Statistical Appendix:

1. *Analysis of non-imaging outcome variables*

a) *Basic linear mixed model for repeated measures in a single group*

The model is a standard random slopes model relating the outcome to time since randomization, with the addition of random site and site-by-visit interactions. Site effects were incorporated in this fashion since plots of the means by site and visit suggested that such parameters, rather than site specific random slopes, best describe the data.

$$y_{ij} = (\alpha + a_i) + (\beta + b_i)t_{ij} + c_{s(i)} + d_{s(i)j} + e_{ij} \quad (1)$$

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}\right),\ c_{s(i)} \sim N(0, \sigma_c^2),\ d_{s(i)j} \sim N(0, \sigma_d^2),\ e_{ij} \sim N(0, \sigma_e^2)$$

where y_{ij} is the value of the outcome variable for the ith person at the jth visit, t_{ij} is the time of that visit relative to randomization (in years) and $s(i)$ is the subject's site ($s(i) = 1,2,3,4$).

Implied variance of a single outcome measure:

$$\mathrm{Var}(y_{ij}) = \sigma_a^2 + 2t_{ij}\sigma_{ab} + t_{ij}^2\sigma_b^2 + \sigma_c^2 + \sigma_d^2 + \sigma_e^2 \quad (2)$$

Implied covariance between two measures on the same person (j not equal to k):

$$\mathrm{Cov}(y_{ij}, y_{ik}) = \sigma_a^2 + (t_{ij} + t_{ik})\sigma_{ab} + t_{ij}t_{ik}\sigma_b^2 + \sigma_c^2 \quad (3)$$

Implied variance of difference between two measures on the same person:

$$\mathrm{Var}(y_{ij} - y_{ik}) = (t_{ij} - t_{ik})^2\sigma_b^2 + 2\sigma_d^2 + 2\sigma_e^2 \quad (4)$$

Implied variance of a rate of change derived from a difference between two measures on the same person:

$$\mathrm{Var}\left((y_{ij} - y_{ik})/(t_{ij} - t_{ik})\right) = \sigma_b^2 + \left((2\sigma_d^2 + 2\sigma_e^2)/(t_{ij} - t_{ik})^2\right) \quad (5)$$

Implied covariance between two measures on different people from the same site at the same/different visits (j not equal to k):

$$\mathrm{Cov}\left(y_{i_1 j}, y_{i_2 j}\right) = \sigma_c^2 + \sigma_d^2 \qquad (5) \qquad \mathrm{Cov}\left(y_{i_1 j}, y_{i_2 k}\right) = \sigma_c^2 \qquad (6)$$

b) *Adjustment for covariates and other modelling details*

To improve the extent to which normality assumptions were satisfied UHDRS-TMS was square root transformed, UHDRS-TFC was square root transformed after values were subtracted from 13 (the maximum value) and the speeded tapping inter-onset interval standard deviation was log transformed. For the SDMT and speeded tapping outcomes frequent convergence problems were encountered with the bootstrap (see below) and so the centre by visit interaction terms ($d_{s(i)j}$), whose variance was usually estimated to be zero, was omitted from all models.

Models which that allow the intercept and slope to depend upon a vector of covariates ($\boldsymbol{x}$) were also fitted to permit investigation of the effect of stratification on such covariates on required sample sizes. These involve adding appropriate fixed effects ($\boldsymbol{\gamma}^T \boldsymbol{x}_i + \boldsymbol{\partial}^T \boldsymbol{x}_i t_{ij}$) to the model to give the following.

$$y_{ij} = (\alpha + \boldsymbol{\gamma}^T \boldsymbol{x}_i + a_i) + (\beta + \boldsymbol{\partial}^T \boldsymbol{x}_i + b_i) t_{ij} + c_{s(i)} + d_{s(i)j} + e_{ij} \qquad (7)$$

In the early-HD group covariates that were investigated were age at baseline, CAG repeat length, disease burden, gender, stage (I or II) and education (measured on a standardized seven-point scale as in earlier TRACK-HD publications[5-8]). Of these there were statistically significant effects of age, CAG repeat length, disease burden and stage on the slope in the early-HD group for at least one of the four non-imaging variables considered here and so these four factors were included in the models from which required sample sizes were estimated. In the control group covariates that were investigated were age, gender and education. Of these there were only statistically significant effects of age on slopes and so only age was included in the models from which required sample sizes were estimated.

