RAPID PUBLICATION

AMERICAN JOURNAL OF medical genetics PART B
Neuropsychiatric Genetics

Indexing Disease Progression at Study Entry With Individuals At-Risk for Huntington Disease

Ying Zhang,[1] Jeffrey D. Long,[2] James A. Mills,[2] John H. Warner,[3] Wenjing Lu,[1] Jane S. Paulsen[2]* and the PREDICT-HD Investigators and Coordinators of the Huntington Study Group

[1]Department of Biostatistics, University of Iowa, Iowa City, Iowa

[2]Department of Psychiatry, University of Iowa, Iowa City, Iowa

[3]CHDI Foundation, Princeton, New Jersey

Received 22 March 2011; Accepted 2 August 2011

The identification of clinical and biological markers of disease in persons at risk for Huntington disease (HD) has increased in efforts to better quantify and characterize the epoch of prodrome prior to clinical diagnosis. Such efforts are critical in the design and implementation of clinical trials for HD so that interventions can occur at a time most likely to increase neuronal survival and maximize daily functioning. A prime consideration in the examination of prodromal individuals is their proximity to diagnosis. It is necessary to quantify proximity so that individual differences in key marker variables can be properly interpreted. We take a data-driven approach to develop an index that can be viewed as a proxy for time to HD diagnosis known as the CAG-Age Product Scaled or CAP_S. CAP_S is an observed utility variable computed for all genetically at-risk individuals based on age at study entry and CAG repeat length. Results of a longitudinal receiver operating characteristic (ROC) analysis showed that CAP_S had a relatively strong ability to predict individuals who became diagnosed, especially in the first 2 years. Bootstrap validation provided evidence that CAP_S computed on a new sample from the same population could have similar discriminatory power. Cutoffs for the empirical CAP_S distribution can be used to create a classification for mutation-positive individuals (*Low–Med–High*), which is, useful for comparison with the naturally occurring mutation-negative *Control* group. The classification is an improvement over the one currently in use as it is based on observed data rather than model-based estimated values.

Key words: survival analysis; prodromal Huntington disease; PREDICT-HD study

How to Cite this Article:
Zhang Y, Long JD, Mills JA, Warner JH, Lu W, Paulsen JS, the PREDICT-HD Investigators, Coordinators of the Huntington Study Group. 2011. Indexing Disease Progression at Study Entry With Individuals At-Risk for Huntington Disease.
Am J Med Genet Part B.

INTRODUCTION

Huntington disease (HD) is an autosomal dominant illness of the brain caused by the trinucleotide cytosine–adenine–guanine (CAG) expansion in the gene of the protein huntingtin. People affected with HD, known as mutation-positive individuals, have a CAG repeat length between 36 and 250 [Kremer et al., 1994; Brinkman et al., 1997; Bruland et al., 1999; Nance et al., 1999]. There is an inverse relationship between CAG repeat length and age of HD diagnosis with longer lengths being associated with earlier diagnosis. Various statistical models have been developed to account for this phenomenon [Andrew et al., 1993; Stine et al., 1993; Lucotte et al., 1995; Rubinsztein et al., 1996, 1997; Brinkman et al., 1997; Squitieri et al., 2000; Gutiérrez and MacDonald, 2002, 2004; Maat-Kievit et al., 2002; Langbehn et al., 2004, 2010; Langbehn and Paulsen, 2007; Andresen et al., 2007a,b].

The PREDICT-HD study is an ongoing observational study including 32 sites from the United States, Canada, Australia, Germany, Spain, and the United Kingdom [Paulsen et al., 2006]. Recruits consist of a large number of participants who have undergone genetic testing for the HD gene mutation, but were not clinically diagnosed with the disease at the time they entered the study. These at-risk participants for HD are described as "prodromal" because they are mutation-positive (CAG repeat length $\geq$36) and show evidence of disease progression based on several key clinical and biological markers [Paulsen et al., 2008,

Grant sponsor: National Institutes for Health; Grant sponsor: National Institute of Neurological Disorders and Stroke; Grant number: NS40068; Grant sponsor: CHDI Foundation, Inc..
*Correspondence to:
Prof. Jane S. Paulsen, Ph.D., The University of Iowa, Roy J. and Lucille A. Carver College of Medicine, Psychiatry Research, 1-305 MEB, Iowa City, IA 52242-1000. E-mail: jane-paulsen@uiowa.edu
Published online 00 month 2011 in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/ajmg.b.31232

2010a; Paulsen, 2010]. PREDICT-HD is a prevalent cohort design as it involves cross-sectional sampling criteria for study inclusion. Participants must be adults ($\geq$18 years), and those who are mutation-positive for HD must not show manifest signs at study entry as determined by the absence of a diagnosis.

One goal of prodromal research is to better characterize the natural history of HD so that preventive intervention can be initiated at the time best suited to maximize success. Previous research using various clinical and biological markers suggests that the prodromal phase will include an insidious decline followed by a precipitous deterioration about 12 years prior to clinical diagnosis [Campodonico et al., 1996; Paulsen et al., 2006, 2008]. Since the specific epoch of time considered during the relatively long prodrome of HD appears to vary, methods to classify prodromal individuals are essential to understanding the success or failure of interventions.

The prevalent cohort design of PREDICT-HD presents a challenge for statistical analysis as inferences can be biased if certain study characteristics are ignored. Individuals with HD are mutation-positive at birth, and this time might be considered the initiating event and the natural time origin for studying disease progression. A problem with using birth as the time origin is that a prevalent cohort design like PREDICT-HD has *length bias* [Zelen and Feinleib, 1969; Zelen, 2005]. Individuals with a greater time to diagnosis since birth will have a higher probability of being selected for the study, and the selected sample will not be representative of the general population of HD.

Though birth might be considered the natural initiating event, participants are not observed until entry into the study. Study entry is when participants join the cohort, and more importantly, become part of the risk set for determining survival probabilities. For these reasons, the time origin for prevalent cohort designs is commonly defined as the point of entry into the study. The time metric is *duration*, defined as the current age in years minus the age at study entry. Upon entering the study, individuals are tracked over a number of years. Interest often focuses on changes in a key marker, such as striatal volume, over the study duration.

To minimize length bias and make proper inferences, it is desirable to have a proxy variable of HD progression at the time of study entry. One participant might have a shorter duration to diagnosis than another due to a greater progression, and this must be taken into account in the interpretation of statistical results. Two variables important in characterizing disease progression are CAG repeat length and age [Langbehn et al., 2010].

A number of proxy variables for baseline HD progression have been suggested, with the work of Langbehn et al. [2004, 2010] being the most pertinent for the present purposes. The Langbehn et al. [2010] approach is based on a logistic survival model incorporating age and CAG length. In most applications [see, e.g., Aylward et al., 2010], an ordered categorical variable is used with groups representing three levels of estimated time to HD diagnosis (TTD). The groups are (i) *Far*: estimated TTD $>$15 years; (ii) *Middle* (*Mid*): 9 years $<$estimated TTD $\leq$15 years; and (iii) *Near*: estimated TTD $\leq$9 years.

Langbehn et al. had the laudable goal of developing a general model of TTD when no prospective study of prodromal HD existed. PREDICT-HD has conducted comprehensive annual assessments for prodromal HD participants for over 10 years, and there are currently 137 individuals prospectively diagnosed. This wealth of longitudinal data, in concert with the prospectively diagnosed patients evaluated over the course of their transition from presymptomatic to diagnosed, offer a rich resource from which to construct a new classification system.

There are two goals of this article. The first goal is to develop a "utility variable" for PREDICT-HD participants and other participants in the HD prodrome, which is a summary of two key variables observed at study entrance, age and CAG repeat length. The utility variable can be treated as a proxy variable of TTD for the duration of the study and used as a covariate or control variable to account for variation in disease progression in order to properly interpret the results of statistical analysis.

The second goal is to update the classification of study participants based on the utility variable. Rather than emphasize the distance to HD diagnosis that *Far–Mid–Near* implies, the classification of *Low–Med–High* is suggested to denote the amount of cumulative "disease burden" that participants have at the time they enter the study.

The utility variable is developed from a data-driven approach using a standard parametric survival model, the accelerated failure-time (AFT) model [Klein and Moeschberger, 2003]. It is argued that the adopted model has scientifically useful parameter interpretations and is highly predictive of TTD. We refer to the utility variable as the *CAG-Age Product* (CAP), as it is computed by multiplying age at entry (Age_0) by a scaling of CAG repeat length, as explained below. A scaled version of CAP, CAP_S, is also discussed that is especially useful for comparing mutation-negative individuals with mutation-positive individuals. Such comparison is a major preoccupation of PREDICT-HD analysis [see, e.g., Aylward et al., 2010] and other prodromal HD research, and prepares the way for preventive clinical trials.

METHODS

Overview

Two statistical analyses were performed. The first was concerned with development of CAP and CAP_S, and an evaluation of predictive performance. The second analysis was concerned with development of the *Low–Med–High* classification and comparison with the *Far–Mid–Near* groups. The analysis sections provide an overview of the methods and details can be found in the Appendix.

Participants

The first analysis was based on N = 730 prodromal individuals with at least 1 year and up to 7 years of follow-up data. These were participants who tested positive for the HD mutation, but did not have motor features indicating onset of diagnosable HD at the time of enrollment. Participants were seen yearly by clinicians experienced in the evaluation of movement disorders and specifically trained in administration of the Unified Huntington's Disease Rating Scale (UHDRS) for PREDICT-HD [Huntington Study Group, 1996]. Of the 730 prodromal individuals, 137 received an HD diagnosis over the course of the study.

