NINDS Huntington's Disease Common Data Element (CDE) Recommendations
Cognitive Subgroup
Summary Statement

The aim of the NINDS HD CDEs Cognitive working group was to recommend cognitive instruments for use in clinical trials of manifest and premanifest HD. An extensive preliminary list of instruments for consideration was generated based on tasks that have been used in multi-centre studies of HD, including TRACK-HD, PREDICT-HD, CAPIT-HD, REGISTRY and COHORT. Our final recommendations are based on the knowledge and expertise of the working group members, taking into account the available evidence, both published and unpublished.

Many cognitive outcome measures yield strong cross-sectional effect sizes when individuals with manifest or premanifest HD are compared with healthy controls and cross-sectional data may be informative for researchers investigating symptomatic rather than disease modifying treatments. However, in evaluating the available evidence, our strongest recommendations are reserved for measures that have shown both cross-sectional and longitudinal effects, and are therefore suitable for treatments aimed at modifying the course of disease.

Recommended instruments have been classified as core, supplementary and exploratory:

Core instruments are strongly recommended for use in clinical studies and have the best evidence of longitudinal sensitivity in manifest and / or premanifest HD. We recommend that investigators use at least one of the core recommendations.

The majority of supplemental instruments also show good longitudinal and cross sectional sensitivity. Our recommendations in this category cover a wide range of cognitive abilities including psychomotor function, memory (verbal learning and visual working memory), sequencing, emotion processing and executive function.

Exploratory instruments are those for which preliminary evidence indicates good cross-sectional sensitivity but for which longitudinal data is not currently available.

The data upon which the recommendations are based are primarily from studies of premanifest and early-to-mid stage HD, since there is limited evidence of longitudinal sensitivity of cognitive measures in advanced HD. Although many of the tasks may be suitable for use in more advanced disease stage there is limited evidence regarding longitudinal sensitivity. Similarly, the recommendations apply to adult-onset HD populations since there is a lack of evidence regarding sensitivity of cognitive measures in juvenile-onset HD.

It is important that the cognitive instruments are administered by an individual with an appropriate level of training and experience. We recommend that the instruments are administered by or under the supervision of a suitably qualified psychologist/neuropsychologist. However, if this is not possible the administrator should consult an appropriately qualified psychologist/ neuropsychologist.

It is important to consider the issue of practice effects when designing any study that involves serial cognitive assessments, particularly when these occur over relatively short intervals. Methods to control for / minimise practice effects include the use of alternate forms or the inclusion of repeated baseline exposure to instruments prior to the study intervention (e.g. Beglinger et al., 2005). An alternative approach is to utilise the practice effect size as a predictor of cognitive outcome (Duff et al., 2007).

Where available, data on the reliability and validity has been included in the instrument summaries. However, it should be noted that this information is largely derived from non-HD, non-patient populations. Effects of age,

gender and education are commonly observed in neuropsychological measures. Where normative data indicate particular effects these have been noted in the instrument summaries. However, it is important that researchers are mindful of these potential confounds when matching groups and in data analysis.

We have included as a supplemental measure the Montreal Cognitive Assessment (MoCA), which is a brief cognitive screening instrument. Although there is currently limited data on this instrument in premanifest HD and no published longitudinal data the available evidence indicates that the MoCA is significantly more sensitive to cognitive dysfunction in HD than the commonly used Mini Mental State Examination. This is currently our 'best recommendation' for a brief cognitive screening assessment in HD.

Beglinger LJ, Gaydos B, Tangphao-Daniels O, Duff K, Kareken DA, Crawford J, Fastenau P, Siemers E. (2005) Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of Clinical Neuropsychology,* 20: 517-529.

Duff K, Beglinger LJ, Schulz S, Moser DJ, McCaffrey R, Haase R, Westervelt H, Langbehn DR, Paulsen J. (2007) Practice effects in the prediction of long-term cognitive outcome in three patient samples: a novel prognostic index. *Archives of Clinical Neuropsychology,* 22: 15-24.

Summary Table – Instruments by Evidence and Disease Staging

		Cross-Sectional Evidence of Sensitivity	Cross-Sectional Evidence of Sensitivity	Longitudinal Evidence of Sensitivity	Longitudinal Evidence of Sensitivity	Insufficient evidence but promising based on expert opinion	Insufficient evidence but promising based on expert opinion	Administration Feasible in Disease Stage	Administration Feasible in Disease Stage
Instrument Name	**Classification**	**Pre-Manifest**	**Mild**	**Pre-Manifest**	**Mild**	**Pre-Manifest**	**Mild**	**Pre-Manifest**	**Mild**
Circle Tracing	Supplemental	x	x		x			x	x
Cued Movement Sequence	Supplemental	x		x			x	x	
Emotional Recognition	Supplemental	x	x	x	x			x	x
Hopkins Verbal Learning Test-Revised	Supplemental	x	x					x	x
Map Search Task	Exploratory	x	x			x	x	x	x
Mental Rotation	Exploratory	x	x			x	x	x	x
Montreal Cognitive Assessment	Exploratory					x	x	x	x
Phonemic Verbal Fluency	Supplemental	x	x	x	x			x	x
Self-Paced Tapping	Core	x	x	x	x			x	x

