in some models. We illustrate them both in Figure 3, as either may be the more relevant, depending on context. There is little difference between the mean and median estimates for the clinical PC scores, but discrepancy appears in each of the brain volume scores. For example, among those with a caudate-putamen score 0, the mean TFC is about 11.5, and a few subjects already have severe TFC loss. However, the median is still nearly 13—about half of such subjects have no TFC loss. Note that, once TFC scores begin to decline, the mean and median rates are nearly linear relative to each of the PC scores.

eAcknowledgments

We thank the following for their contribution to the TRACK-HD and TrackOn studies:

Name		Institution
Campbell	JC	Monash University, Australia
Campbell	M	Monash University, Australia
Cash	DM	IXICO, London, UK
Desikan	M	University College London, UK
Fox	NC	University College London, UK
Frajman	E	Monash University, Australia
Hicks	S	Oxford University, UK
Hoffman	A	Ruhr University, Germany
Keenan	S	University College London, UK
Kennard	C	Oxford University, UK
Kloppel	S	Freiburg University, Germany
Kraus	P	Ruhr University, Germany
Lee	S	MGH, USA
Mayer	I	Ulm University
Milchman	C	Monash University, Australia
Mincova	L	Freiburg University
Monaco	W	MGH, USA
Ordidge	R	University College London, UK
Orth	M	Ulm University
Patel	A	University College London, UK
Pourchot	C	Monash University, Australia
Razi	A	University College London, UK

eReferences

1. Tabrizi SJ, Langbehn DR, Leavitt BR, et al. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *Lancet. Neurol.* 2009;8(9):791-801.