All aspects of the study were approved by the Institutional Review Board at each participating institution, and all aspects of the study are in compliance with the declaration of Helsinki.

In accordance with clinical practice, diagnosis was made on the basis of "an otherwise unexplained characteristic movement disorder," operationally defined as a score of 4 on the HD Diagnostic Rating Scale of the UHDRS. A rating of 4 indicates that the clinician had ≥99% certainty that the participant showed "unequivocal presence of an otherwise unexplained extrapyramidal movement disorder." This method of diagnosis is "standard" in the respect of being approved by authorities such as the Huntington Study Group (HSG) [1996]. Participants were excluded from the analysis if they received a rating of 4 at baseline.

Portions of the second analysis added 233 control participants, for a total sample size of N = 963. The control participants were offspring of at least one HD-diagnosed parent, but tested negative for the HD mutation (CAG repeat length <36). Additional information about the participants and the PREDICT-HD study can be found in [Duff et al., 2010a; Paulsen et al., 2010b; Nopoulos et al., 2011; Stout et al., 2011].

STATISTICAL ANALYSIS I: MODEL AND UTILITY VARIABLE DEVELOPMENT

The first step in the analysis was a data-driven selection of an appropriate AFT model. The response variable was duration, which was the years since entry to HD diagnosis for individuals who received a diagnosis, and the years since entry to the last observation time for individuals who did not receive a diagnosis. All possible subsets of models with predictors of Age_0, CAG, and their interaction (i.e., $Age_0 \times CAG$) were considered. Reduced models were defined by having one or two of the predictor terms in the structural portion of the AFT. For example, one reduced model had Age_0 as the only predictor, and another had Age_0 and $Age_0 \times CAG$ as the predictors, etc.

In applications of the AFT, the exact distribution of the error term is typically unknown and various possibilities are considered. Therefore, the statistical analysis consisted of fitting models defined by combinations of number of predictors and different types of error distributions. There were 7 structural models and 5 error distributions for a total of 35 fitted models. Each model was estimated using maximum likelihood methods assuming right censoring. To assess model fit, Akaike's information criterion (AIC) was computed for each fitted model. In addition, the prediction error (PE) of the fitted model for the 137 HD diagnosed individuals was also computed ($PE = \sum |\hat{Y} - Y|$). All models were estimated using the survival package [Therneau and original Splus->R port by Thomas Lumley, 2009] of the R computer program [R Development Core Team, 2010].

The best fitting model according to the AIC had the predictors Age_0 and $Age_0 \times CAG$ with a normal error distribution, AIC = 888.1599, PE = 539.9115. The second best fitting model according to the AIC had the same predictors, but a logistic error distribution, AIC = 888.4643, PE = 496.6704. The difference in the AIC values ($\Delta AIC = 0.3044$) indicated the second best fitting model had essentially equal fit as the first, based on the common evaluation criterion of $\Delta AIC < 2$ [Burnham and Anderson, 2002]. The logistic model had a smaller PE value indicating less PE for those actually diagnosed. Therefore, the logistic AFT model was selected as the working model for continued development.

Having selected the model with the predictors Age_0 and $Age_0 \times CAG$, certain simplifications were made to yield a scientifically meaningful model with useful parameter interpretations (see the Appendix). The final AFT prediction model was

$$\hat{Y} = \exp[\alpha + \beta(Age_0)(CAG + C)] \quad (1)$$

Parameter estimates obtained via maximum likelihood are shown in Table I. Since C is a ratio of two parameters, the delta method was used to compute its standard error [see Casella and Berger, 2002].

The parameter estimates of Table I provide useful information. The sign of the estimate and the CI limits for C are negative. Assuming the parameter C is negative, CAG + C can be thought of as a type of calibrated or corrected CAG, with C being the correction factor. Thus, we define the corrected CAG repeat length, CAG_C, as

$$CAG_C = CAG + C \quad (2)$$

Substantively, CAG_C indexes the toxicity of the mutant huntingtin gene. Individuals with higher values have greater toxicity and those with lower values have less toxicity.

In addition to mutation toxicity, the other important consideration is Age_0. Age_0 is an index of the length of exposure to the mutation toxicity from birth to study entry. It is also the truncation time required to account for length bias [Wolkewitz et al., 2010]. Taking toxicity and length of exposure into account, we define the CAG and Age_0 product (CAP) as

$$CAP = Age_0 \times (CAG + C) = Age_0 \times CAG_C \quad (3)$$

Based on the data, $CAP = Age_0 \times (CAG - 33.6600)$. CAP is interpreted as an index of the *cumulative toxicity of mutant huntingtin* at study entry and is nearly identical to the measure, sometimes called "genetic burden," that was introduced by Penney et al. [1997]. Hence, participants with a larger cumulative genetic toxicity at study entry may likely develop HD sooner. It is emphasized that the CAP score is an *observed score* computed on an individual's actual age at study entry and CAG repeat length.

TABLE I. Parameter Estimates, Standard Errors (SEs), and 95% Confidence Intervals (CIs) for the Model of Equation (3)

Parameter	Estimate[a]	SE	95% CI
α	4.4196	0.3364	3.7602, 5.0790
β	−0.0065	0.0007	−0.0079, −0.0051
C	−33.6600	0.7046	−35.0409, −32.2790
$\log(\sigma)$	−0.8451	0.0710	−0.9843, −0.7060

[a] $P < 0.0001$ for all estimates.

Given the equations above, the CAP model for estimated TTD is written as

$$\hat{Y} = \exp(\alpha + \beta \times \mathrm{CAP}) \quad (4)$$

Using the estimates from Table I, we have $\hat{Y} = \exp(4.4196 - 0.0065 \times \mathrm{CAP})$. An important feature of Equation (4) is that estimated TTD is completely determined by CAP based on the fitted model. For this reason, CAP can be regarded as a proxy variable for estimated TTD. The utility is that CAP is computed based on the data, yet it is a substitute for the estimated TTD that is potentially computed from the model.

To enhance the utility of CAP, it can be scaled based on the estimated survival probabilities. Survival probabilities are convenient for studying the process of a disease and are a common means of communicating likelihood of an event, such as HD diagnosis [Kleinbaum and Klein, 2005].

For the survival probability scaling, 5 years is considered a reasonable landmark time, as it allows for a sufficient period of disease progression and is the common duration of clinical trials. Focus is on the key survival probability of 0.5, representing a 50-50 chance of a diagnosis. Given the assumptions mentioned in the Appendix, the scaled CAP, denoted as CAPS, is computed as

$$\mathrm{CAP_S} = \frac{\mathrm{CAP}}{\left[\frac{\alpha - \log(5)}{-\beta}\right]} \quad (5)$$

Based on the data, $\mathrm{CAP_S} = \mathrm{CAP}/432.3326$. $\mathrm{CAP_S}$ has a convenient interpretation: $\mathrm{CAP_S} < 1$ indicates a 5-year diagnosis probability of less than 0.5; $\mathrm{CAP_S} = 1$ indicates a 5-year probability equal to 0.5; and $\mathrm{CAP_S} > 1$ indicates a 5-year probability greater than 0.5. $\mathrm{CAP_S}$ then, is interpreted as an index of the *scaled cumulative mutation toxicity*, with the scaling being in reference to a 50-50 chance of diagnosis by 5 years.

An additional practical issue is that mutation-negative individuals (controls) are often considered along with mutation-positive individuals in PREDICT-HD analyses. Controls provide a convenient reference group for evaluating progression of HD in prodromal individuals. The control individuals have at least one parent diagnosed with HD, but their CAG repeat length is less than 36, indicating they will never develop HD. Control individuals have no mutation toxicity, which is represented by setting $\mathrm{CAP_S} = 0$ for these individuals. Suppose the probability of diagnosis after 5 years is D_5. Then the final definition of $\mathrm{CAP_S}$ is

$$\begin{array}{ll} \mathrm{CAP_s} = 0 & \text{if mutation-negative}(D_5 = 0), \\ \mathrm{CAP_s} < 1 & \text{if } D_5 < 0.5, \\ \mathrm{CAP_s} = 1 & \text{if } D_5 = 0.5, \\ \mathrm{CAP_s} > 1 & \text{if } D_5 > 0.5. \end{array} \quad (6)$$

Table II shows $\mathrm{CAP_S}$ along with other relevant scores and probabilities for hypothetical individuals aged 50 at study entry and having various CAG repeat length. The CAG repeat lengths are classified as mutation-negative (no penetrance), mutation-positive but reduced penetrance (36–40), and mutation-positive but full penetrance (>40). These penetrance classifications are based on proportions of individuals in the population expected to display HD symptoms sufficient to warrant diagnosis at some point [Langbehn et al., 2004].

The AFT model of Equation (4) can be expressed in terms of $\mathrm{CAP_S}$. Setting $\beta^* = -(\alpha - \log(5))$, the $\mathrm{CAP_S}$ AFT model is

$$\hat{Y} = \exp(\alpha + \beta^* \times \mathrm{CAP_S}) \quad (7)$$

Based on the data, $\hat{Y} = \exp(4.4196 - 2.8102 \times \mathrm{CAP_S})$.

The parameters of Equation (7) have the following interpretations. By definition, $\mathrm{CAP_S} = 0$ for mutation-negative individuals (see Eq. 6) meaning that α and β^* are irrelevant for these individuals. For mutation-positive individuals, $\mathrm{CAP_S} > 0$, and α is the intercept and β^* is the slope or acceleration of their scaled cumulative toxicity. As indicated by the CI in the second row of Table I, β is assumed to have a negative sign. Thus, as $\mathrm{CAP_S}$ increases, the expected TTD decreases. The acceleration is estimated to be $\hat{\beta}^* = -2.8102$, meaning a unit increase in $\mathrm{CAP_S}$ is accompanied by a 2.8102 decrease in log duration.

