Survival Endpoints for Huntington's Disease Trials Prior to Motor Diagnosis

Jeffrey D. Long, James A. Mills, Blair R. Leavitt, Alexandra Durr, Raymund A. Roos, Julie C. Stout, Ralf Reilmann, Bernhard Landwehrmeyer, Sarah Gregory, Rachael I. Scahill, Douglas R. Langbehn, Sarah J. Tabrizi and the Track-HD and Track-On investigators

Author Affiliation, Contact, and Email

Jeffrey D. Long: Departments of Psychiatry and Biostatistics, Carver College of Medicine, University of Iowa, jeffrey-long@uiowa.edu

James A. Mills: Department of Psychiatry, Carver College of Medicine, University of Iowa, jim-mills@uiowa.edu

Blair R. Leavitt: Centre for Molecular Medicine and Therapeutics, Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada, bleavitt@cmmt.ubc.ca

Alexandra Durr: Department of Genetics and Cytogenetics, and INSERM UMR S679, APHP, ICM Institute, Hôpital de la Salpêtrière, Paris, France, alexandra.durr@upmc.fr

Raymund A. Roos: Department of Neurology, Leiden University Medical Centre, Leiden, Netherlands, R.A.C.Roos@lumc.nl

Julie C. Stout: School of Psychology and Psychiatry, Monash University, Melbourne, VIC, Australia, julie.stout@monash.edu

Ralf Reilmann: Department of Neurology, University of Mu¨nster, Mu¨nster, Germany, ralf.reilmann@ghi-muenster.de

Bernhard Landwehrmeyer: CHDI Foundation, University of Ulm, Department of Neurology, bernhard.landwehrmeyer@uni-ulm.de

Sarah Gregory: Huntington’s Disease Research Centre, UCL Institute of Neurology, London, UK, s.gregory@ucl.ac.uk

Rachael I. Scahill: University College London Institute of Neurology, University College London, London, UK, Scahill, r.scahill@ucl.ac.uk

Douglas R. Langbehn: Departments of Psychiatry and Biostatistics, Carver College of Medicine, University of Iowa, Langbehn, douglas-langbehn@uiowa.edu

Sarah J. Tabrizi: UCL Huntington's Disease Centre, Dept of Neurodegenerative Disease, University College London Institute of Neurology, s.tabrizi@ucl.ac.uk

Corresponding Author

Jeffrey D. Long, Carver College of Medicine, 500 Newton Road Iowa City, IA 52242-1000, jeffrey-long@uiowa.edu

Keywords

Huntington's disease, motor diagnosis, progression free survival

Key Points

Question: Can survival endpoints for pre-diagnosis Huntington's disease (HD) trials be developed to provide feasible sample sizes?

Findings: Progression-free survival using a composite of motor diagnosis or a HD progression event yields much smaller sample sizes than using the event of motor diagnosis alone. The HD progression events show good external validity when applied to an independent study.

Meaning: Progression-free survival provides feasible sample sizes for clinical trial planning with HD gene expansion mutation carriers who have not yet received a motor diagnosis.

Tweet

Progression-free survival drastically reduces the sample size required for planning clinical trials in Huntington’s disease.

Abstract

Importance: Predictive genetic testing in Huntington's disease (HD) enables therapeutic trials in HD gene expansion mutation carriers prior to motor diagnosis. Progression-free survival (PFS) is defined as the composite of motor diagnosis or a HD progression event, whichever comes first.

Objective: Determine if PFS provides feasible sample sizes for trials with mutation carriers who have not yet received a motor diagnosis.

Design: Pre-diagnosis mutation carriers are analyzed from the Track-HD (collected 2008-2011) and Track-On (collected 2011-2014) cohort studies, with up to 7 years of follow-up. Clinically meaningful change of a Unified Huntington's Disease Rating Scale (UHDRS) variable is used to define a progression event. Results are externally validated with data from the Cooperative Huntington Observational Research Trial (COHORT; collected 2006-2011) cohort study. Required sample size is estimated for a two-arm pre-diagnosis clinical trial.

Setting: The Track studies are longitudinal cohort studies conducted like clinical trials, but without intervention, with four sites. COHORT is a longitudinal cohort study with 38 sites.

Participants: Track had 167 pre-diagnosis mutation carriers and 156 non-carriers. Inclusion criteria were 18 to 65 years of age, toleration of MRI/biosample collection, no major psychiatric disorder/head injury, no therapeutic trial enrollment, and no limiting medical conditions. Non-carriers were age- and gender-matched spouses or partners of participants or non-carrier siblings. COHORT had 366 pre-diagnosis mutation carriers recruited with broader inclusion criteria (wider age range, no imaging).

Main outcomes and measures: The primary endpoint is PFS. HD progression events are defined for the UHDRS total motor score, total functional capacity, symbol digit modalities test, and Stroop word test.

Results: PFS survival curves of the Track mutation carriers (*N* = 156, 56% female, mean age = 40.06) show good external validity with the COHORT mutation carriers (*N* =

366, 63% female, mean age = 42.21) after adjusting for initial progression. For required sample size, PFS with motor diagnosis or TMS progression requires about 4 times fewer participants than motor diagnosis alone. Including additional cognitive progression events further reduces the number.