2. *Analysis of imaging variables*

a) *Basic linear mixed model for repeated "direct" measures of change in a single group*

The general formula for a measured difference between two visits is as follows. This is the model proposed by Frost and colleagues[25] with the addition of random site-by-visit interaction terms.

$$c_{ijk} = (\beta + b_i)(t_{ik} - t_{ij}) - u_{ij} + u_{ik} - v_{s(i)j} + v_{s(i)k} + w_{ijk} \quad (8)$$

$$b_i \sim N(0, \sigma_b^2),\ u_{ij} \sim N(0, \sigma_u^2),\ v_{s(i)j} \sim N(0, \sigma_v^2),\ w_{ijk} \sim N(0, \sigma_w^2)$$

where c_{ijk} is the measured change in the outcome variable for the *i*th person between the *j*th and the *k*th visits, other notation as above.

However for the data considered here all changes were measured from visit 1 at time 0, so the suffix *j* can be dropped from equation (7) giving

$$c_{ik} = (\beta + b_i)t_{ik} - u_i + u_{ik} - v_{s(i)} + v_{s(i)k} + w_{ik} \quad (9)$$

$$b_i \sim N(0, \sigma_b^2),\ u_i \text{ and } u_{ik} \sim N(0, \sigma_u^2),\ v_{s(i)} \text{ and } v_{s(i)k} \sim N(0, \sigma_v^2),\ w_{ik} \sim N(0, \sigma_w^2)$$

Model (9) can be written using more standard notation by defining $m_{s(i)} = -v_{s(i)}$, $n_{s(i)k} = v_{s(i)k}$, $p_i = -u_i$ and $q_{ik} = u_{ik} + w_{ik}$.

$$c_{ik} = (\beta + b_i)t_{ik} + m_{s(i)} + n_{s(i)k} + p_i + q_{ik} \quad (10)$$

$$b_i \sim N(0, \sigma_b^2),\ m_{s(i)} \sim N(0, \sigma_m^2),\ n_{s(i)k} \sim N(0, \sigma_n^2),\ p_i \sim N\left(0, \sigma_p^2\right),\ q_{ik} \sim N(0, \sigma_q^2)$$

For strict equality between equations (9) and (10) σ_q^2 should be constrained to be larger than σ_p^2. However, in practice when model (10) was fitted it was frequently found that parameter estimates did not satisfy this constraint, so this was not enforced. For convenience we also did not constrain σ_m^2 and σ_n^2 to be equal, as strict agreement between models (9) and (10) would imply.

Implied variance of a single direct measure of change:

$$\mathrm{Var}(c_{ik}) = t_{ik}^2\sigma_b^2 + \sigma_m^2 + \sigma_n^2 + \sigma_p^2 + \sigma_q^2 \quad (11)$$

Implied variance of a rate derived from a single direct measure of change:

$$\mathrm{Var}(c_{ik}/t_{ik}) = \sigma_b^2 + \left(\sigma_m^2 + \sigma_n^2 + \sigma_p^2 + \sigma_q^2\right)/t_{ik}^2 \quad (12)$$

Implied covariance between two direct measures of change on the same person (k not equal to l):

$$\mathrm{Cov}(c_{ik}, c_{il}) = t_{ik}t_{il}\sigma_b^2 + \sigma_m^2 + \sigma_p^2 \quad (13)$$

Implied covariance between two direct measures of change on different people from the same site at the same/different visits (j not equal to k):

$$\mathrm{Cov}\left(c_{i_1k}, c_{i_2k}\right) = \sigma_m^2 + \sigma_n^2 \quad (14)$$

b) <u>*Adjustment for covariates and other modelling details*</u>

Both whole brain and caudate volumes were analysed on three different scales: i) absolute changes in volume (in mls), ii) percentage changes using the logarithmic approach described by Frost and colleagues[25] where the outcome variable is defined to be log(1 + (direct measure of change/baseline volume))and iii) changes as a percentage of total intra-cranial volume (ICV) where (for consistency with ii)) the outcome variable is defined to be log(1 + (direct measure of change/ICV)).

Models that allow the slope to depend upon a vector of covariates ($\boldsymbol{x}$) were also fitted to permit investigation of the effect of stratification on such covariates on required sample sizes. These involve adding appropriate fixed effects ($\boldsymbol{\partial}^T\boldsymbol{x}_i(t_{ik} - t_{ij})$) to the model. The same covariates as for the non-imaging variables together with the baseline measure of the respective volume were investigated. Of these there were statistically significant effects of age, CAG repeat length, disease burden and baseline volume on the slope in the early HD group for at least one of the six imaging variables considered here and so these four factors were included in the models for which required sample sizes were estimated. In controls age and baseline brain volume were included in such models.