Predictive Performance and Bootstrap Validation

An important issue in the development of the $\mathrm{CAP_S}$ utility variable is the extent to which it predicts HD diagnosis. $\mathrm{CAP_S}$ is suggested as a proxy for cumulative disease burden, which implies that individuals with higher scores are closer to diagnosis than individuals with lower scores. A popular means of assessing predictive performance is based on the receiver operating characteristic (ROC) curve. In this context, the area under this curve (AUC) is an index of the ability of $\mathrm{CAP_S}$ to discriminate between diagnosed and non-diagnosed individuals. AUC varies between 0.5 and 1.0 for reasonable models, with higher values being better [Miller et al., 1993].

Since PREDICT-HD is a prospective study, diagnosis accumulates among the cohort over time. This time dependence should be taken into account when constructing the ROC curve. To evaluate the predictive performance of $\mathrm{CAP_S}$, longitudinal AUC was computed based on censored survival times using the methods of [Heagerty and Zheng, 2005]. The R package risksetROC was used for this analysis [Heagerty and packaging by Saha, 2011]. In addition to considering AUC as a function of study duration, an overall index was computed that summarizes the AUC over the time span, known as the integrated AUC (iAUC).

It is well known that predictive performance will decrease when a statistical model developed in one sample is applied to an independent sample. It is desirable to adjust for this "optimism" to provide a more realistic appraisal of predictive performance for out-of-sample data originating from the same population. One method of adjustment is the enhanced bootstrap [Efron, 1986; Efron and Tibshirani, 1993], the details of which are found in the Appendix. The enhanced bootstrap produces adjusted AUC and iAUC. The extent of similarity of the bootstrap-adjusted and unadjusted values is an indication of the internal validity of the AFT and $\mathrm{CAP_S}$ [Harrell et al., 1996]. Close agreement suggests that

TABLE II. Scores Probabilities, and Estimated TTD for Hypothetical Individuals Aged 50 at Study Entry With Various CAG Repeat Length

Age_0	CAG	CAG_C	CAP	CAP_S	$\hat{D}_5$	$\hat{Y}$
Mutation-negative (no penetrance)						
50	≤35	—	—	0.00	0.00	—
Reduced penetrance						
50	36	2.34	117.00	0.27	0.01	38.74
50	37	3.34	167.00	0.39	0.02	27.97
50	38[a]	4.34	217.00	0.50	0.04	20.19
50	39	5.34	267.00	0.62	0.08	14.57
50	40	6.34	317.00	0.74	0.15	10.52
Full penetrance						
50	41[b]	7.34	367.00	0.85	0.27	7.59
50	42[c]	8.34	417.00	0.97	0.45	5.48
50	43	9.34	467.00	1.08	0.63	3.96
50	44[d]	10.34	517.00	1.20	0.79	2.86
50	45	11.34	567.00	1.32	0.89	2.06
50	46	12.34	617.00	1.43	0.94	1.49
50	47	13.34	667.00	1.55	0.97	1.07
50	48	14.34	717.00	1.66	0.99	0.78
...	...	...	...	...	...	...
50	61[e]	27.34	1367.00	3.17	> 0.99	0.01

[a]Sample minimum.
[b]Lower quartile.
[c]Median.
[d]Upper quartile.
[e]Maximum.

when CAP_S is computed based on a new sample from the same population, it will function similarly as in the original sample.

Bootstrap-adjusted and unadjusted AUC and iAUC were examined using the N = 730 mutation-positive individuals. There were N = 137 participants who received a diagnosis over the course of the study (18.77%), and N = 593 individuals who did not (81.23%).

Results of the analysis show the overall unadjusted index was iAUC = 0.7203 and the bootstrap-adjusted value was slightly smaller, 0.7172. Figure 1 shows the AUC as a function of duration. The unadjusted values are depicted by a solid line and the bootstrap-adjusted values with a dashed line. As the figure shows, the AUC (adjusted and unadjusted) was relatively high (AUC > 0.75) for

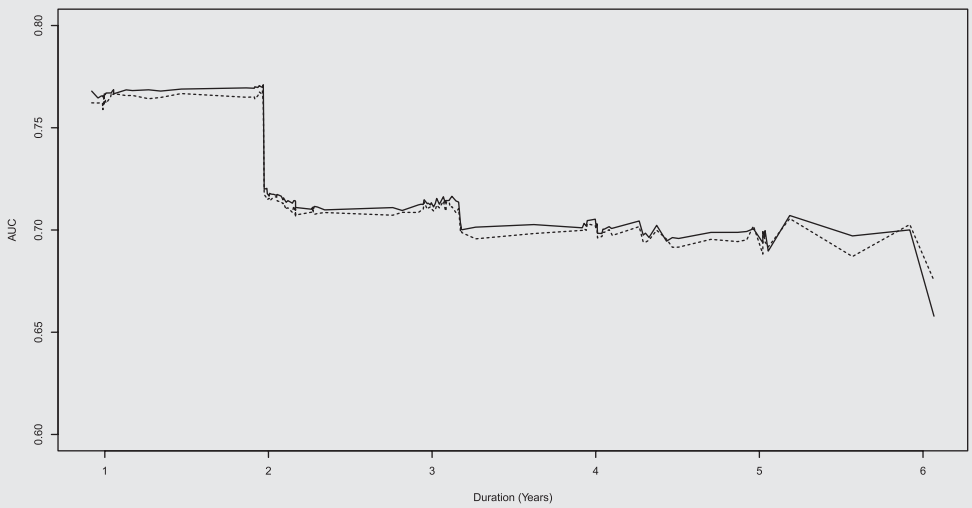

FIG. 1. AUC as a function of duration using CAP_S as the baseline predictor. Unadjusted AUC is depicted by a solid line, and bootstrap-adjusted AUC is depicted by a dashed line.

2 years duration, then dropped to around 0.70 until the last time point. The precipitous drop at 2 years was due to a small group of participants with the highest CAP_S values whose survival probability changed from 1 to 0 at this time. The bootstrap-adjusted AUC was only slightly lower that the unadjusted version indicating that CAP_S may have similar discriminatory power for a new sample from the same population.

STATISTICAL ANALYSIS II: GROUP CLASSIFICATION

CAP_S can be used whenever an index of proximity to diagnosis at study entry is needed. For example, CAP_S might be used in a regression model as a predictor of striatal volume to investigate prodromal indicators of disease progression [see Aylward et al., 2010].

There are situations in which researchers might consider using groups based on a categorization of the CAP_S distribution. When mutation-negative individuals (controls) and mutation-positive individuals (cases) are to be analyzed together, grouping has advantages. Mutation-negative individuals constitute a naturally occurring group, and there is no within-group variability of cumulative HD toxicity or proximity to diagnosis. This is reflected by the fact that controls always have $CAP_S = 0$.

Since the controls constitute a well-defined group, it is perhaps natural for some researchers to form similar mutation-positive groups for comparison. A dichotomous distinction, such as control/case groups, is too coarse a formulation as it ignores the CAP_S variability among mutation-positive individuals (only CAG repeat length is needed for a dichotomous classification). To take advantage of potentially important between-subjects variability in CAP_S, it is desirable to categorize the CAP_S distribution for mutation-positive individuals.

As mentioned in the introduction, there have been previous attempts to formulate groups representing proximity to diagnosis, the *Far–Mid–Near* of Langbehn et al. being the most relevant. In previous analyses [Beglinger et al., 2008; Paulsen et al., 2008, 2010a,b; Biglan et al., 2009; Klöppel et al., 2009; Aylward et al., 2010; Duff et al., 2010a; Nopoulos et al., 2010; Paulsen, 2010; Rowe et al., 2010; Stout et al., 2011], the formulation of the *Far–Mid–Near* groups was based on predicted values rather than observed scores with the criteria for categorizing being unclear. To remedy this, we propose using CAP_S and estimated TTD as the basis for grouping. This will help ensure maximal potential group differences of estimated primary outcomes. For continuity with the previous analysis, three groups will be considered. To distinguish the new groups from the old, the descriptors *Low*, *Med* (*Medium*), and *High* are used.

The cutoffs for group membership among mutation-positive individuals were determined by the algorithm outlined in the Appendix. This resulted in the rounded cutoffs values of $CAP_S = 0.67$ (lower cutoff), and $CAP_S = 0.85$ (upper cutoff). Based on Equation (3), these values corresponded to estimated durations of $\hat{Y} = 12.78$ and $\hat{Y} = 7.59$ years, respectively. Based on the cutoffs, the mutation-positive study participants were classified into the three groups. The fourth *Control* group consisted of the mutation-negative individuals with $CAP_S = 0$.

Examination of Group Differences

There are key variables used by PREDICT-HD researchers to study the progression of HD. These include baseline imaging, motor, cognitive, and psychiatric measures. Descriptive statistics for the key variables and some demographic variables (gender, age, CAG repeat length, and education) by classification (*Low–Med–High* and *Far–Mid–Near*) were computed. The N = 233 control participants (*Control* group) were included for a more thorough comparison (total N = 963).

