Appendix E-1

Methods

Randomization and Enrollment: The computer-generated randomization plan was prepared by an independent programmer at the University of Rochester and shared with the Clinical Materials Services Unit (CMSU, Rochester, NY) who packaged and labeled the study medication. The randomization scheme used permuted blocks and was stratified by site. Sites enrolled subjects via a secure internet portal that provided the appropriate drug kit numbers. Nobody other than the independent programmer, an independent statistician who interacted in closed session meetings with the DSMB, and CMSU personnel had access to treatment assignments.

- Additional Eligibility Criteria: Inclusion: able to give informed consent and comply with trial procedures (for minor subjects under the age of 18, both parental permission and child assent required); able to take oral medication; identification of an informant or caregiver willing and able to supervise the daily dosing of study medications and to maintain control of study medications in the home, if necessary. Exclusion: creatinine > 2.0 mg/dl, liver function tests greater than 3 times the upper limit of normal, absolute neutrophil count ≤ 1000/µl, platelet concentration < 100,000/µl, hematocrit level < 33 for female or < 35 for male subjects, or coagulation tests > 1.5 times the upper limit of normal; known allergy to the color

additive FD&C #5 yellow lake (also known as tartrazine) or any other ingredient in the study drug.

Study Visits: CoQ was measured using methods described by Tang et al (S1), performed at the Weill Medical College of Cornell University, Department of Neurology/Neuroscience, New York, NY, USA. For visits at which blood was drawn for plasma CoQ analysis, subjects were instructed not to take study medication on the day of the visit until after the sample was drawn. The HD CAG repeat length was determined with a modified PCR amplification assay using a fluorescent oligonucleotide primer pair flanking the repeat for automated allele calling after capillary electrophoresis on an ABI3730XL DNA Analyzer (S2). The set of HD CAG allele standards for this process was determined by DNA sequencing as reported by Perlis et al (S3). CAG analysis was performed by the Molecular Neurogenetics Unit, Massachusetts General Hospital, Boston, MA, USA. After randomization, subjects received the first dose of study drug (one 300 mg wafer) and were monitored for one hour for adverse events. Study drug was then dispensed with instructions about titration.

Subjects were allowed to remain in the study at the discretion of the Investigator if they chose to stop study medication (e.g., for reasons such as intolerability or preference). Subjects being followed off study medication were encouraged to continue having their visits in person at the study center; however, for purposes of retention, subjects could be followed by telephone for follow-up visits. Whenever

possible, the subject's final visit was conducted in person. Activities to be conducted during the modified telephone visits included TFC (by the investigator) and review of concomitant medications.

Modifications of the Dosage of the Study Intervention: Dosage reductions were allowed in the event of intolerability, for which doses would be halved from 4 wafers twice daily to 2 wafers twice daily. Subjects could be completely taken off of study medication but remain in the trial, if willing. Suspensions due to intolerability were also allowed, handled on a case-by-case basis, with provisions for re-challenge of study drug towards the full dosage. Re-challenge was allowed after 48 hours of tolerability on the reduced dosage.

Secondary Outcome Variables: Secondary outcome variables included changes from baseline to month 60 in UHDRS subscale scores (Total Motor score, maximal dystonia score, maximal chorea score, Functional Checklist score, Independence Scale score, total behavioral frequency score, total behavioral frequency × severity score, Symbol Digit Modalities Test, Verbal Fluency Test, and Stroop Interference Test scores) and times to HD milestones (two-point decline in TFC score or death, three-point decline in TFC score or death, TFC score of 6 or less or death, institutionalization or death, and death).

