Reliability and Factor Structure of the Short Problem Behaviors Assessment for Huntington's Disease (PBA-s) in the TRACK-HD and REGISTRY studies

Jenny Callaghan, B.Sc., Cheryl Stopford, Ph.D., Natalie Arran, B.Sc., D.ClinPsy, Marie-Francoise Boisse, Allison Coleman, B.Sc., Rachelle Dar Santos, B.Sc., Eve M. Dumas, Ph.D., Ellen P. Hart, M.Sc., Damian Justo, Ph.D., Gail Owen, Ph.D., Joy Read, M.Sc., Miranda J. Say, D.Psych, Alexandra Durr, M.D., Ph.D., Blair R. Leavitt, M.D., C.M., Raymund A. C. Roos, M.D., Ph.D., Sarah J. Tabrizi, M.B., Ch.B., Ph.D., Anne-Catherine Bachoud-Levi, M.D., Ph.D., Catherine Bourdet, M.D., Ph.D., Erik van Duijn, M.D., Ph.D., David Craufurd, M.B., B.S., M.Sc.

The authors report the inter-rater reliability and factor structure of the Short Problem Behaviors Assessment (PBA-s), a semistructured interview to measure severity and frequency of behavioral problems in Huntington's disease. Video recordings of 410 PBA-s interviews were rescored by an independent rater, and Cohen's kappa calculated to assess inter-rater reliability. The mean kappa was 0.74 for severity and 0.76 for frequency scores, whereas weighted kappa (allowing scores to differ by 1 point) was 0.94 for severity and 0.92 for frequency scores. The results of factor analysis were consistent with previous studies using other measures. The authors conclude that the PBA-s is a reliable measure.

Huntington's disease (HD) is an inherited neurodegenerative disorder characterized by progressive motor and cognitive decline, and a variety of behavioral abnormalities. Onset is usually during the fourth or fifth decade of life, and the condition follows a progressive course over 15–25 years. Although attention has traditionally focused on the movement disorder, recent research has highlighted the major impact of behavioral symptoms on HD sufferers' functional capacity and quality of life.[1–3] Behavioral problems in HD include apathy, irritability, and depression[4–8] but these are typically rather variable, and the affective symptoms in particular do not correlate well with overall disease progression.[2,9]

The Problem Behaviors Assessment for HD (PBA-HD)[4] is a 40-item semistructured interview designed to elicit information about behavioral symptoms relevant to HD. The short version (PBA-s) was developed by the Behavioral Phenotype Working Group of the European Huntington's Disease Network (EHDN) for use in the REGISTRY study,[10] where behavioral symptoms were not the primary focus and a shorter instrument was required. The PBA-s contains 11 items, each measuring a different behavioral problem which is rated for both severity and frequency on a 5-point scale; severity and frequency ratings are then multiplied to provide an overall score for each symptom. Interviews are conducted with the patient and a knowledgeable informant such as a spouse or professional carer; the final rating being made in the light of all available information

including the interviewer's own observations of the patient's behavior.

The reliability of the original 40-item PBA-HD was described in 2001.[4] Thirty-eight patients in Manchester were interviewed sequentially by two trained PBA raters who achieved good levels of inter-rater agreement, with mean weighted kappa values of 0.86 for severity and 0.84 for frequency scores. Test-retest reliability was also assessed in a subset of 15 patients by repeating the interview with the same rater 2 weeks later, yielding mean weighted kappas of 0.94 for severity and 0.92 for frequency scores. Kingma et al.[8] replicated the initial inter-rater reliability study in a larger sample in The Netherlands, using a Dutch translation of the PBA-HD; they reported mean weighted kappas of 0.82 for severity and 0.73 for frequency scores. Both studies[4,8] included a factor analysis and obtained very similar results.