As for the key variables, the imaging measure was striatal volume (as a ratio to total intracranial volume) [Paulsen et al., 2010a], and the motor measure was the UHDRS total motor score. There were two cognitive measures, the Symbol Digit Modalities Test (SDMT), which is a measure of processing speed, and the Hopkins Verbal Learning Test-Revised (HVLT-R), which is a test of immediate recall memory. Three psychiatric measures were included, the Frontal System Behavior Scale (FrSBe) yielding a total score and an executive subscale score, and the Schedule of Compulsions, Obsessions, and Pathologic Impulses (SCOPI) total score. The FrSBe is a self-report symptom inventory that measures behaviors associated with damage to frontal systems of the brain; the executive subscale evaluates problems with working memory, planning, problem solving, and insight [Grace and Malloy, 2001; Duff et al., 2010b]. The SCOPI is a multidimensional self-report measure composed of 47 items with a total score that is the sum of the obsessive checking, obsessive cleanliness, and compulsive rituals subscales [Watson and Wu, 2005; Beglinger et al., 2008].

The difference between the new and old classifications is summarized in the contingency table shown in Table III. As indicated by the values above the diagonal, the new grouping tended to classify study participants into more severe disease groups compared to the *Far–Mid–Near* classification. For example, the new classification moved 93 of the 275 participants in the *Far* group (34.55%) to the *Med* group, and 113 of the 276 in the *Mid* group (41.30%) to the *High* group.

Descriptive statistics for the key variables by classification are shown in Table IV. Note this is a comparison of *Control–Far–Mid–Near* and *Control–Low–Med–High*, with the *Control* group consisting of the same individuals in both classifications. One of the dramatic differences between the classifications is the sample size of the groups as the percentage of the total. Whereas the *Near* group

TABLE III. The Relationship Between the *Low–Med–High* Classification and the *Far–Mid–Near* Classification

Current classification	Proposed classification			
	Low	Med	High	Total
Far	182	93	0	275 (37.67%)
Mid	0	163	113	276 (37.81%)
Near	0	0	179	179 (24.52%)
Total	182 (24.93%)	256 (35.07%)	292 (40.00%)	730

TABLE IV. Descriptive Statistics of Key Variables for Two Classifications

Old groups	Control	Far	Mid	Near		
New groups	Control	Low	Med	High	Statistic/*P*-value	Group contrasts
N (% of total)	233 (24.20%)	275 (28.56%)	276 (28.66%)	179 (18.59%)	—	—
	233 (24.20%)	182 (18.90%)	256 (26.58%)	292 (30.32%)	—	
Gender (% F)	63.52	67.64	61.96	58.66	Chi-sq = 4.11, $P = 0.25$	ns
	63.52	67.03	65.23	59.25	Chi-sq = 3.58, $P = 0.31$	ns
Age mean (SD)	43.48 (11.63)	36.54 (8.13)	42.17 (9.74)	44.98 (10.16)	$F = 29.36$, $P < 0.0001$	C, N, M > F
	43.48 (11.63)	35.52 (7.49)	41.77 (9.32)	44.33 (9.76)	$F = 33.93$, $P < 0.0001$	C, H > M > L
CAG mean (SD)	20.01 (3.49)	41.00 (1.57)	42.55 (2.14)	44.23 (3.05)	$F = 4347.48$, $P < 0.0001$	C < F < M < N
	20.45 (3.49)	40.76 (1.56)	42.00 (1.99)	43.72 (2.77)	$F = 4271.10$, $P < 0.0001$	C < L < M < H
Education mean (SD)	14.65 (2.66)	14.55 (2.50)	14.35 (2.80)	14.16 (2.74)	$F = 1.36$, $P = 0.25$	ns
	14.65 (2.66)	14.52 (2.52)	14.43 (2.68)	14.24 (2.76)	$F = 1.09$, $P = 0.35$	ns
Striatal volume mean (SD)[a]	15.99 (2.14)	14.92 (2.19)	12.99 (2.33)	10.94 (2.45)	$F = 150.93$, $P < 0.0001$	C > F > M > N
	15.99 (2.14)	15.22 (2.15)	13.69 (2.25)	11.59 (2.52)	$F = 142.91$, $P < 0.0001$	C > L > M > H
UHDRS motor score mean (SD)	2.65 (3.35)	3.37 (3.73)	4.71 (4.76)	7.87 (6.81)	$F = 48.69$, $P < 0.0001$	C, F < M < N
	2.65 (3.35)	2.99 (3.33)	4.29 (4.40)	6.82 (6.35)	$F = 41.70$, $P < 0.0001$	C, L < M < H
Symbol digit mean (SD)	54.03 (8.95)	54.12 (11.30)	49.25 (10.46)	44.06 (11.05)	$F = 49.28$, $P < 0.0001$	C, F > M > N
	54.03 (8.95)	56.41 (11.44)	51.20 (10.46)	45.42 (10.93)	$F = 50.33$, $P < 0.0001$	C, L > M > H
HVLT mean (SD)	28.11 (4.43)	27.76 (4.24)	26.18 (5.13)	23.84 (5.42)	$F = 33.46$, $P < 0.0001$	C, F > M > N
	28.11 (4.43)	28.03 (4.04)	26.77 (4.95)	24.56 (5.37)	$F = 30.58$, $P < 0.0001$	C, L > M > H
FrSBe total mean (SD)	55.19 (12.90)	59.79 (19.78)	61.62 (19.87)	57.09 (18.87)	$F = 6.00$, $P = 0.0001$	M > C, N; F > C
	55.19 (12.90)	59.42 (20.05)	60.96 (19.76)	59.07 (19.30)	$F = 4.27$, $P = 0.0055$	C < M
FrSBe executive mean (SD)	24.59 (6.79)	26.53 (10.00)	27.47 (9.65)	25.80 (9.54)	$F = 4.40$, $P = 0.0044$	C < M
	24.59 (6.79)	25.23 (9.98)	26.89 (9.53)	26.85 (9.86)	$F = 3.36$, $P = 0.0183$	C < M, H
SCOPI mean (SD)[b]	73.54 (18.91)	76.66 (21.81)	77.52 (21.30)	71.01 (21.29)	$F = 2.02$, $P = 0.1128$	ns
	73.54 (18.91)	75.00 (21.39)	79.30 (22.12)	72.65 (20.88)	$F = 2.34$, $P = 0.0724$	ns

[a]Striatal volume comparison is based on a sample size of N = 798 (Control = 192, Low = 163, Med = 218, High = 225).
[b]SCOPI comparison is based on a sample size of N = 546 (Control = 175, Low = 113, Med = 130, High = 128).

consists of 18.59% of the total in the old classification, the comparable *High* group consists of 30.22% in the new classification.

The descriptive statistics of the other demographic variables were similar for the two classifications. The same can be said for motor score, symbol digit, HVLT, and striatal volume. There were differences between the two classifications for one of the psychiatric variables. For FrSBe Executive, there was a statistically significant difference between the *High* group and *Control* in the new classification, but not a similar difference for the two most distant groups in the old classification (i.e., *Near* and *Control*). For the remaining psychiatric variables (FrSBe Total and SCOPI), the *Mid* group had the largest mean value for the old classification, which was also true for the *Med* group. However, differences among the FrSBe scores were less pronounced for the new classification.

DISCUSSION

The first aim of this article was to develop a proxy of disease severity or ''disease burden'' in the form of proximity to diagnosis at study entry based on observed age at entry and CAG repeat length. The result was CAP_S, which is a scaled index of cumulative huntingtin mutation toxicity. The second aim was to develop a new classification of mutation-positive individuals that can be used for comparisons with mutation-negative individuals. The result was a categorization of the observed CAP_S distribution into the groups *Control–Low–Med–High*.

The predictive performance indexed by the longitudinal AUC values (see Fig. 1) suggest that CAP_S is a reasonable index for proximity to diagnosis, especially within the first 2 years. CAP_S is especially convenient for analysis in PREDICT-HD and other prevalent cohort studies because it is an *observed score*. CAP_S is computed from quantities that are observed with high accuracy at study entrance, and CAP_S does not require a statistical model for its estimation. The bootstrap-adjusted AUC in Figure 1 provides evidence that the scaling constants are relatively inconsequential, in the sense that when estimated on a new sample from the same population, the predictive performance of CAP_S is essentially the same.

It was demonstrated that the new classification of *Low–Med–High* based on CAP_S in relation to the old (*Far–Mid–Near*) represents a migration of individuals from less severe disease categories to more severe categories (see Table III). This migration seems desirable as the PREDICT-HD cohort has aged since the older classification, which has increased the risk of diagnosis in the cohort. As Table IV illustrates, the *Low–Med–High* classification using CAP_S is clinically meaningful, in that, it is correlated with differences in other widely recognized markers of prodromal HD. Not only does the classification strongly predict TTD, but it also

groups prodromal HD participants into potentially more homogeneous cohorts for the study of clinical markers.

Baseline clinical markers are very important in the study of prodromal HD as they are indicators of disease progression and potentially provide information for understanding the relevant mechanisms involved [Beglinger et al., 2008; Paulsen et al., 2008, 2010a,b; Biglan et al., 2009; Aylward et al., 2010; Duff et al., 2010a; Nopoulos et al., 2010; Paulsen, 2010; Stout et al., 2011]. Figure 2 is a hypothesized illustration of HD disease progression. CAP_S represents a necessary but not sufficient condition for the progression of HD, namely a sufficiently extended CAG repeat length and a sufficiently long exposure period. As opposed to other hypothesized variables, CAP_S is an observed variable that is hypothesized to exert influence at the beginning of the causal change depicted in Figure 2.