Summary Table – Instruments by Evidence and Disease Staging

		Cross-Sectional Evidence of Sensitivity	Cross-Sectional Evidence of Sensitivity	Longitudinal Evidence of Sensitivity	Longitudinal Evidence of Sensitivity	Insufficient evidence but promising based on expert opinion	Insufficient evidence but promising based on expert opinion	Administration Feasible in Disease Stage	Administration Feasible in Disease Stage
Instrument Name	**Classification**	**Pre-Manifest**	**Mild**	**Pre-Manifest**	**Mild**	**Pre-Manifest**	**Mild**	**Pre-Manifest**	**Mild**
Simple and Two-Choice Reaction Time	Supplemental	x		x			x	x	x
Speeded Tapping Test	Core	x	x	x			x	x	x
Spot the Change	Supplemental	x	x		x			x	x
Stroop Tests	Naming and Reading: Core (select 1); Interference: Supplemental	x	x	x	x			x	x
Symbol Digit Modality Test	Core	x	x	x	x			x	x
Trail Making A and B	Core	x	x	x	x			x	x
Verbal Fluency Tests	Supplemental		x			x	x		

Description of Circle Tracing for HD Common Data Elements

Instrument Name:	Circle Tracing
Classification:	Supplemental
Short Description of Instrument:	Summary/ Overview of Instrument: For a circle-tracing task, patients are instructed to start at the vertical apex of a predrawn annulus (on a tablet laptop or some other technological device) and to trace circles within the annulus as quickly and accurately as possible in the clockwise direction. Patients may have multiple practice trials in order to ensure that they understand the instructions. Also, direct and indirect conditions may be applied (i.e., patients can directly observe their hand and the path they are to follow, or the patient's arm as well as the circle they are to trace are obscured from view). Typically, three trials of direct tracing and three trials of indirect tracing are administered. Each terminates after 45 seconds. Construct measured: Visuomotor integration deficits Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): Assessment of cognitive function in HD cross-sectional and longitudinal studies. Means of administration (paper and pencil, computerized): computerized Location of administration (clinic, home, telephone, etc): Clinic Intended respondent (patient, caregiver, etc.): Patient # of items: N/A # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): Scores are generated by a software program and the score computed is the length (in centimeters of ink laid within the annulus for all trial) and is computed separately for the direct and indirect trials. Other measures are computed by the program but were not as sensitive to disease status (i.e., control compared to premanifest or to early HD). These include amount of ink laid either outside or inside the annulus and moving either away from or toward the annulus (adapted from Lemay et al., 2005). Standardization of scores to a reference population (z scores, T scores, etc): N/A If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc). N/A

Description of Circle Tracing for HD Common Data Elements

Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups. Floor effects may be expected within a moderate- advanced HD sample.
Psychometric Properties	Reliability: Test-retest or intra-interview (within rater) reliability (as applicable): N/A Inter-interview (between-rater) reliability (as applicable): N/A Internal consistency: N/A Statistical methods used to assess reliability: test-retest correlation Validity: Content validity: N/A Construct validity: N/A Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): In the TRACK-HD study, late premanifest HD and early HD performance differed from that of healthy controls, and this was true in both direct and indirect conditions (Say et. al, 2011). Unpublished internal analyses (Stout et al preparation) show that 24 month rates of change in Early HD (but not premanifest HD) differed from rates of change in controls in both conditions. Known Relationships to Other Variables (e.g. gender, education, age, etc): Performance related to age, gender and education; change in performance not related to age, gender or education. Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups): N/A
Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)	Strengths: HD Toolkit meta-analysis suggests that tracing tasks and movement to target tasks have promising cross-sectional effect sizes. The tracing and movement to target tasks tap an error correction mechanism that will likely have minimal redundancy with other measures. Circle tracing has advantages over other movement to target paradigms. For example, the set-up does not require specialized robotics, the duration of the task is short, and the sensitive dependent measure is computationally simple. Weaknesses: Availability (copyright): Adapted by Julie Stout's lab as part of the TRACK-HD study based on Lemay et al (2005) and Shepard and Metzler (1971). http://hdresearch.ucl.ac.uk/completed-studies/track-hd/ Special Requirements for administration: This task requires a tablet laptop with stylus. In the indirect condition only, it requires an additional computer monitor as well as a means of hiding the shoulder, arm, hand, and tablet from the subject's view (e.g., placing the tablet in an open ended box and draping a cape