Conclusions and Relevance: Reasonably sized pre-diagnosis HD trials can be planned with PFS and there is evidence of generalization of the approach.

Introduction

Huntington's disease (HD) is a devastating neurodegenerative disorder caused by a cytosine-adenine-guanine (CAG) expansion on the HTT gene of chromosome 4[1]. HD is autosomal dominant with reduced penetrance for $36\text{-}39$ repeats, and full penetrance for ≥ 40 repeats[2]. Progression leads to a triad of signs and symptoms including motor, cognitive, and behavioral features[3]. A reliable predictive genetic test is available that can be used to establish whether an individual is a HD gene expansion mutation carrier prior to the emergence of any signs or symptoms. Early identification of carrier status enables trials to examine whether a therapeutic intervention might prevent or delay the pathological processes that lead to disease onset.

A landmark event in HD is motor diagnosis, which is determined after the standard motor examination scored with the Unified Huntington's Disease Rating Scale (UHDRS)[4]. Motor diagnosis is defined as the highest rating on the UHDRS Diagnostic Confidence Level (DCL) that states the rater is at least 99% confident that the motor abnormalities displayed by the patient are unequivocal signs of HD. Despite the prominent role of motor diagnosis in HD research, there is a reluctance to use it as an endpoint in clinical trials[2]. The reason is that HD is a slow progressing disease and there are a small number of motor diagnosis events for cohorts that are followed over the short periods typical of clinical trials.

We are reaching a time in HD therapy development when preventative clinical trials for mutation carriers who are not yet diagnosed are being planned. It is crucial to have better defined endpoints with this population over time periods that are feasible and cost-effective. The number of events can be increased by considering a secondary variable that is correlated with the definitive event of motor diagnosis, but has a faster rate of change. This alternative approach is known as progression-free survival (PFS), which is widely used in oncology trials[5–7]. PFS is the time elapsed since treatment initiation to the first event of HD progression or motor diagnosis, whichever comes first.

The goal of this study is to examine PFS using data from 7 years of the Track observational study[8–10] and evaluate the extent of reproducibility using data from a

separate independent study. Our hypothesis is that PFS will provide sufficient events for the planning of feasible pre-diagnosis clinical trials and show reasonable generalization.

Methods

Study population

Primary analysis involved HD gene expansion mutation carriers and non-carriers from two phases of the Track study: Track-HD and Track-On. Track-HD is a longitudinal cohort study of pre-diagnosis and post-diagnosis mutation carriers, and healthy non-carriers, with four sites in four countries (CAN, FRA, GBR, NE), and data collected 2008-2011[8–10]. Inclusion criteria were: 18 to 65 years of age; toleration of MRI and biosample collection; absence of major psychiatric disorder or history of significant head injury; not active in an experimental therapeutic trial; and no comorbid medical conditions preventing assessment. Non-carriers were selected from spouses or partners of carriers or non-carrier siblings. Non-carriers were age- and gender-matched to the carriers.

Track-On is a longitudinal cohort study of mutation carriers and non-carriers, most of whom transitioned from Track-HD. 74% of the pre-diagnosis carriers and 64% of the non-carriers transitioned over to Track-On, and data collection spanned 2011-2014. Participants who transitioned over could have a maximum of 7 years of data. Table 1 shows descriptive statistics for key variables at study entry (baseline) by carrier status.

The external validation involved participants from the Cooperative Huntington Observational Research Trial (COHORT)[11]. COHORT is a longitudinal cohort study of HD gene expansion mutation carriers and non-carriers with 38 sites in three countries (AUS, CAN, USA), and data collected 2006-2011. Enrollment was open to people who had tested positive for the HD gene expansion mutation (pre-diagnosis or post-diagnosis), or people who were untested but had a family history of HD. Non-carriers were family members verified by genetic testing to not have the expansion mutation. Only confirmed pre-diagnosis mutation carriers were considered for the validation analysis. Table 1 shows descriptive statistics for the COHORT sample along with

statistical comparisons to the Track carrier group. It is noted that the COHORT sample is significantly more progressed at baseline, as indicated by the clinical measures.

Standard protocol approvals, registrations, and patient consents

Study activities were reviewed and approved by local ethics committees (for Track) and institutional review boards (for COHORT). Informed consent according to the Declaration of Helsinki was obtained for each participant, including consent for the distribution of de-identified data for research purposes.

Measures

PFS requires a definitive endpoint, which is motor diagnosis. PFS also requires at least one secondary variable that is correlated with motor diagnosis, but potentially has a faster rate of change. The secondary variables considered here are the UHDRS total motor score (TMS), total functional capacity (TFC), symbol digit modalities test (SDMT), and Stroop word test (SWT).

Endpoints

Two types of survival endpoints are examined. The endpoint for traditional survival analysis is motor diagnosis (DCL = 4), and time to first occurrence is analyzed. PFS is a type of composite event that is triggered by either a progression event or motor diagnosis, whichever comes first. Thus, the time to the composite event is used for the PFS survival analysis.