3. *Sample sizes for designs of interest*

A number of authors including Dawson[39] and Frost and colleagues[16] have shown how required sample sizes for a particular clinical trial design can be computed provided that there are postulated values for the parameters in the model that is to be used for the trial analysis. We explain this using (slightly revised) material from Frost and colleagues[16].

The analysis model for early HD patients is extended to a "treatment trial" model appropriate when there are two groups of patients receiving different treatments, with the slope allowed to differ between the two groups. For example, extending the model specified in equation (1) in this way gives the following.

$$y_{ij} = (\alpha + a_i) + (\beta + \tau g_i + b_i)t_{ij} + c_{s(i)} + d_{s(i)j} + e_{ij} \quad (15)$$

where g_i takes the values 0 and 1 in placebo and intervention groups respectively.

Once the analysis model is specified sample size requirements follow from the theory of linear mixed models as follows.

A general formulation for a linear mixed model is

$$\boldsymbol{Y}|\boldsymbol{u} \sim \boldsymbol{N}[\boldsymbol{X\beta} + \boldsymbol{Zu}\,;\,\boldsymbol{R}] \text{ for } \boldsymbol{u} \sim \boldsymbol{N}[\boldsymbol{0}\,;\,\boldsymbol{G}]. \quad (16)$$

This implies, marginally, that

$$\boldsymbol{Y} \sim \boldsymbol{N}[\boldsymbol{X\beta}\,;\,\boldsymbol{\Sigma}] \text{ where } \boldsymbol{\Sigma} = \boldsymbol{R} + \boldsymbol{ZGZ^T}. \quad (17)$$

Here $\boldsymbol{Y}$ is the vector of outcome variables, $\boldsymbol{X}$ is the design matrix, $\boldsymbol{\beta}$ is the vector of fixed effects and $\boldsymbol{\Sigma}$ is the variance-covariance matrix for the residuals. If a linear mixed model is to be used to analyse a randomised controlled trial then one of the elements of $\boldsymbol{\beta}$ will correspond to a treatment effect (τ in equation (15)). Without loss of generality consider this to be β_1.

Provided that there is a postulated fixed value for the variance-covariance matrix then

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T\Sigma^{-1}X})^{-1}\boldsymbol{X^T\Sigma^{-1}Y} \quad (18)$$

and

$$\mathrm{V}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Sigma}^{-\mathbf{1}}\boldsymbol{X})^{-1}. \quad (19)$$

Equation (19) defines a covariance matrix that permits calculation of the standard error of the treatment effect $\left(s = \sqrt{\left[V(\hat{\beta})\right]_{11}}\right)$ for any design matrix and postulated $\boldsymbol{\Sigma}$. To determine sample size requirements for a complex design a computationally convenient approach is to first use equation (19) to estimate the standard error of the treatment effect for a hypothetical 'two subject trial' involving one person in each of two arms $\left(s_{two\ subject\ trial}\right)$. Since the standard error for a trial with *N* subjects in each arm is $s_{two\ subject\ trial}/\sqrt{N}$ it follows from standard theory that the necessary sample size to identify a postulated treatment difference τ with 90% statistical power, using a two-sided 5% significance level is

$$N = \left[\frac{(1.96+1.282)s_{two\ subject\ trial}}{\tau}\right]^2. \quad (20)$$

The sample size formula depends upon the design matrix $\mathbf{X}$ (which is dependent upon the number and spacing of the trial visits), and also on the various components of variance and covariance which are assumed to be the same in the treatment trial model and analogous analysis model (for the early HD patients).

The ratio of the treatment effect to its standard error for the two subject trial ($\tau/s_{two\ subject\ trial}$) can be thought of as a unit-free effect size, easily convertible to a sample size using equation (20).

To switch to 80% statistical power the 90th percentile of the N(0,1) distribution (1.282) should be replaced in equation (20) by the 80th percentile (0.842), multiplying the required sample size by 0.747 $(0.747 = (1.960 + 0.842)^2/(1.960 + 1.282)^2)$.

For each of our basic models (without adjustment for covariates such as age) we considered 'two subject trials' where each person is from the same site, hence giving sample sizes that relate to a trial that is stratified by site. For imaging outcomes we consider designs where "direct" measures of change are only available from baseline to each follow-up visit (as in TRACK-HD). "Direct" measures of change between other

pairs of visits could also calculated, and incorporated in the design matrix and statistical analysis, but gains in efficiency are likely to be small and we did not pursue this here.