	from the subject's shoulder over the box). Administration Time: Approximately 10 minutes to administer both conditions Translations available (e.g. Spanish, French, Other languages): English, French, Dutch
References:	Key Reference: Lemay M, Fimbel E, Beuter A, Chouinard S, Richer F. Sensorimotor mapping affects movement correction deficits in early Huntington's disease. Experimental Brain Research 2005, 165(4), 454–460. Other References: Say MJ, et al. Visuomotor integration deficits precede clinical onset in Huntington's disease. Neuropsychologia 2011; 49:264-270.

Description of Cued Movement Sequence (Buttons) for HD Common Data Elements

Instrument Name:	Cued Movement Sequence
Classification:	Supplemental
Short Description of Instrument:	Summary/ Overview of Instrument: The cued movement sequence task requires participants to press circles that are displayed in 12 vertical-pairs along the bottom of a touch screen. One circle of each vertical pair is illuminated at a time, sequentially from left to right. Participants press the circles as they are illuminated. Three cue conditions provide different levels of advance information. In the low-level cue condition, the next circle is illuminated when the finger is lifted from the current circle. In the medium-level condition, the next circle is illuminated when the finger presses the current circle. In the high-level condition, the next button is illuminated as the finger presses the current circle and the circle two over is also illuminated when the finger is lifted from the current circle. Construct measured: Planning and movement sequencing. Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): Assessment of planning and sequencing abilities in HD cross-sectional and longitudinal studies of prodromal HD and HD. Means of administration (paper and pencil, computerized): Computerized. Location of administration (clinic, home, telephone, etc): Clinic Intended respondent (patient, caregiver, etc.): Patient # of items: N/A # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): If a button is incorrectly pressed or pressed twice, the trial is terminated and the occurrence recorded as an error. In each cue condition, mean time to complete a sequence (movement time) and the standard deviation of movement time is recorded for accurate trials. Standardization of scores to a reference population (z scores, T scores, etc): This task has not been standardized. If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc).

Description of Cued Movement Sequence (Buttons) for HD Common Data Elements

Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous. If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups. No ceiling or floor effects.
Psychometric Properties	Reliability: Test-retest or intra-interview (within rater) reliability (as applicable): Inter-interview (between-rater) reliability (as applicable): N/A Internal consistency: Statistical methods used to assess reliability: Validity: Content validity: Construct validity: Stepwise increases in movement times for low, medium, and high-level cue conditions suggest that the experimental manipulation influences planning demands. Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): In published cross-sectional (Stout et al., 2011) and internal analyses (PREDICT-HD), all 3 cue conditions are sensitive to changes in prodromal HD, especially in individuals who are closer to an expected diagnosis. Unpublished internal analyses of 7-year longitudinal data (PREDICT) show longitudinal changes in rate of change in prodromal HD for the low and high cue level conditions. Known Relationships to Other Variables (e.g. gender, education, age, etc): Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups):
Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)	Strengths: Task is highly sensitive to changes in prodromal HD, both cross-sectionally and longitudinally. Task has been tested at sites in the United States, Canada, United Kingdom, Australia, Germany, and Spain. Task is easy to administer. Weaknesses: Touch screens may not always be sensitive to responses. External hardware interface devices may be more reliable for recording responses, and have been used in early studies of HD (see N. Georgiou and colleagues). Availability (copyright): Available in the public domain. This task was adapted for computerized presentation by Julie Stout's lab (julie.stout@monash.edu) as part of the PREDICT-HD study based on Georgiou, Phillips, Chiu, and Bradshaw (1995). Special Requirements for administration: A computer with a touch screen or a hardware interface device similar to ones reported in the literature (see Georgiou and colleagues).

Description of Cued Movement Sequence (Buttons) for HD Common Data Elements

	Administration Time: Varies on the ability of the patient. 6-8 minutes. Translations available (e.g. Spanish, French, Other languages): There are no standardized instructions; the task can be administered in any language.
References:	Key Reference: Georgiou N, Bradshaw JL, Phillips JG, Chiu E, Bradshaw JA. Reliance on advance information and movement sequencing in Huntington's disease. Movement Disorders 1995; 10(4):472-481. Stout, J. C., Paulsen, J. S., Queller, S., Solomon, A. C., Whitlock, K. B., Campbell, J. C. et al. (2011). Neurocognitive signs in prodromal Huntington disease. *Neuropsychology., 25,* 1-14. Other References: Gladwin TE, 't Hart BM, de Jong R. Dissociations between motor-related EEG measures in a cued movement sequence task. Cortex. 2008 May;44(5): 521-536. Epub 2007 Dec 23.