The progression event is a change from baseline of sufficient size to be deemed important, referred to as clinically meaningful change (CMC)[13]. Statistical methods for CMC estimation are described below. A progression event occurs for an individual if the change on the secondary variable meets or exceeds the CMC. For example, consider CMC = 3 for the TMS, and an individual has four time points with TMS = 8, 9, 12, 13, at baseline (0), 1, 2, and 3 years of follow-up. The TMS difference from baseline is 0, 1, 4, 5, and we assume that DCL = 4 does not occur. The third and fourth values meet the criterion for a progression event, and time to the PFS event is 2 years because it is the

first instance of the progression event (time points after the event are ignored). PFS requires the composite event to occur after baseline; if there is no composite event over the observed epoch, the individual is considered right-censored, meaning the composite event happens sometime in the future.

Statistical Analysis

Clinically Meaningful Change (CMC)

Track data is used to develop the CMC. In order for the CMC to accurately reflect disease effects as opposed to normal aging effects, our approach is to use an extreme score based on analysis with only the non-carriers. Change from baseline is computed for each follow-up time, which represents change due to natural aging (though there are practice effects for SDMT and SWT). Extremes of these changes are computed based on quantile mixed models for follow-up after baseline. The quantile mixed model accounts for the dependency due to repeated measures, but unlike traditional mixed models, a quantile is estimated as a function of time rather than the mean[14]. Change from baseline is regressed onto time on study, age at baseline, and their interaction. The 99th quantile is estimated for TMS because an increase from baseline indicates greater HD progression. Conversely, the 1st quantile is estimated for TFC, SDMT, and SWT because a decrease from baseline indicates greater progression.

External Validation

To assess the replicability (generalizability) of the endpoints, the CMC developed in Track is applied to the COHORT data. Mutation carriers from both studies are combined, study membership is coded, and then the survival profiles of the studies are compared using two statistical methods. The first method is the Wald test of study difference (z-test) using Cox regression, and the second method is the likelihood ratio test of study difference based on smoothed cubic spline survival models[15]. The null hypothesis for both tests is that the survival curves of the studies are equivalent, possibly adjusting for covariates. Thus, evidence of the reproducibility of the PFS survival curves based on the Track CMC will be provided if there is no statistically significant study difference (significance is defined as $p < 0.05$ for all results). Spline

modeling has the advantage of providing smooth survival curves that are not unduly affected by final times being event times (see Figure 1). Study differences are examined without and with adjustment for covariates. The covariates are all baseline variables that show a significant study difference from Table 1, except for follow-up.

Required Sample Size

A popular test of the equivalence of survival curves among groups is the log-rank test[16]. Sample size can be estimated from standard formulas when testing the null hypothesis of equivalent survival curves, under the assumptions of proportional hazards and exponentially distributed survival times[17]. Sample size estimates for the log-rank test require an estimate of the survival proportion at study end. Survival proportions are estimated based on the cubic spline survival curves. Sample sizes are estimated using the conventional type I error rate of 5% (two-tailed test), and type II error rate of 20% (power = 80%). Estimates are for a 3-year two-arm parallel trial with equal group sample size. In order to allow for attrition, the total sample size, N, is adjusted by the dropout rate, w. The adjustment for dropout is $N_w = N/(1-w)$, where $w = 0,\ 0.10,\ 0.20$.

To provide a benchmark for judging the performance of the non-parametric log-rank test, sample size is also estimated using the two-group Mann-Whitney-Wilcoxon (MWW) test of TMS at only the last time point. The MWW is a test of difference in group TMS medians when the group distributions are identical except for a location shift[18]. When the assumptions of the MWW test hold, it is more efficient than the log-rank test and will yield a smaller sample size[19].

The treatment effect size is defined as the hypothetical proportion reduction (π) in the treatment TMS mean (μ_T) relative to the placebo mean (μ_P) at the study terminus; $\pi = |(\mu_P - \mu_T)|/\mu_P$. The Track pre-diagnosis mutation carriers are treated as a proxy for the placebo group, and the hypothetical improvement in the treatment TMS mean is computed as $\mu_T = \mu_P(1-\pi)$. The quantity π is related to Cohen's d as $d = \mu_P(\pi)/\sigma$, where σ^2 is the common group variance. For the Track data, the 3 year visit mean and SD is estimated to be $\hat{\mu}_P = 6.59$ and $\hat{\sigma} = 5.86$, so that a 50% TMS mean reduction produces $d = 6.59(0.50)/5.86 = 0.56$. Cohen's d is not appropriate for the non-

parametric log-rank test and MWW test. Therefore, Cohen's d is transformed into the area under the curve (AUC) of the ROC curve in medical diagnostic testing[20]. AUC is a non-parametric effect size on the 0-1 scale and has a convenient probability interpretation for the log-rank test. AUC is the probability that a randomly sampled patient from the treatment group will delay a HD progression event longer than a randomly sampled patient from the placebo group[19]. There is no treatment effect when $AUC = 0.50$ because there is an equal chance of longer delay for both groups; it is only when $AUC > 0.50$ that we have longer delay for the treatment group. For the aforementioned 50% TMS mean reduction (with $d = 0.56$), the effect size is $AUC = 0.65$.