REGISTRY is a longitudinal observational study conducted at more than 100 EHDN sites in 18 European countries.[10] Patients with manifest HD and presymptomatic gene carriers are evaluated annually using a battery of motor, cognitive, functional, and behavioral assessments; the latter include either the Unified Huntington's Disease Rating Scale (UHDRS) behavioral assessment or the PBA-s. Participants in version 3 of REGISTRY at the Manchester site completed the PBA-s. TRACK-HD is another multicenter, longitudinal observational study utilizing a more extensive battery of motor, cognitive,

TABLE 1. Numbers of PBA-s Interviews Rescored at Each Site

	Leiden	London	Paris	Vancouver	Manchester
TRACK-HD Baseline assessment	34	43	86	46	0
TRACK-HD 12-month follow-up	28	46	0	46	0
TRACK-HD 24-month follow-up	0	1	0	5	0
TRACK-HD 36-month follow up	0	1	0	1	0
REGISTRY study	0	0	0	0	73
Total	62	91	86	98	73

behavioral, and imaging assessments (including the PBA-s) to assess the utility of these measures as potential biomarkers for clinical trials.[2,11] TRACK-HD was designed and conducted as if it were a clinical trial, with rigorous quality control and blinded data analysis. All PBA-s interviews in the TRACK-HD study were video recorded, and a proportion were rescored by an independent rater for quality control purposes.

The aims of the present study were 1) to assess the inter-rater reliability of the PBA-s in three languages (English, Dutch, and French) using the TRACK-HD recordings, and 2) to compare the factor structure of the PBA-s to that extracted from the PBA-HD in a very similar clinical population using the Manchester REGISTRY data.

METHODS

Participants

A total of 366 individuals participated in the TRACK-HD study, at four sites: Leiden (The Netherlands), London (UK), Paris (France), and Vancouver (Canada). Subjects were divided into three groups: 123 early manifest HD patients (disease stage I and II, equating to a Total Functional Capacity Score[12] of 7–13), 120 premanifest HD gene carriers, and 123 controls. Premanifest participants all had a total UHDRS motor score of less than 5 and a disease burden score greater than 250; the latter was calculated from the formula [age × (CAG-35.5)] and gives an estimation of lifetime exposure to the disease, without predicting its course.[13] The 123 control participants were age- and gender- matched to the combined gene-positive groups, and were either spouses, partners, or gene-negative siblings of the premanifest and early HD patients. Subjects attended an annual assessment day, which included a cognitive battery, MRI scan, motor, functional and quality of life assessments, and a behavioral interview (the PBA-s). Baseline assessments were conducted between January and August 2008, with follow-up at 12, 24, and 36 months. Written informed consent was obtained from all participants, including consent to video record parts of the assessment for quality control, research, and training purposes. The study was described in detail by Tabrizi et al.[11]

A total of 732 TRACK-HD PBA-s interviews were conducted during the first 2 years, of which 329 were selected for rescoring. In addition, follow-up assessments from subsequent years were screened for severity scores of ≥2 for the least-commonly reported symptoms (paranoid thinking, hallucinations, disorientation), to improve reliability assessment of these items; eight such interviews were identified.

A total of 230 manifest HD patients attending the Manchester REGISTRY site for baseline assessment were included in the factor analysis. Recorded PBA-s interviews were available from 73 individuals and included in the reliability study because they encompassed a wider range of disease severity and PBA-s scores. Written informed consent to have their interviews video recorded for research and training purposes was obtained from all patients. The reliability study, thus, included data from 410 PBA-s interviews, conducted with a total of 308 participants (Table 1).

The PBA-s

The version of the PBA-s used initially at the baseline assessments comprised 10 items taken from the original 40-item PBA-HD: depression, suicidal ideation, anxiety, irritability, aggressive behavior, apathy, perseverative thinking, paranoid thinking, hallucinations, and behavior suggestive of disorientation. A list of all the items included in the original PBA-HD is shown in Table 2 for comparison. An additional PBA-HD item measuring obsessive-compulsive symptoms was added to the PBA-s at the 12-month