Description of Emotion Recognition for HD Common Data Elements

Instrument Name:	Emotion Recognition
Classification:	Supplemental
Short Description of Instrument:	Summary/ Overview of Instrument: Patients are asked to view a subset of 70 Ekman and Friesen faces on a computer display. For each trial, a photograph of a face depicting an emotional or neutral expression is displayed and seven emotion labels (anger, disgust, fear, happiness, neutral, sadness, surprise) are presented simultaneously beneath the stimulus. Seven practice trials (one for each emotion) are completed and 70 total test trials are then completed. Patients are instructed to decide which emotion the person was feeling based on his/her facial expression and to respond by touching the selected emotion with the dominant index finger. Construct measured: Emotion recognition Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): Assessment of cognitive function in HD cross-sectional and longitudinal studies Means of administration (paper and pencil, computerized): Computerized and pencil and paper versions of the task are available. Location of administration (clinic, home, telephone, etc): Clinic Intended respondent (patient, caregiver, etc.): Patient # of items: N/A # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): The score most supported by the evidence is the number of correct responses summing across the negative emotions (fear, sad, angry and disgust). The number correct for each individual emotions can also be analyzed. Standardization of scores to a reference population (z scores, T scores, etc): N/A If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc). N/A

Description of Emotion Recognition for HD Common Data Elements

Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous scale is used for all scores. Ceiling effects can be expected, especially for the emotions of happiness and surprise (Johnson et al., 2007). If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups.
Psychometric Properties	Reliability: Test-retest or intra-interview (within rater) reliability (as applicable): N/A Inter-interview (between-rater) reliability (as applicable): N/A Internal consistency: N/A Statistical methods used to assess reliability: test-retest correlation Validity: Content validity: N/A Construct validity: N/A Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): These results are for the number of correct for negative emotions. In TRACK-HD, cross sectional differences from controls were found for both premanifest and early HD (Tabrizi et al, 2009); Longitudinal rate of change (annualized over 24 months) differed from change in controls for early HD but not premanifest HD (Tabrizi et al., 2011; Stout et al., in submission). PREDICT-HD detected both cross-sectional differences from controls and longitudinal changes over time (7 years) in premanifest HD. The TRACK-HD premanifest participants may be less likely to show cognitive effects than the PREDICT-HD premanifest participants because 1) they are further from estimated onset based on CAG repeat length and age (Langbehn et al., 2004) and 2) they are potentially less progressed in actuality because the TRACK-HD study excluded premanifest subjects based on UHDRS motor scores >= 5. Generally speaking, cognitive tests will be more effective metrics in studies of premanifest HD when the focus is on subjects that are close to onset. Known Relationships to Other Variables (e.g. gender, education, age, etc): Performance related to age but not to gender and education; change in performance not related to age, gender or education. Medication (i.e., neuroleptic) has shown to modulate emotion recognition abilities in early HD (Labuschagne et al., in submission). Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups): N/A

Description of Emotion Recognition for HD Common Data Elements

Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)	Strengths: This task is engaging for participants Weaknesses: The recommended score (number of negative emotions correctly identified) typically meets assumptions for ANOVA. Subjects tend to score at the maximum number correct on some of the individual emotions. Availability (copyright): Stimuli copyrighted and available at paulekman.com, use stimulus set Pictures of Facial Affect (POFA), copyright 1993. https://face.paulekman.com/face/productdetail.aspx?pid=1 Computerized task was developed by Julie Stout's lab as described in Johnson et al (2007) as part of the TRACK-HD study based on Ekman and Friesen (1976). http://hdresearch.ucl.ac.uk/completed-studies/track-hd/ Special Requirements for administration: A touch computer with a LCD stylus-sensitive screen (e.g., Lenovo ThinkPad X61 tablet PC) and a stylus. Administration Time: Approximately 9 minutes Translations available (e.g. Spanish, French, Other languages): French, Dutch and English
References:	Key Reference: Johnson SA, et al. Beyond disgust: impaired recognition of negative emotions prior to diagnosis in Huntington's disease. Brain 2007. 130:732-744. Other References: Ekman P, Friesen WV. Pictures of facial affect. Palo Alto, CA: Consulting Psychological Press; 1976. Tabrizi SJ, et al. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. Lancet Neurology 2009; 8: 791-801 Tabrizi SJ, et al. Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: the 12-month longitudinal analysis. Lancet Neurology 2011; 10: 31-42. Stout JC et al. Evaluation of longitudinal 12- and 24-month cognitive outcomes in premanifest and early Huntington's disease. In submission 2011. Labuschagne I, et al. Emotional face recognition deficits in pre-manifest through stage-II Huntington's disease. In submission 2011.