Results

Results of the CMC analysis using the Track non-carrier group are shown in Table 2. The table shows the point estimates and 95% bootstrap confidence intervals (CIs) for select ages. Because the UHDRS variables take only integer values, the CMC point estimates can be rounded up ignoring sign. For example, the CMC for TMS is 3 for age 30 and 35, and 4 for age 40-50, etc. The CMC for TFC is -1 (loss of 1) for age 35 and older.

Results of the study survival curve comparison are presented in Table 3. Kaplan-Meier curves and fitted spline curves without covariate adjustment are shown in Figure 1. Kaplan-Meier probabilities descend to 0 for some curves because the final time is an event time (spline curves are unaffected by this occurrence). Figure 1 shows that individuals from COHORT have a greater risk of an event than those from Track, with the TMS curves being most similar. Likewise, the upper portion of Table 3 indicates significant study survival curve differences without covariate adjustment for all endpoints except TMS. The bottom portion of Table 3 shows that after adjusting for baseline variables, the study differences are no longer significant, except for SDMT (which is not considered further for this reason).

Estimated total sample size is shown in Table 4 for a hypothetical 3-year parallel arm clinical trial with 10% attrition. As expected, traditional survival analysis based on DCL

has the largest estimated sample size. PFS with TMS progression shows substantially lower sample sizes than DCL alone, being almost 4 times smaller for most effect sizes. However, PFS based on TMS progression has sample sizes that are approximately 1.5 times larger than the smallest possible sizes of the MWW test. Combining TMS with SWT progression lowers the sample sizes to the point of being only about 1.3 times larger than the MWW test.

Discussion

The goal of this study was to define clinically meaningful change (CMC) and a HD progression event for use in preventative trials with HD gene expansion mutation carriers prior to motor diagnosis. The Track study is favorable for CMC analysis as it was conducted like a clinical trial, but without an active treatment group[8–10]. CMC developed in Track was used to define progression-free survival (PFS) endpoints, and for the most part, the survival curves were found to be similar for the COHORT data, especially after adjusting for progression differences at study entry. Therefore, the CMC values developed here appear to be reasonable general indexes for defining HD progression events in clinical trial planning.

The required sample size for a clinical trial can be greatly reduced when a TMS progression event is used in combination with motor diagnosis, which is consistent with our explicit hypothesis. Motor diagnosis based on the DCL is perhaps the closest the field has to a gold standard for a landmark progression event in HD. As expected, survival analysis with time to motor diagnosis yielded the largest required sample size, whereas the PFS endpoints offered substantial reduction. If one wants to retain the definitive outcome of a motor diagnosis while providing enough HD progression events for a reasonably sized trial, then PFS can help. TMS progression, for example, requires approximately 4 times fewer participants for the range of effect sizes considered. It is notable that PFS is approved by the FDA as a surrogate endpoint for cancer trials[21]. Such endorsement is an encouraging sign that PFS might eventually be successful in pivotal trials of HD.

The potential advantage of PFS is that it incorporates motor diagnosis and produces enough events to increase the feasibility of pre-diagnosis trials. This is not to say that PFS necessarily yields smaller sample sizes than traditional methods. PFS involves dichotomizing continuously measured UHDRS variables, which reduces information, and can lead to lower efficiency[22]. Our results show that though PFS can drastically lower the sample size as compared to using motor diagnosis alone, the sample size is still larger than a method that compares the TMS among groups at the last visit. Therefore, if efficiency is the only criterion for choosing an endpoint, then PFS may be less attractive than traditional methods.

The primary appeal of PFS is that it is anchored to the event of motor diagnosis, which is a well-established landmark in the progression of HD. PFS may also be desirable because its effect size can be expressed as a probabilistic statement of potential patient benefit. For instance, AUC $= 0.60$ means that a given person in the treatment group has a 60-40 chance of delaying a HD progression event relative to a person in the untreated group. HD prevention therapy will likely be demanding for participants (e.g., repeated lumbar punctures), and treatment may only be desirable to pursue if the chance of benefit is sufficient in the minds of both the participants and researchers. The AUC effect size provides a clear means of articulating a minimum potential benefit.

The question remains as to what minimum AUC is acceptable. Pilot data can provide an indication of effect sizes that are in reach, or HD stakeholders can decide on a minimum probability. We offer the opinion that AUC values smaller than 0.60 seem to get uncomfortably close to a 50-50 chance of delay, which is just a coin flip's chance. Furthermore, our results indicate that AUC $= 0.60$ requires total $N = 364$ for the TMS endpoint in a two-arm study with 10% dropout. When AUC is less than 0.57, more than 500 participants total will be required. The feasibility of a particular sample size depends on many factors, but we feel that planning for a study with no greater than 400 participants will increase the likelihood of a trial being conducted. The lower bound of AUC $= 0.60$ is consistent with this goal.

The PFS endpoint is a composite by definition. There is recent increased interest in composite endpoints because of the potential of smaller studies and lower costs[25]. An

advantage of PFS is that the composite is defined in a clear manner, not requiring weights for combining variables. Because key variables are collected at the standard UHDRS exam, it is feasible to use such composite endpoints at little cost. The caveat here, as with any composite endpoint, is a potential lack of clarity regarding the nature of the effects. PFS will not distinguish an individual who has a TMS progression event from another who has a SWT event, for example. Assuming a treatment has a benefit, the benefit must be reported in terms of potentially delaying the package of HD progression events. Perhaps PFS using TMS is most clear, as the composite focuses only on motor signs and a treatment benefit can be expressed broadly as a delay in HD motor progression.