TABLE 2. Item Structure of the Original PBA-HD

Depressed mood, sadness[a]	Inflexible, uncooperative
Initial insomnia (difficulty falling asleep)	Passive, too compliant
Wakening too early	Obsessional thoughts or mental images[c]
Sleepy or drowsy during daytime	Compulsive behaviors[c]
Pessimistic, low self-esteem	Perseverative thinking or behavior[a]
Anxiety[a]	Irritability[a]
Tense, unable to relax	Verbal aggression[d]
Suicidal ideation[a]	Physical aggression[d]
Lack of energy, fatigue	Medically unexplained symptoms
Lacks interest in appearance, self-neglect[b]	Loss of libido
Eating less	Sexually disinhibited behavior
Eating more	Sexually demanding behavior
Eats too quickly, bolting food	Paranoid thinking/delusions[a]
Change in food preference (e.g., sweet tooth)	Delusional jealousy
Needs prompting to do things[b]	Auditory hallucinations[e]
Fails to complete tasks[b]	Visual hallucinations[e]
Poor self-monitoring, fails to correct errors	Tactile hallucinations[e]
Impulsive, makes poor decisions	Olfactory hallucinations[e]
Reduced rapport, emotionally blunted[b]	Gustatory hallucinations[e]
Self-centered, demanding	Abnormal behavior re temperature regulation

[a] These items were retained in the PBA-s.
[b] These items were incorporated into the scoring criteria for Apathy in PBA-s.
[c] Obsessional thoughts and compulsions were combined as a single item in PBA-s.
[d] Verbal and Physical Aggression combined as a single item in PBA-s.
[e] These items were combined into a single hallucinations item in PBA-s.

follow-up interview, to help raters differentiate between perseverative phenomena and obsessive-compulsive behaviors. Interviews were conducted jointly with the subject and a knowledgeable informant (usually the spouse or partner) who was then briefly re-interviewed afterward to elicit any additional information which (s)he had felt unable to discuss openly during the joint interview; raters were instructed to make an overall judgment based on all available information, including their own observations as well as the answers provided by subject and informant. Suggested prompt questions were provided for each of the 11 items, and raters were instructed to ask whatever additional questions they deemed necessary until they had enough information to determine the correct scores. Severity and frequency during the previous 4 weeks were rated separately for each symptom on a 5 point (0–4) scale; these were then multiplied to yield an overall symptom score for each item. The criteria for rating frequency were the same for all items, and ranged from 0 (not present) to 4 (present all the time). Detailed criteria for rating severity were provided for each symptom, based on the following general principles: 0=symptom absent; 1=trivial, or rater is not completely convinced that the symptom is present; 2=symptom definitely present, but not causing a practical problem or interfering with daily life; 3=symptom is causing problems or distress; 4=symptom causing severe problems or making normal life impossible.

All the TRACK-HD interviewers were trained in person during a site investigator meeting at the outset of the study, which involved watching and scoring videos of PBA-s interviews and discussing scores. PBA-s interviews were video recorded and uploaded to the TRACK-HD web portal for rescoring by an independent expert rater (JC) who was familiar with the long-version PBA-HD. Any discrepancies were resolved by a third expert rater (DC) and feedback provided by telephone or e-mail to the original interviewer. Once a site rater was judged to have achieved an acceptable performance, the proportion selected for rescoring was reduced to 25%, and eventually 10%; site raters were blind to which interviews would be selected. In the case of interviews conducted in languages other than English, rescoring was carried out by expert raters (CB, M-FB, EvD) who were native French or Dutch speakers respectively,

TABLE 3. Percentage of Interviews Where the Initial Rater Gave a Severity Score of at Least 1

	English	Dutch	French
Depression	57.2	48.4	54.7
Suicidal ideation	13.2	10.9	14.0
Anxiety	64.6	64.1	74.4
Irritability	64.1	51.6	57.0
Aggression	50.4	40.6	30.2
Apathy	47.2	26.6	40.7
Obsessive-compulsive behaviors	14.8	0.0	n/a
Perseveration	38.6	24.2	25.6
Paranoid thinking/delusions	5.1	3.2	1.2
Hallucinations	3.1	0.0	0.0
Disorientation	15.6	11.3	5.8

and later on by an alternate PBA-s rater at the Leiden site. In all cases, the rater doing the rescoring was blind to the scores assigned during the original interview, and to the group status (affected, premanifest, or control) of the patient. The Manchester interviewers were trained in a similar fashion, and rescored by the same expert rater (JC) as the English TRACK-HD videos.