Description of Hopkins Verbal Learning Test for HD Common Data Elements

Instrument Name:	Hopkins Verbal Learning Test – Revised (HVLT-R)
Classification:	Supplemental
Short Description of Instrument:	Summary/ Overview of Instrument: Assesses verbal short-term learning and memory performance. The test includes three learning trials, a delayed recall (25 minute delay), and a yes/no recognition trial. However, note that the three learning trials only are recommended as HD CDEs. Six distinct forms of the HVLT-R are available, minimizing practice effects on repeated administrations. Construct measured: Memory Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): Assessment of cognitive function in HD cross-sectional studies Means of administration (paper and pencil, computerized): Verbal with paper and pencil response form. Location of administration (clinic, home, telephone, etc): Clinical Setting Intended respondent (patient, caregiver, etc.): Patients # of items: N/A # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): Raw scores are derived for Total Recall (# items correctly recalled on each of 3 trials, maximum 36), Delayed Recall (# items correctly recalled following delay, maximum 12), Retention (% of items initially recalled that are also recalled after delay), and a Recognition Discrimination Index (# hits minus # false positives in 12 target + 12 foil recognition test, Solomon et al, 2007). Standardization of scores to a reference population (z scores, T scores, etc): Published manual norms are available. T scores can be are derived from raw scores for different age ranges. If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc). General population.

Description of Hopkins Verbal Learning Test for HD Common Data Elements

Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous. If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups. HVLT floor and ceiling effects do not appear to be a concern for controls, early HD or premanifest HD populations (see Stout et al. 2011).
Psychometric Properties	Reliability: Test-retest or intra-interview (within rater) reliability (as applicable): Based on an interval of 6 weeks, reliability coefficients for the primary variables are: .74 for Total Recall, .66 for Delayed Recall, .39 for retention %, and .40 for Recognition discrimination index (Benedict et al., 1998). Practice effects were minimal. Inter-interview (between-rater) reliability (as applicable): Internal consistency: Statistical methods used to assess reliability: Pearson correlations, paired t-tests, and Wilcoxon test. Reliability data from the CAB study will be available for the immediate recall measure of this task by end of 2012 for 100 control, 100 premanifest, and 50 early HD subjects. Validity: Content validity: HVLT correlates well with the California Verbal Learning Test (Lacritz, 2001). Construct validity: Good construct validity (Woods et al., 2005). Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): In published cross-sectional (Stout et al., 2011) and internal analyses (PREDICT-HD), raw scores for total learning and delayed recall reveal differences between premanifest HD and controls, especially in individuals who are closer to an expected diagnosis. Unpublished internal analyses of 7-year longitudinal data (PREDICT) do not show significant changes in longitudinal rate of change in premanifest HD. Known Relationships to Other Variables (e.g. gender, education, age, etc): Aging has a large effect on performance. Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups): N/A
Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)	Strengths: Total learning and delayed recall are sensitive to subtle changes in verbal learning and memory in premanifest HD. The task is easy to administer and score and is well-tolerated even by significantly impaired individuals. Task has been tested at sites in the United States, Canada, United Kingdom, Australia, Germany, and Spain. Multiple forms are available. Weaknesses: There are no significant differences in longitudinal rate of change for individual with premanifest HD vs. controls. The task is for use in subjects 16 years and older but the CDE recommendations do not inform juvenile HD (as

	indicated in the summary document). Availability (copyright): Available for purchase through PAR: http://www4.parinc.com/Products/Product.aspx?ProductID=HVLT-R Special Requirements for administration: None Administration Time: Approximately 30 minutes Translations available (e.g. Spanish, French, Other languages): German, French, Dutch, Italian, Norwegian, Spanish.
References:	Key Reference: Brandt J. The Hopkins Verbal Learning Test: development of a new verbal memory test with six equivalent forms. *The Clinical Neuropsychologist*. 1991; 5: 125-142. Brandt, J. & Benedict, R. (2001). Hopkins Verbal Learning Test-Revised. Lutz, FL: Psychological Assessment Resources. Benedict, R.H.B., & Zgalijardic, D.J. (1998). Practice effects during repeated administration of memory tests with and without alternate forms. Journal of Clinical and Experimental Neuropsychology, 20, 339-352. Lacritz, L. H., Cullum, C. M., Weiner, M. F., & Rosenberg, R. N. (2001). Comparison of the hopkins verbal learning test-revised to the California verbal learning test in Alzheimer's disease. *Appl.Neuropsychol., 8,* 180-184. Stout, J. C., Paulsen, J. S., Queller, S., Solomon, A. C., Whitlock, K. B., Campbell, J. C. et al. (2011). Neurocognitive signs in prodromal Huntington disease. *Neuropsychology., 25,* 1-14. Woods, S.P., Cobb Scott, J., Dawson, M.S., Morgan, E.E., Carey, C.L., Heaton, R.K., Grant, I., HNRC Group (2005). Construct validity of Hopkins Verbal Learning Test-Revised component process measures in an HIV-1 sample. Archives of Clinical Neuropsychology, 20, 1061-1071 Strauss E, Sherman EMS, Spreen O. *A compendium of neuropsychological tests: administration, norms, and commentary, 3rd ed*. New York: Oxford University Press; 2006, p. 760-769.