Assuming the hypothesized model in Figure 2, the *Low–Med–High* classification may be important for the design of clinical trials, which is a future goal of PREDICT-HD. The second set of results show that 40% (292/730) of prodromal HD participants in the PREDICT-HD study were classified into the *High* group (see Table IV). For the prodromal HD individuals in this group, the average estimated time to HD diagnosis was 5.47 years with a standard deviation of 1.41, and the average estimated 5-year probability of HD diagnosis was 0.47 with a standard deviation of 0.16. This group may serve as a potential focus group for future clinical trials on therapeutic treatment designed to delay HD progression. For example, suppose a balanced treatment-control clinical trial is designed with the aim of reducing the 5-year HD diagnosis by 40%. If the study participants were enrolled from the *High* group, the trial would require only 140 participants for each experimental group to demonstrate such treatment efficacy in a two-sided 0.05 level test powered at 0.90. As the PREDICT-HD study has already enrolled 292 individuals that are classified in the *High* group within 7 years, it is anticipated that the recruitment of study participants for such clinical trials with 5-year follow-up can be achieved in <7 years with a multicenter setting.

An important caveat to be highlighted is that CAP_S is intended as a utility variable for analyses in the PREDICT-HD and similar prevalent cohort studies. It is not intended as a true or literal measure of an individual's distance from HD diagnosis. CAP_S summarizes two of the key variables related to the timing of HD diagnosis: age and CAG repeat length. However, there are undoubtedly numerous other important variables—most unobserved—that influence the TTD, as depicted at the left in Figure 2. PREDICT-HD participants are observed only once a year. Assuming there is a threshold for passing from prodromal to diagnosed, it is impossible to determine the exact TTD based on yearly measures (monthly or daily measures would be required). In addition, there is debate as to the validity of the threshold of CAG $\geq$36 for classifying individuals as mutation-positive. The validity of the threshold is not a concern for the development of CAP_S in this article because no PREDICT-HD participants had CAG repeated lengths in the range of 35–37 inclusive. However, for samples that include numerous individuals with CAG lengths in this range, the estimated AFT and CAP_S will be affected. Therefore, we caution the reader to not over-interpret CAP_S, as *it does not represent the true TTD for prodromal HD individuals*. The same can be said for the *Low–Med–High* classifications.

The AFT model developed in the analysis has the desired characteristic of simplicity as compared to related models [see Langbehn et al., 2010]. The predictive performance of the model, as reflected by the discriminatory ability of CAP_S (see Fig. 1), is comparable to that of some other sample-developed models in other areas of medical research [e.g., Bleeker et al., 2003]. The bootstrap validation suggests that CAP_S might be useful for new samples from the same population; the population consisting of prodromal individuals in prevalent cohort studies. However, the bootstrap adjustment does not address the issue of performance for data from a different population. The utility of CAP_S for individuals with characteristics that differ from those in PREDICT-HD is an open question and probably depends on the similarity of key characteristics [Laupacis et al., 1997].

CONCLUSION

In conclusion, CAP_S and its associated classification can be used as an index of proximity to HD diagnosis at study entry for individuals in the PREDICT-HD study and in similar prevalent cohort studies. An index of proximity is important for prodromal analyses, so that differences among prodromal HD individuals can be properly interpreted. It is hoped that the use of CAP_S and the *Control–Low–Med–High* classification will lead to better prediction and greater understanding of the processes of HD.

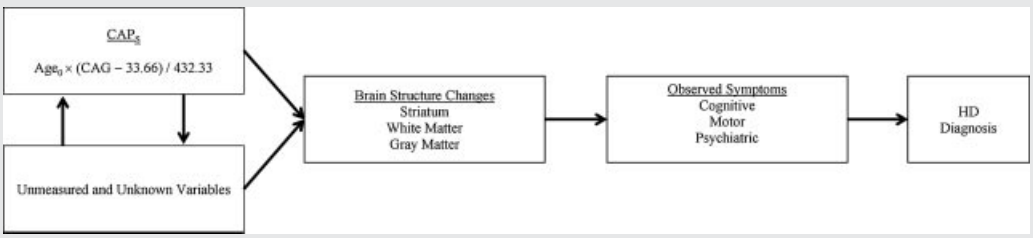

FIG. 2. Hypothesized diagram for HD progression.

ACKNOWLEDGMENTS

We thank the PREDICT-HD sites, the study participants, and the National Research Roster for Huntington disease Patients and Families. This research is supported by the National Institutes for Health, National Institute of Neurological Disorders and Stroke (NS40068) and CHDI Foundation, Inc. PREDICT-HD Investigators, Coordinators, Motor Raters, Cognitive Raters—Active: September 2009 to August 2010. Thomas Wassink, MD, Stephen Cross, BA, Nicholas Doucette, BA, Mycah Kimble, BA, Patricia Ryan, MSW, LISW, MA, Jessica Wood, MD, PhD, Eric A. Epping, MD, PhD, and Leigh J. Beglinger, PhD (University of Iowa, Iowa City, Iowa, USA); Edmond Chiu, MD, Olga Yastrubetskaya, PhD, Joy Preston, Anita Goh, D.Psych, Chathushka Fonseka, and Liz Ronsisvalle (St. Vincent's Hospital, The University of Melbourne, Kew, Victoria, Australia); Phyllis Chua, MD, and Angela Komiti, BS, MA (The University of Melbourne, Royal Melbourne Hospital, Melbourne, Australia); Lynn Raymond, MD, PhD, Rachelle Dar Santos, BSc, and Joji Decolongon, MSC, CCRP (University of British Columbia, Vancouver, British Columbia, Canada); Adam Rosenblatt, MD, Christopher A. Ross, MD, PhD, Barnett Shpritz, BS, MA, OD, and Claire Welsh (Johns Hopkins University, Baltimore, Maryland, USA); William M. Mallonee, MD, Greg Suter, BA, and Judy Addison (Hereditary Neurological Disease Centre, Wichita, Kansas, USA); Ali Samii, MD, and Alma Macaraeg, BS (University of Washington and VA Puget Sound Health Care System, Seattle, Washington, USA); Randi Jones, PhD, Cathy Wood-Siverio, MS, Stewart A. Factor, DO, and Claudia Testa, MD, PhD (Emory University School of Medicine, Atlanta, Georgia, USA); Roger A. Barker, BA, MBBS, MRCP, Sarah Mason, BSC, Anna Goodman, PhD, Rachel Swain, BA, and Anna DiPietro (Cambridge Centre for Brain Repair, Cambridge, UK); Elizabeth McCusker, MD, Jane Griffith, RN, Clement Loy, MD, David Gunn, BS, and Linda Stewart, RN (Westmead Hospital, Sydney, Australia); Bernhard G. Landwehrmeyer, MD, Michael Orth, MD, PhD, Sigurd Süβmuth, MD, RN, Katrin Barth, RN, and Sonja Trautmann, RN (University of Ulm, Ulm, Germany); Kimberly Quaid, PhD, Melissa Wesson, MS, and Joanne Wojcieszek, MD (Indiana University School of Medicine, Indianapolis, IN); Mark Guttman, MD, Alanna Sheinberg, BA, and Irita Karmalkar, BSc (Centre for Addiction and Mental Health, University of Toronto, Markham, Ontario, Canada); Susan Perlman, MD and Arik Johnson, PsyD (University of California, Los Angeles Medical Center, Los Angeles, California, USA); Michael D. Geschwind, MD, PhD, Jon Gooblar, BA, and Gail Kang, MD (University of California San Francisco, California, USA); Tom Warner, MD, PhD, Maggie Burrows, RN, BA, Marianne Novak, MD, Thomasin Andrews, MD, BSC, MRCP, Elisabeth Rosser, MBBS, FRCP, and Sarah Tabrizi, MD, PhD (National Hospital for Neurology and Neurosurgery, London, UK); Anne Rosser, MD, PhD, MRCP, Kathy Price, RN, and Sarah Hunt, BSc (Cardiff University, Cardiff, Wales, UK); Frederick Marshall, MD, Amy Chesire, LCSW-R, MSG, Mary Wodarski, BA, and Charlyne Hickey, RN, MS (University of Rochester, Rochester, New York, USA); Oksana Suchowersky, MD, FRCPC, Sarah Furtado, MD, PhD, FRCPC, and Mary Lou Klimek, RN, BN, MA (University of Calgary, Calgary, Alberta, Canada); Peter Panegyres, MB, BS, PhD, Elizabeth Vuletich, BSC, Steve Andrew, and Rachel Zombor, MPSYC (Neurosciences Unit, Graylands, Selby-Lemnos & Special Care Health Services, Perth, Australia); Joel Perlmutter, MD, Stacey Barton, MSW, LCSW, and Amy Schmidt (Washington University, St. Louis, Missouri, USA); Zosia Miedzybrodzka, MD, PhD, Sheila A. Simpson, MD, Daniela Rae, RN, and Mariella D'Alessandro, PhD (Clinical Genetics Centre, Aberdeen, Scotland, UK); David Craufurd, MD, Ruth Fullam, BSC, and Elizabeth Howard, MD (University of Manchester, Manchester, UK); Pietro Mazzoni, MD, PhD, Karen Marder, MD, MPH, and Paula Wasserman, MA (Columbia University Medical Center, New York, New York, USA); Rajeev Kumar, MD and Diane Erickson, RN (Colorado Neurological Institute, Englewood, Colorado, USA); Vicki Wheelock, MD, Terry Tempkin, RNC, MSN, Nicole Mans, BA, MS, and Kathleen Baynes, PhD (University of California Davis, Sacramento, California, USA); Joseph Jankovic, MD, Christine Hunter, RN, CCRC, and William Ondo, MD (Baylor College of Medicine, Houston, Texas, USA); Justo Garcia de Yebenes, MD, Monica Bascunana Garde, Marta Fatas, BA, and Asuncion Martinez-Descales (Hospital Ramón y Cajal, Madrid, Spain); Wayne Martin, MD, Pamela King, BScN, RN, and Satwinder Sran, BSC (University of Alberta, Edmonton, Alberta, Canada); Anwar Ahmed, PhD, Stephen Rao, PhD, Christine Reece, BS, Janice Zimbelman, PhD, PT, Alexandra Bea, BA, Emily Newman, BA, and Alex Bura, BA (Cleveland Clinic Foundation, Cleveland, Ohio, USA).