Statistical Analyses

Inter-rater reliability was assessed using the Cohen's kappa (k) statistic to correct for the possibility of chance agreement.[14] Weighted kappa (k_w) was devised as a way of describing inter-rater reliability in situations where some disagreements between raters are of greater gravity than others[15]; for example, in the case of an interview like the PBA-s, a symptom close to the borderline between a 'mild' or 'moderate' score could be rated as '2' or '3', but should not receive a score of '1' or '4'. The 'clinically significant' k_w used here accepts cases where two raters differ by ±1 point as agreement, whereas k requires both raters to give exactly the same score. Both k and k_w were calculated for the present study.

Factor Analysis

To investigate the factor structure of the PBA-s, we conducted a principal components analysis with Varimax rotation, using REGISTRY 3 data from Manchester rather than TRACK-HD because the range of disease severity more closely resembled

TABLE 4. Percentage of Participants Scoring Severity and Frequency Levels of 0–4 by Item (Numbers of Valid Interviews in brackets)

	Severity Score					Frequency Score				
Symptom	0	1	2	3	4	0	1	2	3	4
Depression	47.3 (194)	11.2 (46)	32.9 (135)	7.3 (30)	1.2 (5)	47.1 (193)	22.4 (92)	17.1 (70)	10.5 (43)	2.9 (12)
Suicide	88.1 (362)	3.9 (16)	6.3 (26)	1.5 (6)	0.0 (0)	88.0 (360)	6.6 (27)	2.7 (11)	1.7 (7)	0.7 (3)
Anxiety	36.7 (150)	22.2 (91)	28.4 (116)	11.7 (48)	1.0 (4)	36.8 (150)	24.3 (99)	21.6 (88)	14.7 (60)	2.7 (11)
Irritability	42.4 (174)	29.5 (121)	15.4 (63)	11.2 (46)	1.5 (6)	42.5 (174)	21.0 (86)	26.2 (107)	7.8 (32)	2.4 (10)
Anger	59.8 (245)	13.2 (54)	13.7 (56)	10.2 (42)	3.2 (13)	59.8 (245)	30.7 (126)	7.6 (31)	1.5 (6)	0.5 (2)
Apathy	59.2 (239)	8.9 (36)	17.8 (72)	9.7 (39)	2.0 (8)	59.5 (239)	7.2 (29)	8.0 (32)	8.5 (34)	14.4 (58)
OCD	88.9 (128)	3.5 (5)	6.3 (9)	1.4 (2)	0.0 (0)	90.8 (128)	2.1 (3)	0.0 (0)	5.0 (7)	2.1 (3)
Perseveration	68.8 (278)	8.7 (35)	15.3 (62)	5.2 (21)	1.2 (5)	69.3 (278)	8.5 (34)	9.7 (39)	10.0 (40)	1.7 (7)
Delusions	96.8 (392)	1.5 (6)	1.5 (6)	0.2 (1)	0.0 (0)	96.8 (392)	2.0 (8)	1.0 (4)	0.2 (1)	0.0 (0)
Hallucinations	98.0 (397)	0.2 (1)	0.7 (3)	0.5 (2)	0.5 (2)	98.0 (397)	1.0 (4)	0.7 (3)	0.2 (1)	0.0 (0)
Disorientation	88.4 (328)	8.4 (31)	2.7 (10)	0.3 (1)	0.0 (0)	88.4 (328)	5.9 (22)	2.7 (10)	1.6 (6)	1.1 (4)

TABLE 5. Cohen's Kappa and 'Clinically Significant' Weighted Kappa for Rater Pairs and Mean k and k_w Values for the English, French, and Dutch Speaking Sites