Assumptions Underlying the Sample Size Determination: In the original trial design, the sample size was based on the primary outcome variable of the change

in TFC score from baseline to Month 60. The assumptions for the sample size calculation were based on data from the CARE-HD trial (S4), which suggested that the mean change in TFC score decreased and the standard deviation (SD) of the change increased, both approximately linearly over time, in those not treated with CoQ. Extrapolation of the changes in TFC score over time to month 60 yielded assumptions of a mean (± SD) change of -5.0 ± 3.5 in the placebo group for the present trial. A treatment effect of a 20% slowing of the decline in TFC score over 60 months was considered to be of minimal clinical significance based on a survey of 70 HD clinicians and researchers at the 2003 American Neurological Association meeting. Thus, the CoQ group was assumed to have a mean change of -4.0 over 60 months.

Statistical Analysis: The original primary analysis was to be performed using a repeated measures analysis of covariance model (i.e., the mixed model repeated measures, or MMRM, analysis strategy) (S5), with terms for treatment group, center, baseline TFC score, time (treated as a categorical variable), and interaction terms for baseline TFC score and time and for treatment group and time. The covariance matrix for the within-subject observations was to be modeled using an unstructured pattern. This model would yield adjusted estimates of treatment effect, along with associated confidence intervals and p-values, for each time point, with the Month 60 time point being of primary interest. It would also use maximum likelihood to estimate the model parameters from all available data from all

randomized subjects; thus, it would accommodate missing data under the missing at random (MAR) assumption (S6).

The independent NINDS-appointed Data and Safety Monitoring Board (DSMB), on the basis of the examination of accumulating data on deaths at bi-annual meetings, recommended to the trial leadership that they reconsider how they accommodate missing data from subjects who die in their primary analysis of the change in TFC score, i.e., to consider death as a poor outcome rather than as an ordinary subject withdrawal. The trial leadership decided, without knowledge of the incidence of death by treatment group, to change the primary analysis to that of a joint rank approach. In this analysis, subjects are ranked from worst to best outcome with subjects who die being assigned the worst ranks (and ranked according to the time of death) and subjects who survive being ranked more favorably in the order of the change from baseline to month 60 in TFC score. Special considerations are required to assign ranks for subjects who prematurely withdraw from the trial: pair-wise comparisons between a subject who withdrew and all other subjects who survived up until the time of the withdrawal of the subject in question are made with respect to the change in TFC score at the latest time point at which both subjects have a TFC score recorded. Subjects who withdrew from the trial but were later found to have died prior to their scheduled Month 60 visit were considered to have died for the purposes of the primary analysis. This analysis yields an estimated probability that a randomly selected subject treated with CoQ

has a better outcome than a randomly selected subject treated with placebo, along with its associated 95% confidence interval (CI) and p-value.

The joint rank analysis was applied to the UHDRS functional outcomes (TFC score, Functional Checklist score, and Independence Scale score). These outcome variables were also analyzed using the repeated measures analysis of covariance model described above, but with a value of zero imputed for all visits that occurred after the time of the subject's death. Other UHDRS subscale scores (motor, behavioral, and cognitive) were analyzed using the repeated measures analysis of covariance model described above, treating subjects who died as having withdrawn from the trial.

Cox proportional hazards models were used to compare the treatment groups with respect to the time from randomization to the HD milestones of interest (time to a two-point decline in TFC score or death, time to a three-point decline in TFC score or death, time to a TFC score of 6 or less or death, time to institutionalization or death, time to death). Each model included adjustment for the baseline TFC score. Results are reported as hazard ratios (HR) and associated 95% confidence intervals and p-values.

Data on disposition, dosage, safety, compliance, and CoQ levels are presented descriptively.

Interim Analyses: The DSMB met twice annually to review data on trial performance and subject safety, and also received safety data for review on a quarterly basis. Interim analyses for futility were scheduled to take place after 50% of the subjects had completed 36 months of follow-up (summer of 2012) and again after all subjects had completed 36 months of follow-up (summer of 2015), at which time an interim analysis for efficacy was also scheduled and approximately 70% of subjects would have completed 60 months of follow-up. The NINDS requested an additional interim analysis for futility in the summer of 2011, just prior to renewal of funding for the trial. Shortly after the modification of the primary analysis to accommodate death as a poor outcome, the DSMB requested an unscheduled interim analysis for futility in May, 2014. This revealed a conditional power of < 5% for the primary analysis, and the trial was halted in July, 2014. Only data collected prior to the time that this decision was communicated to the trial investigators (July, 2014) were included in the final statistical analyses.