A particular CMC threshold does not necessarily speak to the importance of the progression event in the experience of a mutation carrier. In clinical trials where PFS is used, it is typical to have an expert panel convened to confirm the appropriateness of the CMC as the basis for a progression event. For example, panels of oncology experts have been convened to determine the criteria for a solid tumor increase to define a tumor progression event, resulting in the Response Evaluation Criteria in Solid Tumors (RECIST) guidelines for cancer trials[26,27]. Our CMC results are not meant to substitute for an expert panel; they are offered as data-based results that might be informative to such a panel.

The focus of this study is on preventative trials involving confirmed mutation expansion carriers who have not yet reached motor diagnosis. It is possible that a mutation carrier might not have a motor diagnosis at pre-screening, but have a diagnosis at the first visit. This is problematic because such a person must be analyzed according to the intent-to-treat principle, even though they have a 0 event time that is uninformative. An approach to address this problem is to disregard DCL status in the analysis and only use time to the TMS progression event as the outcome. The TMS progression event is defined relative to study entry, so there will always be an analyzable event time. Analysis with this modified endpoint is called time-to-progression analysis[5]. Practically speaking, there is no difference in the number of events for time-to-progression and PFS in our analysis (results not presented), because every individual who is eventually assigned a

motor diagnosis has an earlier TMS progression event. Thus, motor diagnosis does not contribute to the event status of PFS, and it is equivalent to time-to-progression. The drawback however, is that the method does not have the definitive endpoint of motor diagnosis (motor diagnosis is ignored).

The finding that the mutation carriers of COHORT were more progressed than those of Track at study entry is likely due to differences in recruitment strategy. Track-HD explicitly aimed to recruit pre-diagnosis carriers who were relatively far from motor diagnosis, whereas COHORT did not. The external validation considered mutation carriers in both studies who were willing to undergo genetic testing. We provide evidence that the participants from the two studies are similar after adjusting for baseline progression. But it is unknown if our results generalize to the broader HD population because most at-risk individuals do not undergo genetic testing[28].

In conclusion, minimum values are proposed for assessing clinically meaningful change over time for HD gene expanded mutation carriers who have not yet received a motor diagnosis. The change values can be used to define progression events that are easy to combine, yield trials of reasonable size, and may apply across studies. The approach is especially appealing when a researcher wants to examine whether a treatment delays motor onset.

Acknowledgments

Track-HD was supported by the CHDI Foundation, a not-for-profit organisation dedicated to finding treatments for Huntington's disease. We thank the Track-HD study participants and their families. We also acknowledge the support of the National Institute for Health Research University College London Hospitals Biomedical Research Centre and the Manchester Biomedical Research Centre.

Samples and/or data from the COHORT Study, which received support from HP Therapeutics, Inc., were used in this study. We thank the Huntington Study Group COHORT investigators and coordinators who collected data and/or samples used in this study, as well as participants and their families, who made this work possible.

The Track-HD investigators are: Australia—C Campbell, M Campbell, E Frajman, I Labuschagne, C Milchman, A O'Regan, J Stout (Monash University, Victoria). Canada—A Coleman, R Dar Santos, J Decolongon, A Sturrock (University of British Columbia, Vancouver). France—E Bardinet, C Jauffret, D Justo, S Lehericy, C Marelli, K Nigaud, R Valabrègue (APHP, Hôpital Salpêtriere, Paris). Germany—N Bechtel, S Bohlen, R Reilmann (University of Münster, Münster); B Landwehrmeyer (University of Ulm, Ulm); A Hoffman, P Kraus (University of Bochum, Bochum). Netherlands—SJA van den Bogaard, E M Dumas, J van der Grond, EP t'Hart, C Jurgens, M-N Witjes-Ane (Leiden University Medical Centre, Leiden). UK—N Arran, J Callaghan, D Craufurd, C Stopford (St Mary's Hospital, Manchester); C Frost, R Jones (London School of Hygiene and Tropical Medicine, London); H Crawford, N C Fox, C Gibbard, N Hobbs, N Lahiri, I Malone, R Ordidge, G Owen, T Pepple, J Read, M J Say, D Whitehead (University College London, London); S Keenan (Imperial College London, London); D M Cash (IXICO, London); C Berna, S Hicks, C Kennard (University of Oxford, Oxford). USA— T Acharya, E Axelson, H Johnson, C Wang, (University of Iowa, Iowa City, IA); B Borowsky (CHDI Foundation, NY); S Lee, W Monaco (Massachusetts General Hospital, Harvard, MA); C Campbell, S Queller, K Whitlock (Indiana University, IN).