Description of Map Search Task for HD Common Data Elements

Instrument Name:	Map Search Task
Classification:	Exploratory
Short Description of Instrument:	Summary/ Overview of Instrument: The Map Search task is a subtest of the Test of Everyday Attention (TEA). It requires patients to identify target symbols among distractor symbols on a visually cluttered map within a specific time interval. The evidence in HD is primarily for a 60 second interval (TRACK-HD). Construct measured: A task of sustained visual attention Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): Assessment of cognitive function in HD cross-sectional and longitudinal studies Means of administration (paper and pencil, computerized): Paper and pencil Location of administration (clinic, home, telephone, etc): Clinic Intended respondent (patient, caregiver, etc.): Patient # of items: N/A # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): Scoring is based on the number of correctly identified symbols within a specific time interval. TRACK-HD used a 60 second time interval. Standardization of scores to a reference population (z scores, T scores, etc): N/A If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc). N/A
Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous. If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups. No ceiling or floor effects have been detected in controls, premanifest and early (stage 1) HD, but task approached floor effects in stage 2 HD.
Psychometric Properties	Reliability: Test-retest or intra-interview (within rater) reliability (as applicable): The test-retest coefficients for this task are .80 to .89 in controls (Strauss et al., 2006) and 0.84 to 0.85 in Stroke patients (Robertson et al., 1994)

	Reliability data from the CAB study will be available by end of 2012 for 100 control, 100 premanifest, and 50 early HD subjects. Validity: Because the Map Search subtest uses everyday items, it provides a feeling of high face validity; however for those with questionable visual acuity, its validity may be undermined if steps are not taken to ensure that extraneous sensory difficulties are not ruled out as a confound factor prior to assessment. Content validity: N/A Construct validity: N/A Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): In TRACK-HD, cross-sectional differences from controls were detected in premanifest HD, even in those estimated to be more than 10 years from onset, and were also detected in early HD (O'Regan et al., 2011). Analysis of TRACK-HD longitudinal data is imminent. Known Relationships to Other Variables (e.g. gender, education, age, etc): Manual indicates that gender is not related to Map Search performance. Age and education relations to Map Search performances are unknown (Strauss et al., 2006). Strauss et al. (2006) indicates performance of Map Search is not related to gender. In TRACK_HD, performance was related to age and education but not gender; change in performance was not related to age, gender or education. Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups): N/A
Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)	Strengths: Advantages are that the test is fast and engaging for the participant. Weaknesses: Although scoring is a simple count of number of target items correctly circled, scorers make a high number of scoring errors. TRACK-HD reduced errors by dividing the scoring template into 6 equal sections and subtotaling counts for each section separately. Availability (copyright): Available for purchase as part of the Test of Everyday Attention (TEA) from PAR: http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8054-458 Modifying the scoring template requires permission from Pearson, see weaknesses section above. Special Requirements for administration: As sold by Pearson, scoring is done by using an erasable plastic sleeve that allows comparison of subject's responses to a scoring template. In standard administration, responses for each subject are erased and the plastic sleeve is reused for the next subject. To retain source documentation, TRACK-HD used a permanent marker and a separate plastic sleeve for each administration of the task and those plastic sleeves then served as source documentation.

	A stopwatch is required for this task The scoring template can be modified to improve scoring accuracy (see Availability). Administration Time: Approximately 3 minutes Translations available (e.g. Spanish, French, Other languages): English, French, Dutch
References:	Key Reference: Standardization of scores to a reference population: normative data for the two-minute administration are available in the published test manual (Robertson et al., 1994) Robertson IH, Ward T, Ridgeway V & Nimmo-Smith I. Test of Everyday Attention, The (TEA). Thames Valley Test Company. Suffolk, England, 1994. Other References: Strauss E, Sherman MS, & Spreen O. *A Compendium of Neuropsychological tests: administration, norms and commentary. 3rd Edition.* 2006. Oxford University Press. O'Regan AM et al. Visuospatial deficits in Huntington's disease: an investigation using two tasks. Clinical Genetics, 80, S1: 51.