When the first two interim analyses were performed, no subjects had reached the Month 60 visit and different subjects had different lengths of follow-up. For these reasons, all analyses for futility were performed using simulation of conditional power (S7), i.e., the probability of rejecting the null hypothesis of no treatment effect at the end of the trial, based on the primary outcome variable, given all of the data observed in the trial to date. It was assumed that the vector of outcomes for each subject across all visits had a multivariate normal distribution. Therefore, for an individual subject, the joint distribution of outcomes after the subject's last

observed value, given the subject's observed outcomes, is multivariate normal with a mean and covariance matrix that depends on (1) the vector of true mean responses across all time points and (2) the true covariance matrix among the responses at all time points (S8). Given assumptions concerning these quantities and the observed data (including treatment group) for the subject, the remaining data for this subject were simulated using the appropriate conditional multivariate normal distribution. This process was repeated for each subject who had not already withdrawn from the trial or died. Logistic regression models were then used to simulate withdrawal and death at each time point after the last observed TFC score for each subject. These models were fit for each treatment group separately and included TFC score at the current time point as a predictor of the event of interest at that time point; generalized estimating equations with an unstructured correlation matrix were used to account for the repeated binary outcomes for each subject in the model fitting. A subject who was deemed to have withdrawn or died at a particular time point had their simulated TFC data set to missing at and after that time point. The assumptions concerning the unknown model parameters (means, covariance matrix, logistic regression coefficients) were based on the current observed data and plausible deviations from the current patterns. This entire process was repeated 1000 times in order to estimate the conditional power.

References

E1) Tang PH, Miles MV, DeGrauw A, et.al. HPLC Analysis of Reduced and Oxidized Coenzyme Q10 in Human Plasma. Clinical Chem 2001; 47:2, 256–265.

E2) Warner JP, Barron LH, Brock DJ. A new polymerase chain reaction (PCR) assay for the trinucleotide repeat that is unstable and expanded on Huntington's disease chromosomes. Mol Cell Probes 1993;7:235–239.

E3) Perlis RH, Smoller JW, Mysore J, et al. Prevalence of incompletely penetrant Huntington's disease alleles among individuals with major depressive disorder. Am J Psychiatry 2010;167:574 –579.

E4) The Huntington Study Group. A randomized, placebo-controlled trial of coenzyme Q10 and remacemide in Huntington's Disease. Neurology 2001;57:397-404.

E5) Molenberghs G, Thijs H, Jansen I, et al. Analyzing incomplete longitudinal clinical trial data. Biostatistics 2004; 5:445-464.