The Track-On investigators are: Australia—I Labuschagne (Monash University, Victoria); Canada—A Coleman, J Decolongon, M Fan, T Petkau (University of British

Columbia, Vancouver). France—C Jauffret, D Justo, S Lehericy, K Nigaud, R Valabrègue (ICM and APHP, Hôpital Salpêtriere, Paris). Germany—N Weber (George Huntington Institute, Münster); I Mayer, M Orth (University of Ulm, Ulm). Netherlands—A Schoonderbeek, EP t'Hart, (Leiden University Medical Centre, Leiden). UK—D Craufurd (St Mary's Hospital, Manchester); A Cassidy, C Frost, R Keogh (London School of Hygiene and Tropical Medicine, London); C Berna, H Crawford, M Desikan, R Ghosh, D Hensman Moss, E Johnson, P McColgan, G Owen, M Papoutsi, J Read, A Razi, D Mahaleskshmi (University College London, London). USA—H Johnson (University of Iowa, Iowa City, IA); B Borowsky (CHDI Foundation, NY).

Financial Disclosures

Jeffrey D. Long has a consulting agreement with NeuroPhage, is a paid consultant for Roche Pharma and Azevan Pharmaceuticals, and receives funding from CHDI, Michael J. Fox Foundation, and NIH.

James A. Mills has no financial disclosures.

Blair R. Leavitt reports research grants from Teva, CHDI, and Lifemax. Personal Consulting fees from Roche, uniQure, Novartis, Lifemax and Raptor.

Alexandra Durr received research grants from the French Agency for Research, French Ministry for Social affairs and Health, Pfizer Inc. and Annapurna Therapeutics. She holds partly a patent B 06291873.5 on Anaplerotic therapy of Huntington’s disease and other polyglutamine diseases.

Raymund A. Roos received grants from The Gossweiler Foundation and TEVA, both via the LUMC administration. He is an advisor for UniQure.

Julie C. Stout has served on an advisory board for Roche, has consulted to Prana Biotechnology, is treasurer and board member for the Huntington's Study Group, Inc., and conducts business implementing cognitive assessments at Stout Neuropsych Pty Ltd, with contracts from Teva, Vaccinex, Omeros, and Ionis.

Ralf Reilmann is founding director and owner of the George-Huntington-Institute, a private research institute focused on clinical and preclinical research in Huntington’s

disease, and QuantiMedis, a clinical research organization providing Q-Motor (quantitative motor) services in clinical trials and research. He holds appointments at the Dept. of Radiology of the University of Muenster and at the Department of Neurodegenerative Diseases and Hertie-Institute for Clinical Brain Research, University of Tuebingen. He provided consulting services, advisory board functions, clinical trial services, quantitative motor analyses, and/or lectures for Teva, Pfizer, uniQure, Ipsen, Vaccinex, Novartis, Raptor, Omeros, Siena Biotech, Neurosearch Inc., Lundbeck, Medivation, Wyeth, ISIS Pharma, Link Medicine, Prana Biotechnology, MEDA Pharma, Temmler Pharma, Desitin, AOP Orphan, and the Cure Huntington's Disease Initiative Foundation. He received grant support from the High-Q-Foundation, the Cure Huntington's Disease Initiative Foundation (CHDI), the Deutsche Forschungsgemeinschaft (DFG), the European Union (EU-FP7 program), the Bundesministerium fur Bildung und Forschung (BMBF), the Deutsches Zentrum fur Neurodegeneration und Entzundung (DZNE), and the European Huntington's Disease Network (EHDN).

Bernhard Landwehrmeyer has provided consulting services, advisory board functions, clinical trial services and/or lectures for Allergan, Alnylam, Aventis, Amarin, AOP Orphan Pharmaceuticals AG, Bayer Pharma AG, Desitin, Genzyme, GlaxoSmithKline, Hoffmann-La Roche, Ipsen, ISIS Pharma, Lundbeck, Neurosearch Inc, Medesis, Medivation, Medtronic, Novartis, Pfizer, Prana Biotechnology, Raptor Pharmaceuticals, Sangamo/Shire, Sanofi-Aventis, Siena Biotech, Temmler Pharma GmbH, Trophos, and TEVA and has received research grant support from the CHDI Foundation, the Bundesministerium fur Bildung und Forschung (BMBF), the Deutsche Forschungsgemeinschaft (DFG), the European Commission (EU-FP7). His study site Ulm has received compensation in the context of the observational REGISTRY-Study of European Huntington's Disease Network (EHDN).

Sarah Gregory has no financial disclosures.

Rachael I. Scahill has no financial disclosures.

Douglas R. Langbehn received research funding from CHDI, from the Neurology Institute of the University college of London, and from NIH/NIDA via grant 5R01AA021165-03. He is also a paid statistical consultant for Roche and Voyager pharmaceutical companies for Huntington's disease clinical trial design.

Sarah J. Tabrizi has served on advisory boards or had consultancies with F. Hoffmann-La Roche Ltd, Ionis Pharmaceuticals, Ixico Technologies, Shire Human Genetic Therapies, Takeda Pharmaceuticals International and TEVA Pharmaceuticals. All honoraria paid for these consultancies and advisory boards go to UCL, Dr. Tabrizi's employer.

References

1. Ross CA, Tabrizi SJ. Huntington's disease: From molecular pathogenesis to clinical treatment. Lancet Neurology. 2011;10:83–98.

2. Huntington Study Group PHAROS Investigators. Clinical-genetic associations in the prospective Huntington at risk observational study (PHAROS). JAMA Neurology. 2016;73:102–110.