Rater Pair	Severity Scores		Frequency Scores	
	k	k_w	k	k_w
Rater pair 1	0.858	1	0.909	0.991
Rater pair 2	0.795	0.976	0.892	0.973
Rater pair 3	0.606	0.925	0.664	1
Rater pair 4	0.772	1	0.877	1
Rater pair 5	0.659	0.93	0.741	0.944
Rater pair 6	0.594	0.917	0.683	0.884
Rater pair 7	0.785	0.99	0.831	0.989
Rater pair 8	0.623	0.967	0.734	0.916
Rater pair 9	0.685	0.94	0.682	0.83
Rater pair 10	0.648	0.899	0.624	0.802
Rater pair 11	0.735	0.926	0.727	0.923
Rater pair 12	0.566	0.712	0.561	0.705
Rater pair 13	1.000	1.000	0.867	0.960
Rater pair 14	0.818	0.948	0.824	0.965
Rater pair 15	1.000	1.000	0.910	0.977
Rater pair 16	0.694	0.911	0.699	0.927
English mean	0.767	0.964	0.803	0.961
Dutch mean	0.692	0.913	0.676	0.863
French mean	0.626	0.826	0.622	0.768

the population used for the original PBA-HD factor analysis. The symptoms 'paranoid thinking' and 'hallucinations' were excluded because fewer than 3% of subjects scored more than zero on these items, leaving nine remaining PBA-s variables for inclusion, and a ratio of>25 subjects per variable. There are no clearly established rules to determine the number of subjects required for a satisfactory factor analysis, but a ratio of at least 20 subjects per variable appears to be a conservative estimate from the published literature.[16]

RESULTS

The proportion of interviews in each language in which the subject exhibited each of the behavioral symptoms measured by the PBA-s is given in Table 3, while the distribution of scores across items (number of participants scoring 0–4 by symptom) is shown in Table 4.

Inter-rater Reliability—Raw Agreement
The percentage of severity and frequency ratings where both the interviewer and rescoring rater agreed was calculated for each of the 16 pairs of raters. On average, raters agreed completely on 87.7% of severity scores (range 76.9%–100%)

and 88.6% of frequency scores (range 80.2%–96.1%). Using the more realistic criterion of agreement within ±1 point, the average percentage agreement for severity scores was 98.9% (range 97.4%–100%), and for frequency scores 98.3% (range 94%–100%). The corresponding figures for the English-speaking sites only (London, Manchester, and Vancouver) were 88.4% for severity scores (99.3% within ±1 point) and 89.8% for frequency scores (99.2% within ±1e point). There was very little difference in the percentage of 'clinically significant' agreement between symptoms, or between languages.

Inter-rater Reliability–Cohen's Kappa
The mean k was 0.74 for severity and 0.76 for frequency ratings. Mean k_w was 0.94 for severity and 0.92 for frequency ratings. Table 4 shows the k and k_w values for each of the 16 pairs of raters, and the mean k and k_w values for each language. When the analysis was restricted to English-speaking sites only, the mean values of k were 0.77 for severity and 0.80 for frequency scores, while the mean k_w values were 0.96 for severity and 0.96 for frequency.

Factor Analysis
After excluding the 'paranoid thinking' and 'hallucinations' items, the remaining data satisfied conventional requirements for factor analysis (Kaiser-Meyer-Olkin measure of sampling adequacy=0.60; Bartlett's test of sphericity p <0.001). Principal components analysis extracted three factors with Eigenvalues greater than one, collectively explaining 57.4% of total variance. The factor loadings after Varimax rotation are shown in Table 6.