Steering Committee—Jane Paulsen, PhD, Principal Investigator, Eric A. Epping, MD, PhD, Hans Johnson, PhD, Megan Smith, PhD, Janet Williams, PhD, RN, FAAN, Leigh Beglinger, PhD, James Mills, MS (University of Iowa Hospitals and Clinics, Iowa City, IA); Elizabeth Aylward, PhD (Seattle Children's Research Institute, WA); Kevin Biglan, MD (University of Rochester, Rochester, NY); Blair Leavitt, MD (University of British Columbia, Vancouver, BC, Canada); Marcy MacDonald, PhD (Massachusetts General Hospital); Martha Nance, MD (Hennepin County Medical Center, Minneapolis, MN); and Cheryl Erwin, JD, PhD (University of Texas Medical School at Houston).

Scientific Sections—Bio Markers: Blair Leavitt, MDCM, FRCPC (Chair) and Michael Hayden, PhD (University of British Columbia); Stefano DiDonato, MD (Neurological Institute "C. Besta," Italy); Ken Evans, PhD (Ontario Cancer Biomarker Network); Wayne Matson, PhD (VA Medical Center, Bedford, MA); Asa Peterson, MD, PhD (Lund University, Sweden), Sarah Tabrizi, MD, PhD (National Hospital for Neurology and Neurology and Neurosurgery, London); Beth Borowsky, PhD (CHDI); Andrew Juhl, BS, James Mills, MS, Kai Wang, PhD (University of Iowa); and David Weir, BSc (University of British Columbia). **Brain**: Jean Paul Vonsattell, PhD (Chair), and Carol Moskowitz, ANP, MS (Columbia University Medical Center); Anne Leserman, MSW, LISW, Lynn Schaul, BA, and Stacie Vik, BA (University of Iowa). **Cognitive**: Deborah Harrington, PhD (Chair), Gabriel Castillo, BS, Jessica Morison, BS, and Jason Reed, BS (University of California, San Diego), Michael Diaz, PhD, Ian Dobbins, PhD, Tamara Hershey, PhD, Erin Foster, OTD, and Deborah Moore, BA (Washington University Cognitive Science Battery Development); Holly Westervelt, PhD (Chair, Quality Control and Training, Alpert Medical School of Brown University), Jennifer Davis, PhD, and Geoff Tremont, PhD, MS (Scientific Consultants, Alpert

Medical School of Brown University); Megan Smith, PhD (Chair, Administration), David J. Moser, PhD, Leigh J. Beglinger, PhD, Kelly Rowe, and Danielle Theriault, BS (University of Iowa); Carissa Gehl, PhD (VA Medical Center, Iowa City, IA); Kirsty Matheson (University of Aberdeen); Karen Siedlecki, PhD (Fordham University); Marleen Van Walsem (EHDN); Susan Bonner, BA, Greg Elias, BA, and Melanie Faust, BS (Rhode Island Hospital); Beth Borowski, PhD (CHDI); Noelle Carlozzi (University of Michigan); Kevin Duff, PhD (University of Utah); Nellie Georgiou-Karistianis (St. Vincent's Hospital, The University of Melbourne, Australia); Julie Stout, PhD (Monash University, Melbourne, Australia); Herwig Lange (Air-Rahazentrum); and Kate Papp (University of Connecticut). **Functional**: Janet Williams, PhD (Chair), Leigh J. Beglinger, PhD, Anne Leserman, MSW, LISW, Eunyoe Ro, MA, Lee Anna Clark, Nancy Downing, Joan Laing, PhD, Kristine Rees, BA, and Stacie Vik, BA (University of Iowa); Rebecca Ready, PhD (University of Massachusetts); Anthony Vaccarino, PhD (Ontario Cancer Biomarker Network); Sarah Farias, PhD (University of California, Davis); Noelle Carlozzi, PhD (University of Michigan); and Carissa Gehl, PhD (VA Medical Center, Iowa City, IA). **Genetics**: Marcy MacDonald, PhD (Co-Chair), Jim Gusella, PhD, and Rick Myers, PhD (Massachusetts General Hospital); Michael Hayden, PhD (University of British Columbia); Tom Wassink, MD (Co-Chair) Eric A. Epping, MD, PhD, Andrew Juhl, BA, James Mills, MS, and Kai Wang, PhD (University of Iowa); Zosia Miedzybrodzka, MD, PhD (University of Aberdeen); and Christopher Ross, MD, PhD (Johns Hopkins University). **Imaging**—*Administrative*: Ron Pierson, PhD (Chair), Kathy Jones, BS, Jacquie Marietta, BS, William McDowell, AA, Greg Harris, BS, Eun Young Kim, MS, Hans Johnson, PhD, and Thomas Wassink, MD (University of Iowa); John Ashburner, PhD (Functional Imaging Lab, London); Steve Potkin, MD (University of California, Irvine); and Arthur Toga, PhD (University of California, Los Angeles). *Striatal*: Elizabeth Aylward, PhD (Chair, Seattle Children's Research Institute). *Surface Analysis*: Eric Axelson, BSE (University of Iowa). *Shape Analysis*: Christopher A. Ross (Chair), MD, PhD, Michael Miller, PhD, and Sarah Reading, MD (Johns Hopkins University); Mirza Faisal Beg, PhD (Simon Fraser University). *DTI*: Vincent A. Magnotta, PhD (Chair, University of Iowa); Karl Helmer, PhD (Massachusetts General Hospital); Kelvin Lim, MD (University of Ulm, Germany-); Mark Lowe, PhD (Cleveland Clinic); Sasumu Mori, PhD (Johns Hopkins University); Allen Song, PhD (Duke University); and Jessica Turner, PhD (University of California, Irvine). *fMRI*: Steve Rao, PhD (Chair), Erik Beall, PhD, Katherine Koenig, PhD, Michael Phillips, MD, Christine Reece, BS, and Jan Zimbelman, PhD, PT (Cleveland Clinic); and April Bryant (University of Iowa). **Motor**: Kevin Biglan, MD (University of Rochester), Karen Marder, MD (Columbia University), and Jody Corey-Bloom, MD, PhD (University of California, San Diego) all Co-Chairs; Michael Geschwind, MD, PhD (University of California, San Francisco); Ralf Reilmann, MD and Zerka Unds (Muenster, Germany); and Andrew Juhl, BS (University of Iowa). **Psychiatric**: Eric A. Epping, MD, PhD (Chair), Nancy Downing, RN, MSN, Jess Fiedorowicz, MD, Robert Robinson, MD, Megan Smith, PhD, Leigh Beglinger, PhD, James Mills, MS, Kristine Rees, BA, Adam Ruggle, Stacie Vik, BA, Janet Williams, PhD, Dawei Liu, PhD, David Moser, PhD, and Kelly Rowe (University of Iowa); Karen Anderson, MD (University of Maryland); David Craufurd, MD (University of Manchester); Mark Groves, MD (Columbia University); Anthony Vaccarino, PhD and Ken Evans, PhD (Ontario Cancer Biomarker Network); Hugh Rickards, MD (Queen Elizabeth Psychiatric Hospital); Eric van Duijn, MD (Leiden University Medical Center, Netherlands); Irina Antonijevic, MD, PhD, and Joseph Giuliano (CHDI); Phyllis Chua (The University of Melbourne, Royal Melbourne Hospital); and Kimberly Quaid, PhD (Indiana University School of Medicine). Core Sections—*Statistics*: James Mills, MEd, MS, Dawei Liu, PhD, Jeffrey Long, PhD, Wenjing Lu, Kai Wang, PhD, and Ying Zhang, PhD (University of Iowa). *Recruitment/Retention*: Martha Nance, MD (Chair, University of Minnesota); Anne Leserman, MSW, LISW, Nicholas Doucette, BA, Mycah Kimble, BA, Patricia Ryan, MSW, LISW, MA, Kelli Thumma, BA, Elijah Waterman, BA, and Jeremy Hinkel, BA (University of Iowa). *Ethics*: Cheryl Erwin, JD, PhD (Chair, McGovern Center for Health, Humanities and the Human Spirit); Eric A. Epping, MD, PhD Janet Williams, PhD, Nicholas Doucette, BA, Anne Leserman, MSW, LISW, James Mills, MS, Lynn Schaul, BA, and Stacie Vik, BA (University of Iowa); Martha Nance, MD (University of Minnesota); and Lisa Hughes, MEd (University of Texas Medical School at Houston). *IT/Management*: Hans Johnson, PhD (Chair), R.J. Connell, BS, Karen Pease, BS, Ben Rogers, BA, BSCS, Jim Smith, AS, Shuhua Wu, MCS, Roland Zschiegner, Erin Carney, Bill McKirgan, Mark Scully, and Ryan Wyse (University of Iowa); Jeremy Bockholt (AMBIGroup). **Program Management**—*Administrative*: Chris Werling-Witkoske (Chair), Karla Anderson, BS, Kristine Bjork, BA, Ann Dudler, Jamy Schumacher, Sean Thompson, BA, Leann Davis, Machelle Henneberry, Greg Ennis, MA, and Stacie Vik, BA (University of Iowa). *Financial*: Steve Blanchard, MSHA, Kelsey Montross, BA, and Phil Danzer (University of Iowa).