3. Ross CA, Aylward EH, Wild EJ, et al. Huntington disease: Natural history, biomarkers and prospects for therapeutics. Nature Reviews Neurology. 2014;10:204–216.

4. Huntington Study Group. Unified Huntington's Disease Rating Scale: Reliability and-consistency. Movement Disorders. 1996;11:136–142.

5. Booth CM, Eisenhauer EA. Progression-free survival: Meaningful or simply measureable? Journal of Clinical Oncology. 2012;30:1030–1033.

6. Saad ED, A. K. Progression-free survival and time to progression as primary endpoints in advanced breast cancer: Often used, sometimes loosely defined. Annals of Oncology. 2009;20:460–464.

7. Soria JC, Massard C, Le Chevalier T. Should progression-free survival be the primary measure of efficacy for advanced NSCLC therapy? Annals of Oncology. 2010;21:2324–2332.

8. 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 Neurology. 2009;8:791–801.

9. Tabrizi S, Scahill R, Durr A, et al. Biological and clinical changes in premanifest and early stage Huntington's disease in the Track-HD study: The 12-month longitudinal analysis. Lancet Neurology. 2011;10:31–42.

10. Tabrizi SJ, Scahill RI, Owen G, et al. Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the Track-HD study: Analysis of 36-month observational data. Lancet Neurology. 2013;12:637–649.

11. Dorsey ER, the Huntington Study Group COHORT Investigators. Characterization of a large group of individuals with huntington disease and their relatives enrolled in the COHORT study. PLoS ONE. 2012;7:1–10.

12. Marder K, Zhao H, Myers RH, et al. Rate of functional decline in Huntington's disease. Neurology. 2000;54:452–458.

13. Crosby RD, L. KR, R. WG. Defining clinically meaningful change in health-related quality of life. Journal of Clinical Epidemilogy. 2003;56:395–407.

14. Geraci M, Bottai M. Linear quantile mixed models. Statistics and Computing. 2014;24:461–479.

15. Royston P, Parmar MKB. Restricted mean survival time: An alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Medical Research Methodology. 2013;13:152–162.

16. Cox DR, Oakes D. Analysis of survival data. London: Chapman; Hall; 1984.

17. Machin D, Campbell MJ, Tan SB, Tan SH. Sample size tables for clinical studies. Hoboken, NJ: Wiley-Blackwell; 2009.

18. Noether GE. Sample size determination for some common nonparametric tests. Journal of the American Statistical Association. 1987;82:645–647.

19. Rahardja D, Zhao YD. Unified sample size calculations using the competing probability. Statistics in Biopharmaceutical Research. 2009;1:323–327.

20. Acion L, Peterson JJ, Temple S, Arndt S. Probabilistic index: An intuitive non-parametric approach to measuring the size of treatment effects. Statistics in Medicine. 2008;25:591–602.

21. FDA. Guidance for industry: Clinical trial endpoints for the approval of cancer drugs and biologics, clinical-medical guidances. [online]. U.S. Department of Health; Human Services; 2007. Accessed at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM259421.pdf.

22. Wason JMS, Dentamaro A, Eisen TG. The power of phase II end-points for different possible mechanisms of action of an experimental treatment. European Journal of Cancer. 2015;51:984–992.

23. Long JD, Paulsen JS, Marder K, Zhang Y, Kim J, Mills JA. Tracking motor impairments in the progression of Huntington's disease. Movement Disorders. 2014;29:311–319.

24. Paulsen JS, Smith MM, Long JD. Cognitive decline in prodromal Huntington: Disease: Implications for clinical trials. Journal of Neurology, Neurosurgery, and Psychiatry. 2013;84:1233–1239.

25. Tomlinson G, Detsky AS. Composite end points in randomized trials: There is no free lunch. Journal of the American Medical Association. 2010;303:267–268.

26. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). European Journal of Cancer. 2009;45:228–247.

27. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. Journal of the National Cancer Institute. 2000;92:205–216.

28. Shoulson I, Young AB. Milestones in Huntington's disease. Movement Disorders. 2011;26:1127–1133.

Table 1. Descriptive statistics for variables measured at study entry (baseline). Mean (SD) for quantitative variables and count (percentage) for categorical variables. Statistical tests are for CAG expanded study participants (Track vs. COHORT).

	Track Non-Carriers	**Track Mutation Carriers**	**COHORT Mutation Carriers**	**Statistic[a]**	**p-value**
N	156	167	366		
Female	87 (56%)	93 (56%)	229 (63%)	-1.50	0.132
Age	45.58 (10.30)	40.06 (8.92)	42.21 (12.48)	-2.00	0.046
CAG Expansion		43.22 (2.42)	42.37 (2.80)	3.35	<.001
Total Motor Score	1.42 (1.65)	3.01 (2.28)	6.23 (8.57)	-4.78	<.001
Total Functional Capacity	12.99 (0.11)	12.86 (0.51)	12.36 (1.54)	4.13	<.001
Symbol Digit Modalities Test	52.49 (9.46)	51.66 (10.17)	44.99 (12.17)	6.17	<.001
Stroop Word Test	105.35 (16.95)	99.95 (16.86)	90.76 (19.89)	5.18	<.001
Follow-Up (Years)	4.00 (2.19)	4.22 (2.11)	2.27 (1.02)	14.41	<.001

Note. [a]*z*-test for female and *t*-test (*df* = 531) for all others; CAG = cytosine-adenine-guanine.