DISCUSSION

We found high levels of inter-rater agreement in both severity and frequency ratings of all PBA-s items across all three languages included in the study. According to Landis and Koch's 1977 interpretations of k values,[17] all 16 pairs of raters achieved at least 'moderate agreement' (0.41–0.60) on severity and frequency scores, whereas most achieved at least 'substantial agreement' (0.61–0.80). The high values of k_w for both severity and frequency scores show that very few disagreements were by more than 1 point. The findings are consistent with the high levels of inter-rater reliability in the PBA-HD reported by previous studies.[4,8]

The percentage of scores within 1 point of each other was similar for all PBA-s items; however, the prevalence of some symptoms (paranoid thinking, hallucinations, and disorientated behavior) was low, which may have produced artificially high levels of agreement because it is easier to achieve agreement between raters when the symptom is absent. With the exception of paranoid thinking and hallucinations, all the PBA-s items in the English-language

TABLE 6. Results of Principle Components Analysis of REGISTRY 3 Data Showing Factor Loadings of Individual PBA-s Items (% Variance Accounted for Shown in Brackets)

Factor 1: Apathy (16.9)	Factor 2: Irritability (26.9)	Factor 3: Affective (13.6)
Lack of initiative (Apathy)	Irritability	Depressed mood
0.755	0.894	0.789
Perseverative thinking or behavior	Angry or aggressive behavior	Suicidal ideation
0.732	0.929	0.664
Disoriented behavior		Anxiety
0.633	0.656	

interviews received a score of 1 or more in at least 10% of interviews; if suicidal ideation, behavior suggesting disorientation, and obsessive-compulsive behaviors are also excluded, the remaining six items were given a positive score in 35%–65% of interviews. A similar pattern of responses was seen in all three languages, with the exception of obsessive-compulsive behaviors where there were no Dutch subjects who scored more than zero, while all the French-language interviews were taken from TRACK-HD visit 1 when obsessive-compulsive behaviors were not included in the instrument. Further research collecting data with positive scores on the psychosis and disorientation items will be needed to fully confirm the inter-rater reliability of these items.

The weighted kappa statistic used in the present study is probably the most appropriate measure of inter-rater agreement for an instrument of this type; by this criterion (agreement within 1 point or better) all but one pair of raters had k_w scores for both severity and frequency ratings in the range (0.81–1.0) conventionally regarded as indicating 'almost perfect agreement.'[17] Even by the more demanding criterion of complete agreement, all but two pairs of raters produced k values in the 'substantial agreement' (0.61–0.80) range or better, and none were unacceptable. The mean values of k_w for both the English-language version and the whole sample were in the 'almost perfect agreement' range, while both were toward the upper end of the 'substantial agreement' range using the more stringent k statistic. Comparison of individual item scores and between languages suggests that no symptom or language version performed poorly relative to the others. The method of rescoring videos employed here means that our analysis of reliability is limited to application of the scoring criteria rather than variations in interview technique, but the latter is likely to be less important.

The results of the factor analysis, in a population very similar to the one used for the original PBA-HD factor analysis, indicate a three factor solution corresponding to Irritability (PBA-s 'irritability' and 'aggression' scores), Apathy (PBA-s 'apathy', 'perseveration' and 'behavior suggestive of disorientation' scores), and Affect (PBA-s 'depressed mood', 'anxiety' and 'suicidal ideation' scores), which is consistent with the original PBA-HD factor structure. Taken together, these data (in combination with previous research using the original PBA-HD) provide strong evidence that the PBA-s is a reliable instrument for assessing behavioral problems in HD.