APPENDIX A

This appendix provides the technical details pertaining to the development of the prediction model and utility variables (CAP and CAP_S), the group classifications (*Low–Med–High*), and the bootstrap adjustment for predictive performance.

Development of the Prediction Model: Suppose that study entrance is the time origin, so that Y is the time in years to HD diagnosis (i.e., duration). Then the full model AFT is

$$\log(Y) = \alpha + \beta_1 \mathrm{Age}_0 + \beta_2 \mathrm{CAG} + \beta_3(\mathrm{Age}_0 \times \mathrm{CAG}) + \sigma\varepsilon \quad (A.1)$$

This is essentially a linear model for Y in the natural logarithm scale with an interaction effect of age at study entrance and CAG repeat length. The model guarantees the positive estimation of TTD [Klein and Moeschberger, 2003]. In Equation (A.1), σ is a scaling parameter and ε is a random variable denoting the model error which may be a proxy for many other potentially important predictors not considered. The exact distribution of ε is typically unknown and empirical AFT modeling involves an exploration of various possibilities, usually the *Exponential*, *Weibull*, *Normal*, *Logistic*, or *Generalized Gamma* distributions.

Equation (A.1) was only one of several potential models with reduced models defined by having one or two of the predictor terms in the structural portion of Equation (A.1). As discussed in the text, the selected model had Age_0 and $Age_0 \times CAG$ as predictors and a logistic error distribution. The prediction formula for the selected model was

$$\hat{Y} = \exp[\alpha + \beta_1 Age_0 + \beta_3(Age_0 \times CAG) \quad (A.2)$$

Certain simplifications were made to yield a scientifically meaningful model with useful parameter interpretations. To simplify Equation (A.2), define $C = \beta_1/\beta_3$, so that $\beta_1 = \beta_3 \times C$. Then the selected working model can be written as

$$\begin{aligned}\hat{Y} &= \exp[\alpha + \beta_3(Age_0 \times C) + \beta_3(Age_0 \times CAG)] \\ &= \exp[\alpha + \beta_3(Age_0)(CAG + C)]\end{aligned}$$

Dropping the subscript for β_3 yields

$$\hat{Y} = \exp[\alpha + \beta(Age_0)(CAG + C)] \quad (A.3)$$

Defining $CAP = Age_0 \times (CAG + C)$ we have Equation (4) in the text.

CAP_S is a scaling of CAP based on the estimated survival probabilities. Assuming a logistic error distribution, the model of Equation (A.3) can be expressed in terms of a survival probability [Klein and Moeschberger, 2003]. The survival probability of being diagnosis-free from study entry to a future time t is given by

$$S_t = \left[1 + \exp\left\{\frac{\log(t) - (\alpha + \beta \times CAP)}{\sigma}\right\}\right]^{-1} \quad (A.4)$$

where all the parameters are as defined above.

Based on the parameter estimates in Table I, the estimated probability of being diagnosis-free at time t is computed as

$$\hat{S}_t = \left[1 + \exp\left\{\frac{\log(t) - (4.4196 - 0.0065 \times CAP)}{\exp(-0.8451)}\right\}\right]^{-1}$$

It follows that the estimated probability of diagnosis at time t is $\hat{D}_t = 1 - \hat{S}_t$.

Focus is on the key survival probability of 0.5, as in this case $\hat{D}_t = \hat{S}_t$ and there is a 50-50 chance of a diagnosis. Inspection of Equation (A.4) reveals that $S_t = D_t = 0.5 = [1 + \exp(0)]^{-1}$, which occurs when $\log(t) - (\alpha + \beta \times CAP) = 0$. Given $t = 5$ and estimates of the parameters, CAP can be solved for in this situation. This solution is then used as a basis for scaling CAP.

Suppose we define $B = CAP$ for this special case and solve for $\hat{B}$ using the sample estimates,

$$\hat{B} = \frac{\hat{\alpha} - \log(5)}{-\hat{\beta}} = \frac{4.4196 - 1.6094}{0.0065} = 432.3326$$

$\hat{B}$ is the sample CAP that produces $\hat{D}_t = \hat{S}_t = 0.5$. Then CAP can be divided by B (or its estimate) to compute CAP_S as in Equation (5). Finally, setting $\beta^* = \beta \times B = -(\alpha - \log(5))$, we obtain the CAP_S AFT model of Equation (6).

Algorithm for determining group membership: Recall that CAP_S can be used as a basis for classifying individuals into groups. Using the sample data, the cutoffs for group membership among mutation-positive individuals were determined by the following steps.

(1) For each study participant, CAP_S was computed along with the estimated time to diagnosis, $\hat{Y}$, based on Equation (4).
(2) To avoid radically different sample sizes among the groups, the potential lower cutoff was constrained to be between the 25th and 40th percentiles of the CAP_S sample distribution.
(3) The potential upper cutoff was constrained to be between the 60th and 75th percentiles of the CAP_S distribution.
(4) The grid of all possible pairs of lower and upper percentiles was considered. For each candidate pair of cutoffs (e.g., 26th and 64th percentiles), the between-group and within-group variation for the estimated TTD was calculated. The ratio of between-group to within-group variation was used as the selection criterion.
(5) The cutoffs for constructing the groups consisted of the pair that yielded the largest value of the ratio.

The optimization algorithm essentially searched for the largest ANOVA *F*-statistic of the estimated TTD based on the CAP_S percentiles. The selected cutoffs indicated the best separation of the study participants in terms of the estimated TTD.

Bootstrap adjustment of predictive performance: To address the issue of optimistic prediction performance, an enhanced bootstrap procedure was used to adjust the AUC and iAUC [see Harrell et al., 1996 for additional details]. The overly optimistic apparent performance (AP) was indexed by the AUC and iAUC computed on the original sample. For each bootstrap replication, a bootstrap sample was drawn by sampling from the original data set with replacement using the same sample size ($N = 730$). For each bootstrap sample, the parameters of Equation (A.2) were estimated, CAP_S was computed, and AUC and iAUC were produced. The latter two were indicators of bootstrap performance (BP). Then CAP_S was computed for the original sample based on the estimates for the bootstrap sample, and AUC and iAUC were produced. In this case, AUC and iAUC were indicators of test performance (TP). The bootstrap procedure was replicated 200 times and the adjusted AUC (AUC_{adj}) was computed as

$$AUC_{adj} = AUC_{AP} - \text{average}(AUC_{BP} - AUC_{TP}) \quad (A.6)$$

where the average was taken over the 200 replications. A similar adjustment was computed for iAUC. The difference $AUC_{BP} - AUC_{TP}$ was an estimate of optimism, and averaging over replications produced a more stable estimate. The average was subtracted from the AUC value computed on the original sample (AUC_{AP}), which constituted an optimism penalty. A relatively few replications (i.e., 200) were sufficient, as the average was used rather than the empirical quantiles required by other bootstrap methods (e.g., a bootstrap confidence interval).

REFERENCES

Andresen JM, Gayan J, Cherny SS, Brocklebank D, Alkorta-Aranburu G, Addis EA, Cardon LR, Housman DE, Wexler NS. 2007a. Replication of twelve association studies for Huntington's disease residual age of onset in large Venezuelan kindreds. J Med Genet 44(1):44–50.

Andresen JM, Gayan J, Djousse L, Roberts S, Brocklebank D, Cherny SS, Cardon LR, Gusella JF, MacDonald ME, Myers RH, et al. 2007b. The relationship between CAG repeat length and age of onset differs for Huntington's disease patients with juvenile onset or adult onset. Ann Hum Genet 71(Pt3):295–301.

Andrew SE, Goldberg YP, Kremer B, Telenius H, Theilmann J, Adam S, Starr E, Squitieri F, Lin B, Kalchman MA, et al. 1993. The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. Nat Genet 4(4):398–403.

Aylward EH, Nopoulos PC, Ross CA, Langbehn DR, Pierson RK, Mills JA, Johnson HJ, Magnotta VA, Juhl AR, Paulsen JS. 2010. Longitudinal change in regional brain volumes in prodromal Huntington disease. J Neurol Neurosurg Psychiatry 82:405–410.

Beglinger LJ, Paulsen JS, Watson DB, Wang C, Duff K, Langbehn DR, Moser DJ, Paulson HL, Aylward EH, Carlozzi NE, et al. 2008. Obsessive and compulsive symptoms in prediagnosed Huntington's disease. J Clin Psychiatry 69(11):1758–1765.

Biglan KM, Ross CA, Langbehn DR, Aylward EH, Stout JC, Queller S, Carlozzi NE, Duff K, Beglinger LJ, Paulsen JS, et al. 2009. Motor abnormalities in premanifest persons with Huntington's disease: The PREDICT-HD study. Mov Disord 24(12):1763–1772.

Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derkson-Lubsen G, Grobbee DE, Moons KGM. 2003. External validation is necessary in prediction research: A clinical example. J Clin Epidemiol 56(9):826–832.

Brinkman RR, Mezei MM, Theilmann J, Almqvist E, Hayden MR. 1997. The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. Am J Hum Genet 60(5):1202–1210.

Bruland O, Almqvist EW, Goldberg YP, Boman H, Hayden MR, Knappskog PM. 1999. Accurate determination of the number of CAG repeats in the Huntington disease gene using a sequence-specific internal DNA standard. Clin Genet 55(3):198–202.

Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: A practical information-theoretic approach. New York: Springer. xxvi, 488 p.