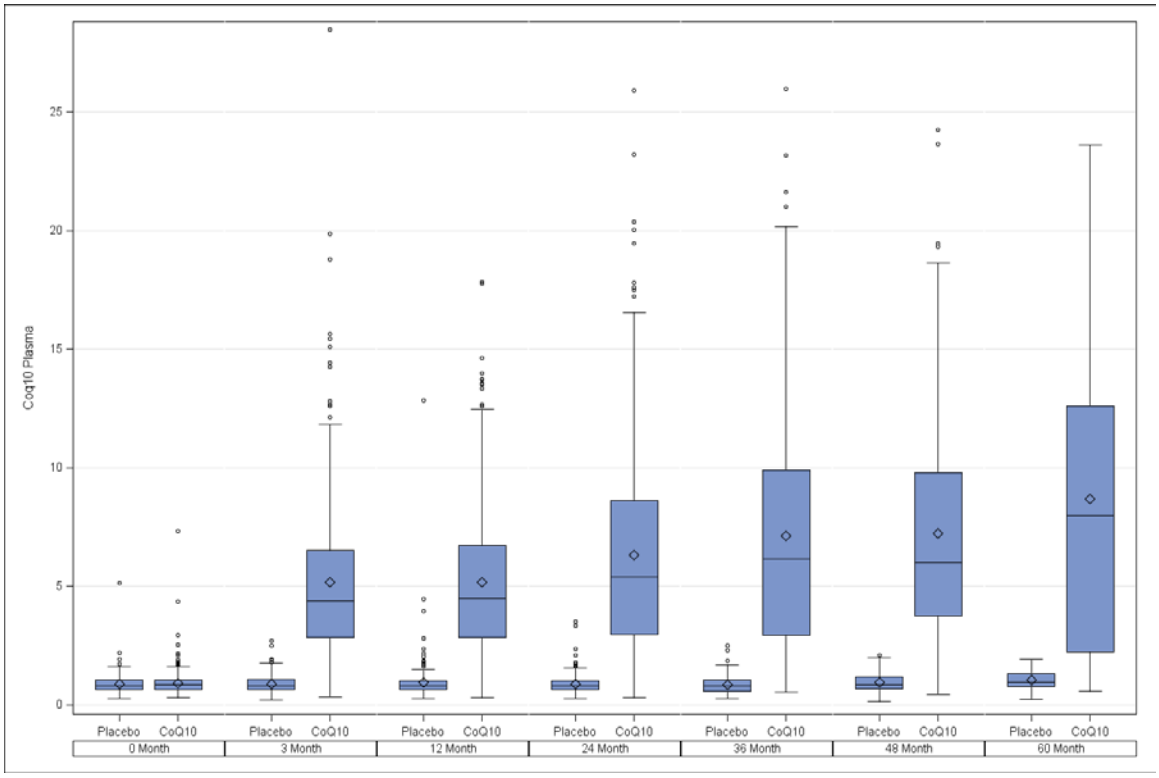
E6) Molenberghs G, Kenward MG. Missing Data in Clinical Studies. Chichester, UK: John Wiley and Sons, 2007).

E7) McDermott MP, Cai X. Interim monitoring of longitudinal outcomes in clinical trials.

Presentation at the Joint Statistical Meetings, Montreal, Canada, 2013.

E8) Anderson TW. An Introduction to Multivariate Statistical Analysis, Second Edition. New York: John Wiley and Sons, 1984.

Figure E-1: Distributions of plasma CoQ levels by treatment group at each visit



The distributions are described using boxplots, where the line in the interior of the box indicates the median CoQ level (µg/mL), the ends of the box indicate the 25th and 75th percentiles, and the exterior lines indicate the range of the data, with the exception of outliers (indicated by individual points) that are 1.5 × IQR larger than the 75th percentile.

IQR = Interquartile range (75th percentile – 25th percentile)

Table E-1. Estimated cumulative event probabilities (Kaplan-Meier estimates) over time by treatment group

Event	Years after Randomization	CoQ (n = 303)	Placebo (n = 306)
TFC decline ≥ 2 points or death	1	33%	35%
	2	57%	59%
	3	73%	72%
	4	80%	76%
	5	90%	84%
TFC decline ≥ 3 points or death	1	18%	18%
	2	37%	41%
	3	56%	57%
	4	65%	65%
	5	77%	76%
TFC score ≤ 6 points or death	1	10%	8%
	2	22%	18%
	3	37%	32%
	4	48%	43%
	5	60%	60%
Institutionalization or death	1	1.4%	1.0%
	2	2.9%	1.7%
	3	6.4%	3.7%
	4	14.2%	8.4%
	5	20.3%	12.3%
Death	1	0.7%	0.7%
	2	1.5%	1.0%
	3	2.6%	1.4%
	4	7.6%	3.4%
	5	11.0%	6.1%

TFC = Total functional capacity