Description of Mental Rotation for HD Common Data Elements

Instrument Name:	Mental Rotation
Classification:	Exploratory
Short Description of Instrument:	Summary/ Overview of Instrument: The mental rotation task is a measure of one's ability to mentally rotate visual stimuli. Participants make judgments about pictures of 3-D objects made of cubes. Two objects are presented at a time. For each pair, the objects are either identical or mirror image reversals of each other. The participant's job is to determine as rapidly as possible whether the objects are the same (i.e., a copy that differs only in rotation angle) or different (i.e., mirror image objects). Whether mirrored or same, the two objects are displaced at various angles ranging from 5^{o} to 305^{o} in 60^{o} steps. Some pairs will be the same image rotated, and others will be mirrored. Construct measured: Visuospatial ability and mental manipulation Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): Assessment of cognitive function in HD cross-sectional and longitudinal studies Means of administration (paper and pencil, computerized): Computer (based on a paper/pencil version of the task) Location of administration (clinic, home, telephone, etc): Clinic Intended respondent (patient, caregiver, etc.): Patient # of items: 48 trials of varying rotational displacement # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): Scores are the number of trials responded to correctly. Standardization of scores to a reference population (z scores, T scores, etc): N/A If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc). N/A
Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous *If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups.* Slight evidence of ceiling in controls but not in premanifest HD nor in early HD

Description of Mental Rotation for HD Common Data Elements

Psychometric Properties	Reliability: Test-retest or intra-interview (within rater) reliability (as applicable): N/A Inter-interview (between-rater) reliability (as applicable): N/A Internal consistency: N/A Statistical methods used to assess reliability: N/A Validity: Content validity: N/A Construct validity: N/A Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): In TRACK-HD, cross-sectional differences from controls were detected in accuracy for late premanifest HD and early HD. Longitudinal data is not currently available. Known Relationships to Other Variables (e.g. gender, education, age, etc): In TRACK-HD, performance was related to age and gender but not education; change in performance relationships to age, gender or education are unknown. Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups): N/A
Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)	Strengths: N/A Weaknesses: Subjects tend to dislike this task because it seems difficult. Availability (copyright): Public domain. This task was adapted for computerized presentation by Julie Stout's lab as part of the TRACK-HD study and based on Shepard and Metzler (1971). http://hdresearch.ucl.ac.uk/completed-studies/track-hd/ Special Requirements for administration: A tablet computer; computer mouse; a holder for the mouse that makes the mouse easier to handle while using the left or right thumbs to respond with the respective mouse buttons; and labels of "Same" and "Different" affixed to the mouse holder that indicate which button corresponds to each response. Administration Time: 8 minutes Translations available (e.g. Spanish, French, Other languages): Dutch, French, and English
References:	Key Reference: Shepard, R.N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701-703. Metzler, J. & Shepard, R.N. (1974). Transformational studies of the internal representation of three-dimensional objects. In R L. Solso (Ed.), *Theories of cognitive psychology: The Loyola symposium* (pp.147-202). Potomac, MD: Lawrence Erlbaum.

Description of Montreal Cognitive Assessment for HD Common Data Elements

Instrument Name:	Montreal Cognitive Assessment (MoCA)
Classification:	Exploratory
Short Description of Instrument:	Summary/Overview of Instrument: Originally developed to screen for mild cognitive impairment (MCI). Covers seven cognitive domains with 30 total possible points. Construct Measured: Visuospatial/ executive functioning, naming, memory, attention, language, abstraction and orientation. Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): Originally developed to screen for MCI. Means of administration: Paper/pencil. Administered by trained examiner. Location of administration: Clinic. home, community Intended respondent (e.g. patient, caregiver, etc): Patient # of items: 30 # of subscales and names of sub-scales: Seven subscales assessing various cognitive domains, including: visuospatial/executive (5 points); naming (3 points); memory (5 points for delayed recall); attention (6 points); language (3 points); abstraction (2 points); and orientation (6 points). # of items per sub-scale: Variable
Scoring	Scoring: Higher scores are better. One point is added if education <=12 years. Standardization of scores to a reference population: The only normative adjustment for the MoCA adds one point to total if education is less than or equal to 12 years. No other norms have been developed, but large population-based study suggests norms are needed (Rossetti et al., 2011). Background: Developed to screen for MCI in the general population. Differentiates general population from MCI and dementia in Alzheimer disease (AD) and in Parkinson's disease (PD).
Measurement	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous If ordinal or continuous, explain if ceiling or floor effect are to be expected if the measure is used in specific HD subgroups: Ceiling effects are likely in cognitively intact individuals. Floor effects not likely except in most severely affected Huntington's disease (HD) patients.
Psychometric Properties:	Feasibility: Time of administration is 10 minutes. Reliability: Good test-retest reliability (r = 0.92 in validation sample). Good internal consistency (Cronbach's alpha = 0.83)