Campodonico JR, Codori AM, Brandt J. 1996. Neuropsychological stability over two years in asymptomatic carriers of the Huntington's disease mutation. J Neurol Neurosurg Psychiatry 61(6):621–624.

Casella G, Berger RL. 2002. Statistical inference. Australia: Pacific Grove, CA/Thomson Learning. xxviii, 660 p.

Duff K, Paulsen J, Mills J, Beglinger LJ, Moser DJ, Smith MM, Langbehn D, Stout J, Queller S, Harrington DL. 2010a. Mild cognitive impairment in prediagnosed Huntington disease. Neurology 75(6):500–507.

Duff K, Paulsen JS, Beglinger LJ, Langbehn DR, Wang C, Stout JC, Ross CA, Aylward E, Carlozzi NE, Queller S. 2010b. "Frontal" behaviors before the diagnosis of Huntington's disease and their relationship to markers of disease progression: Evidence of early lack of awareness. J Neuropsychiatry Clin Neurosci 22(2):196–207.

Efron B. 1986. How biased is the apparent error rate on a prediction rule? J Am Stat Assoc 81(394):461–470.

Efron B, Tibshirani R. 1993. An introduction to the bootstrap. Monographs on statistics and applied probability. New York: Chapman & Hall. 456 p.

Grace J, Malloy P. 2001. Frontal Systems Behavior Scale (FrSBe): Professional manual. Lutz, FL: Psychological Assessment Resources. 109 p.

Gutiérrez C, MacDonald A. 2002. Huntington's Disease and Insurance I: A model of Huntington's disease. Edinburgh: Genetics and Insurance Research Centre (GIRC). 28 p.

Gutiérrez C, MacDonald A. 2004. Huntington's disease, critical illness insurance and life insurance. Scand Actuarial J 4:279–311.

Harrell FE, Lee KL, Mark DB. 1996. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15(4):361–387.

Heagerty PJ, Zheng Y. 2005. Survival model predictive accuracy and ROC curves. Biometrics 61(1):92–105.

Heagerty PJ, packaging by Saha P. 2011. risksetROC: Riskset ROC curve estimation from censored survival data. R package version 1.0.3. http://CRAN.R-project.org/package=risksetROC.

Huntington Study Group. 1996. Unified Huntington's Disease Rating Scale: Reliability and consistency. Mov Disord 11(2):136–142.

Klein J, Moeschberger M. 2003. Survival analysis. Techniques for censored and truncated data. New York: Springer-Verlag.

Kleinbaum DG, Klein M. 2005. Chapter 2: Kaplan–Meier survival curves and the log-rank test. Survival analysis: A self-learning text 2nd edition. New York, NY: Springer. p xv, 590 p.

Klöppel S, Chu C, Tan GC, Draganski B, Johnson H, Paulsen JS, Kienzle W, Tabrizi SJ, Ashburner J, Frackowiak RS, et al. 2009. Automatic detection of preclinical neurodegeneration: Presymptomatic Huntington disease. Neurology 72(5):426–431.

Kremer B, Goldberg P, Andrew SE, Theilmann J, Telenius H, Zeisler J, Squitieri F, Lin B, Bassett A, Almqvist E, et al. 1994. A worldwide study of the Huntington's disease mutation. The sensitivity and specificity of measuring CAG repeats. N Engl J Med 330(20):1401–1406.

Langbehn DR, Brinkman RR, Falush D, Paulsen JS, Hayden MR. 2004. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. Clin Genet 65(4):267–277.

Langbehn DR, Hayden MR, Paulsen JS, Investigators P-H. 2010. CAG-repeat length and the age of onset in Huntington disease (HD): A review and validation study of statistical approaches. Am J Med Genet Part B 153B:397–408.

Langbehn DR, Paulsen JS. 2007. Predictors of diagnosis in Huntington disease. Neurology 68(20):1710–1717.

Laupacis A, Sekar N, Stiell IG. 1997. Clinical prediction rules. A review and suggested modifications of methodological standards. JAMA 277(6):488–494.

Lucotte G, Turpin JC, Riess O, Epplen JT, Siedlaczk I, Loirat F, Hazout S. 1995. Confidence intervals for predicted age of onset, given the size of (CAG)n repeat, in Huntington's disease. Hum Genet 95(2):231–232.

Maat-Kievit A, Losekoot M, Zwinderman K, Vegter-van derVlis M, Belfroid R, Lopez F, Van Ommen GJ, Breuning M, Roos R. 2002. Predictability of age at onset in Huntington disease in the Dutch population. Medicine (Baltimore) 81(4):251–259.

Miller ME, Langefeld CD, Tierney WM, Hui SL, McDonald CJ. 1993. Validation of probabilistic predictions. Med Decis Making 13(1):49–58.

Nance MA, Mathias-Hagen V, Breningstall G, Wick MJ, McGlennen RC. 1999. Analysis of a very large trinucleotide repeat in a patient with juvenile Huntington's disease. Neurology 52(2):392–394.

Nopoulos PC, Aylward EH, Ross CA, Johnson HJ, Magnotta VA, Juhl AR, Pierson RK, Mills J, Langbehn DR, Paulsen JS. 2010. Cerebral cortex structure in prodromal Huntington disease. Neurobiol Dis 40(3): 544–554.

Nopoulos PC, Aylward EH, Ross CA, Mills JA, Langbehn DR, Johnson HJ, Magnotta VA, Pierson RK, Beglinger LJ, Nance MA, et al. 2011. Smaller

intracranial volume in prodromal Huntington's disease: Evidence for abnormal neurodevelopment. Brain 134(Pt1):137–142.

Paulsen J. 2010. Early detection of Huntington's disease. Future Neurol 5(1):85–104.

Paulsen JS, Hayden M, Stout JC, Langbehn DR, Aylward E, Ross CA, Guttman M, Nance M, Kieburtz K, Oakes D, et al. 2006. Preparing for preventive clinical trials: The Predict-HD study. Arch Neurol 63(6):883–890.

Paulsen JS, Langbehn DR, Stout JC, Aylward E, Ross CA, Nance M, Guttman M, Johnson S, MacDonald M, Beglinger LJ, et al. 2008. Detection of Huntington's disease decades before diagnosis: The Predict-HD study. J Neurol Neurosurg Psychiatry 79(8):874–880.

Paulsen JS, Nopoulos PC, Aylward E, Ross CA, Johnson H, Magnotta VA, Juhl A, Pierson RK, Mills J, Langbehn D, et al. 2010a. Striatal and white matter predictors of estimated diagnosis for Huntington disease. Brain Res Bull 82(3–4):201–207.

Paulsen JS, Wang C, Duff K, Barker R, Nance M, Beglinger L, Moser D, Williams JK, Simpson S, Langbehn D, et al. 2010b. Challenges assessing clinical endpoints in early Huntington disease. Mov Disord 25(15): 2595–2603.

Penney JB Jr, Vonsattel JP, MacDonald ME, Gusella JF, Myers RH. 1997. CAG repeat number governs the development rate of pathology in Huntington's disease. Ann Neurol 41(5):689–692.

R Development Core Team. 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0.

Rowe KC, Paulsen JS, Langbehn DR, Duff K, Beglinger LJ, Wang C, O'Rourke JJ, Stout JC, Moser DJ. 2010. Self-paced timing detects and tracks change in prodromal Huntington disease. Neuropsychology 24(4):435–442.

Rubinsztein DC, Leggo J, Chiano M, Dodge A, Norbury G, Rosser E, Craufurd D. 1997. Genotypes at the GluR6 kainate receptor locus are associated with variation in the age of onset of Huntington disease. Proc Natl Acad Sci USA 94(8):3872–3876.

Rubinsztein DC, Leggo J, Coles R, Almqvist E, Biancalana V, Cassiman JJ, Chotai K, Connarty M, Crauford D, Curtis A, et al. 1996. Phenotypic characterization of individuals with 30–40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36–39 repeats. Am J Hum Genet 59(1):16–22.

Squitieri F, Sabbadini G, Mandich P, Gellera C, Di Maria E, Bellone E, Castellotti B, Nargi E, de Grazia U, Frontali M, et al. 2000. Family and molecular data for a fine analysis of age at onset in Huntington disease. Am J Med Genet 95(4):366–373.

Stine OC, Pleasant N, Franz ML, Abbott MH, Folstein SE, Ross CA. 1993. Correlation between the onset age of Huntington's disease and length of the trinucleotide repeat in IT-15. Hum Mol Genet 2(10):1547–1549.

Stout JC, Paulsen JS, Queller S, Solomon AC, Whitlock KB, Campbell JC, Carlozzi N, Duff K, Beglinger LJ, Langbehn DR, et al. 2011. Neurocognitive signs in prodromal Huntington disease. Neuropsychology 25(1):1–14.

Therneau T, original Splus->R, port by, Thomas Lumley. 2009. survival: Survival analysis, including penalised likelihood. R package version 2.35-8.

Watson D, Wu KD. 2005. Development and validation of the schedule of compulsions, obsessions, and pathological impulses (SCOPI). Assessment 12(1):50–65.

Wolkewitz M, Allignol A, Schumacher M, Beyersmann J. 2010. Two pitfalls in survival analyses of time-dependent exposure: A case study in a cohort of Oscar nominees. Am Stat 64(3):205–211.

Zelen M, Feinleib M. 1969. On the theory of screening for chronic diseases. Biometrika 56(3):601–614.

Zelen M. 2005. Forward and backward recurrence times and length biased sampling: Age specific models. Lifetime Data Anal 10(4):325–334.