AUTHOR AND ARTICLE INFORMATION

From the Cardiff University, School of Biosciences, Cardiff, United Kingdom (JC). Institute of Human Development, Faculty of Medicine and Human Sciences, University of Manchester and Manchester Academic Health Science Centre, Manchester, United Kingdom (CS, NA, DC); Manchester Centre for Genomic Medicine, Central Manchester University Hospitals NHS Foundation Trust, Manchester, United Kingdom (CS, DC); Dept. of Neuropsychology, Salford Royal Hospital NHS Foundation Trust, Salford, United Kingdom (NA); Centre de Référence National pour la Maladie de Huntington, AP-HP, Groupe Henri Mondor-Albert Chenevier, Créteil, France (M-FB, A-CB-L, CB); University of British Columbia, The Centre for Huntington's disease, Vancouver, Canada (AC, RDS); Leiden University Medical Centre, Department of Neurology, Leiden, The Netherlands (EMD, EPH, RACR); UPMC UnivParis06, UMR_S975 and Inserm, UMR_S975, CRICM and CNRS UMR7225, F-75013, Paris, France (DJ, AD); University College London, Institute of Neurology, London, United Kingdom (GO, JR, MJS, SJT); Brain and Spine Institute (ICM), F-75013, Paris, France (AD); AP-HP Pitie-Salpetriere Hospital, Dept. of Genetics and Cytogenetics, F-75013, Paris, France (AD); University of British Columbia, Centre for Molecular Medicine and Therapeutics, Vancouver, Canada (BRL); Service de Psychiatrie, 93G18, EPSVE, 93330 Neuilly-sur-Marne, France (CB); and Dept. of Psychiatry, Leiden University Medical Centre, Leiden, The Netherlands (EvD).

Send correspondence to Dr. Craufurd; e-mail: david.craufurd@manchester.ac.uk

REFERENCES

1. Hamilton JM, Salmon DP, Corey-Bloom J, et al: Behavioural abnormalities contribute to functional decline in Huntington's disease. J Neurol Neurosurg Psychiatry 2003; 74:120–122
2. Tabrizi SJ, Scahill RI, Owen G, et al: TRACK-HD Investigators: 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 Neurol 2013; 12:637–649
3. Read J, Jones R, Owen G, et al.: Quality of life in Huntington's Disease: a comparative study investigating the impact for those with premanifest and early manifest disease, and their partners. J Huntington's Dis 2013; 2:159–175
4. Craufurd D, Thompson JC, Snowden JS: Behavioral changes in Huntington Disease. Neuropsychiatry Neuropsychol Behav Neurol 2001; 14:219–226
5. Naarding P, Kremer HP, Zitman FG: Huntington's disease: a review of the literature on prevalence and treatment of neuropsychiatric phenomena. Eur Psychiatry 2001; 16:439–445
6. Rickards H, De Souza J, van Walsem M, et al: European Huntington's Disease Network: Factor analysis of behavioural symptoms in Huntington's disease. J Neurol Neurosurg Psychiatry 2011; 82:411–412
7. van Duijn E: Treatment of Irritability in Huntington's Disease. Curr Treat Options Neurol 2010; 12:424–433
8. Kingma EM, van Duijn E, Timman R, et al: Behavioural problems in Huntington's disease using the Problem Behaviours Assessment. Gen Hosp Psychiatry 2008; 30:155–161
9. Thompson JC, Harris J, Sollom AC, et al: Longitudinal evaluation of neuropsychiatric symptoms in Huntington's disease. J Neuropsychiatry Clin Neurosci 2012; 24:53–60
10. Orth M, Handley OJ, Schwenke C, et al: Investigators of the European Huntington's Disease Network: Observing Huntington's Disease: the European Huntington's Disease Network's REGISTRY. PLoS Curr 2010; 2:RRN1184
11. Tabrizi SJ, Langbehn DR, Leavitt BR, et al: TRACK-HD investigators: Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. Lancet Neurol 2009; 8:791–801

12. Shoulson I, Fahn S: Huntington disease: clinical care and evaluation. Neurology 1979; 29:1–3
13. Langbehn DR, Brinkman RR, Falush D, et al: International Huntington's Disease Collaborative Group: A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. Clin Genet 2004; 65:267–277
14. Cohen J: A coefficient of agreement for nominal scales. Educ Psychol Meas 1960; 20:37–46
15. Cohen J: Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968; 70:213–220
16. Hair J, Anderson RE, Tatham RL, et al: Multivariate data analysis, 4th ed. New Jersey, Prentice-Hall Inc., 1995
17. Koch GG, Landis JR, Freeman JL, et al: A general methodology for the analysis of experiments with repeated measurement of categorical data. Biometrics 1977; 33:133–158