Table 2. Clinically meaningful difference point estimate (95% CI) for a hypothetical 3-year trial. Each point estimate is the extreme quantile (1% or 99%) based on the Track non-mutation carriers (controls).

	Age								
Variable	30	35	40	45	50	55	60	65	70
Total Motor Score	2.30 [1.35,3.21]	2.72 [1.64,3.77]	3.15 [1.95,4.31]	3.57 [2.23,4.89]	4.00 [2.46,5.51]	4.43 [2.72,6.11]	4.85 [2.95,6.72]	5.28 [3.20,7.32]	5.70 [3.48,7.92]
Total Functional Capacity	-0.00 [-0.04,0.01]	-0.02 [-0.06,-0.01]	-0.03 [-0.09,-0.01]	-0.04 [-0.12,-0.01]	-0.06 [-0.15,-0.01]	-0.07 [-0.19,-0.02]	-0.08 [-0.22,-0.01]	-0.10 [-0.25,-0.02]	-0.11 [-0.28,-0.02]
Symbol Digit Modalities Test	-1.78 [-3.33,-0.11]	-3.70 [-5.05,-2.16]	-5.62 [-7.32,-4.10]	-7.54 [-9.69,-5.78]	-9.46 [-12.12,-7.17]	-11.38 [-14.67,-8.54]	-13.30 [-17.32,-9.88]	-15.21 [-19.93,-11.17]	-17.13 [-22.59,-12.44]
Stroop Word Test	-10.39 [-17.04,-6.10]	-13.28 [-20.11,-8.85]	-16.16 [-23.18,-11.38]	-19.05 [-26.86,-13.49]	-21.94 [-30.54,-15.60]	-24.83 [-34.07,-17.60]	-27.72 [-37.73,-19.72]	-30.61 [-41.49,-21.64]	-33.50 [-45.25,-23.46]

Table 3. Training and test study survival comparisons for various endpoints. Upper portion shows the results for study difference not adjusting for covariates, and the lower portion shows study difference adjusting for covariates.

	Cox Model					Spline Model	
Endpoint	**B**	**exp(B)**	**SE**	**Z**	**p**	**Chisq**	**p**
Study Difference Only							
Diagnostic Confidence Level (DCL)	-0.79	0.45	0.22	-3.63	<.001	16.60	<.001
Total Motor Score or DCL	-0.15	0.86	0.13	-1.10	0.269	2.17	0.141
Total Functional Capacity or DCL	-0.71	0.49	0.18	-3.97	<.001	21.03	<.001
Symbol Digit Modalities Test or DCL	-0.55	0.58	0.16	-3.53	<.001	17.26	<.001
Stroop Word Test or DCL	-0.62	0.54	0.18	-3.51	<.001	15.13	<.001
Study Difference + Covariates							
Diagnostic Confidence Level (DCL)	-0.11	0.90	0.25	-0.43	0.671	0.43	0.513
Total Motor Score or DCL	0.07	1.07	0.15	0.49	0.621	0.06	0.811
Total Functional Capacity or DCL	-0.29	0.75	0.20	-1.48	0.139	3.56	0.059
Symbol Digit Modalities Test or DCL	-0.45	0.64	0.17	-2.66	0.008	10.23	0.001
Stroop Word Test or DCL	-0.34	0.72	0.19	-1.72	0.085	3.82	0.051

Note. Covariates are all significant variables ($p < .05$) from Table 1; Chisq has *df* = 1.

Table 4. Required total sample size (*N*) for a 3-year, two-arm parallel trial, as a function of trial condition and endpoint. Multiple response variables denote the composite event of progression-free survival. Calculations are based on equal group allocation, type I error rate = 5%, power = 80%, and 10% dropout.

		Sample Size						
Condition		Survival Analysis	Progression-Free Survival Analysis					MWW Test
TX Effect	AUC	DCL	TMS or DCL	TFC or DCL	SDMT or DCL	SWT or DCL	TMS or SWT or DCL	Median
25%	0.579	2376	646	1450	960	1381	573	468
30%	0.594	1689	457	1030	681	980	406	327
35%	0.610	1269	342	773	511	737	302	242
40%	0.625	993	267	604	399	576	236	187
45%	0.640	801	214	488	321	463	189	149
50%	0.655	661	177	402	266	383	156	122
55%	0.669	557	149	339	223	322	131	102
60%	0.683	477	127	290	191	276	112	87
65%	0.697	412	110	251	166	239	97	74
70%	0.711	361	96	220	144	209	84	66
75%	0.725	320	84	194	128	184	74	58

Note. TX = treatment; AUC = area under the curve; DCL = diagnostic confidence level; TMS = total motor score; TFC = total functional capacity; SDMT = symbol digit modalities test; SWT = Stroop word test; MWW = Mann-Whitney-Wilcoxon.

Figure 1. Kaplan-Meier survival curves (stepped) and cubic spline curves (smooth) by follow-up year and study for different survival endpoints. Curves are unadjusted for covariates.