This basic approach was extended to compute sample size requirements for stratified randomized controlled trials with covariate adjustment for the stratification factors. Extending the model specified in equation (15) to allow intercepts and slopes to depend upon a vector of covariates ($\boldsymbol{x}$) gives the following model.

$$y_{ij} = (\alpha + a_i + \boldsymbol{\gamma}^T \boldsymbol{x}_i) + (\beta + \tau g_i + b_i + \boldsymbol{\partial}^T \boldsymbol{x}_i) t_{ij} + c_{s(i)} + d_{s(i)j} + e_{ij} \quad (21)$$

Again the standard theory of linear mixed models gives a formula for the variance of the treatment effect (τ) in the treatment trial model for the particular design being considered here. Provided it is assumed that covariates are perfectly balanced by randomisation arm the estimates of ($\boldsymbol{\gamma}^T$, $\boldsymbol{\partial}^T$) are orthogonal to those for the treatment effect (τ) and hence neither these estimates, nor their variances and covariances, need be formally computed. The 'two subject trial' method of computing the variance of the treatment effect referred to above can simply be amended such that the two people in question now have with identical values of all stratification factors, with all variances and covariances taken from the model (equation (7)) that adjusts for these factors.

For UHDRS-TMS and UHDRS-TFC the anticipated treatment effect for a randomised controlled trial without stratification for any factor other than site was 20 or 40% of the estimated mean rate of change in early HD ($\hat{\beta}$) whilst for SDMT, speeded tapping and the imaging variables the treatment effects were estimated after first subtracting the analogous estimated mean rate of change in controls from that in early HD.

Defining the anticipated treatment effect for a stratified analyses is slightly more complex because the analysis model (equation (16)) allows the rates of change (although not the treatment effect) to depend upon the stratification factors. Our approach for UHDRS-TFC and UHDRS-TMS was to take the anticipated treatment effects to be 20 or 40% of the mean fitted rates of change ($\sum_{i=1}^{n}\left(\hat{\beta} + \widehat{\boldsymbol{\partial}^T}\boldsymbol{x}_i\right)/n$) using

the notation in equation (7)). For SDMT, speeded tapping and the imaging variables the treatment effects were 20 or 40% of the difference in the mean fitted rates of change in the early HD and control groups. This necessitated (particularly for calculating bootstrap confidence intervals) jointly fitting the models for early HD and controls. Ideally this would have been done allowing all the parameters to be estimated separately in the two groups. However, in practice to avoid frequent non-convergence, the site effects (but not the site-by-visit interaction effects) were assumed to be the same in both groups. This had little impact on the other parameter estimates in the model.

4. *Confidence intervals for sample sizes*

We constructed non-parametric bias-corrected and accelerated (BCa) confidence intervals from 2000 bootstrap samples for each of our sample size estimates. The confidence intervals were constructed on the "effect sizes" described in the previous section (see Tabrizi and colleagues[7] for further details). The distribution of estimated effect sizes is likely to be more symmetric than that of estimated sample sizes and so confidence intervals calculated on this scale are likely to have better coverage properties.

An extension of previously published work is that confidence intervals were constructed around estimated sample sizes to provide a guide to the precision of the estimates. As in previous applications[22, 26] the Bootstrap[27] was used to do this since the sampling distribution of sample size estimates is complex and not readily amenable to approximation with explicit algebraic formulae. The other advantage of utilising bootstrap confidence intervals is that they provide additional robustness if the assumptions of the linear mixed models used in the analysis do not hold exactly.

5. *Adjusting sample sizes for dropout.*

To assess the impact of dropouts we adopt a Pattern-Mixture approach as advocated by Dawson and Lagakos[28, 29] and described by Frost and colleagues[16]. In brief, it is

assumed that subjects will be separated into strata according to missing data patterns, with each stratum first analysed separately. The overall treatment effect is a weighted mean of the stratum specific effects, with the weights equal to the reciprocals of the stratum specific variances. Based on what was observed in TRACK-HD we assumed 5% dropouts in the first year, 5% in the second year and 15% in the third year for the non-imaging outcomes. For the imaging outcomes analogous assumed rates were 15%, 10% and 10%. For simplicity we do not allow for missing data other than that resulting from dropout, *i.e.* we do not make allowance for individuals who might have intermittent missing values during follow-up.