	Reliability data from the CAB study will be available by end of 2012 for 100 control, 100 premanifest, and 50 early HD subjects. Validity: Good discriminant validity (diagnostic groups) for AD, MCI and normal controls (NC). Cross-sectional studies demonstrate sensitivity to disease severity, but there are no longitudinal studies demonstrating sensitivity to change in cognitive function. Two published studies (Mickes et al., 2010; Vidswnovic et al., 2010) of MoCA performance in manifest HD suggest increased sensitivity to cognitive impairment compared to MMSE scores. This increased sensitivity and specificity held for individual items measuring visuospatial, language, memory and orientation as well as total score (Mickes et al., 2011). Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): there are no studies to determine the sensitivity to change of the MoCA in HD. Known relationship to other variables: Original validation study demonstrated relationship to education. This led to education adjustment (add one point if years of education less than or equal to 12). Rossetti and colleagues (2011) demonstrate that normative corrections need to be made for both age and education.
Rationale/ Justification:	Strengths: Widely available, easy to administer, broad range of cognitive domains, capturing domains germane to HD (i.e. executive function); more sensitive as a screening instrument for mild cognitive impairment or dementia; available in over 30 languages; free for non-profit use. Weaknesses: Although the scale assesses cued recall, this item is not considered in the total score; size of stimuli may need adaptation; normative data lacking; ceiling effect may limit utility to assess change in individuals with higher cognitive function. Availability (copyright): Test is available at http://www.mocatest.org. The test is available for no charge for clinical and educational use. Not-for-profit research is at no charge with prior permission. For-profit use requires approval and licensing agreement. Special Requirements for administration: N/A Administration Time: Approximately 10 minutes Translations available (e.g. Spanish, French, Other languages): Translated into over 30 languages and alternate forms available in English and French.
References:	Key Reference: Nasreddine et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005;53:695-699.

	Other References: Gill et al. Montreal Cognitive Assessment as a Screening tool for cognitive impairment in Parkinson's disease. Mov Disord 2008;23:1043-1046 Hoops S, Nazem S, Siderowf AD, Duda JE, Xie SX, Stern MB, Weintraub D. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. Neurology. 2009;73:1738-45. Rossetti HC, Lacritz LH, Cullum M, Weiner MF. Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. Neurology 2011;77:1272-1275. Mickes L, Jacobson M, Peavy G, Eixted JT, Lessig S, Goldstein JL, Corey-Bloom J. A comparison of two brief screening measures of cognitive impairment in Huntington's disease. Movement Disorders 2010;25:2229-22233. Videnovic A, Bernard B, Fan W, Jaglin J, Leurgans S, Shannon KM. The Montreal Cognitive Assessment as a screening tool for cognitive dysfunction in Huntington's disease. Movement Disorders 2010;25:401-404.

Description of Phonemic Verbal Fluency for HD Common Data Elements

Instrument Name:	Phonemic Verbal Fluency (PVF)
Classification:	Supplemental
Short Description of Instrument:	Summary/ Overview of Instrument Verbal fluency (VF) test production of words beginning with a specific letter. There are multiple versions of this test that vary the amount of time given to generate item names, the letters, the retrieval modality, restriction type, and the demand of executive control. One of the most-used versions is known as the Controlled Oral Word Association Test – COWAT. Patients are given three letters, one at a time, and asked to generate a list of words that begin with the specified letter (FAS, CFL, or PRW) within 1 minute. The majority of the evidence in HD uses the COWAT and on that basis alone the COWAT is recommended. Construct measured: Word knowledge, working memory, and executive functioning Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): Assessment of cognitive function in HD cross-sectional and longitudinal studies Means of administration (paper and pencil, computerized): Verbal Location of administration (clinic, home, telephone, etc): Clinic Intended respondent (patient, caregiver, etc.): Patient # of items: Tests three letters # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): Total number of correct words. Occasionally number of perseverations or intrusions are analyzed. Standardization of scores to a reference population (z scores, T scores, etc): Depending on norms used, standard scores may be calculated based on age, gender, and years of education/reading level. If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc). General population.

Description of Phonemic Verbal Fluency for HD Common Data Elements

Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups. No ceiling effects. Performance may be close to floor level in individuals with advanced stage HD.
Psychometric Properties	Reliability: Test-retest or intra-interview (within rater) reliability (as applicable): High test-retest reliability 'generally over 0.70' (Strauss et al., 2006) in the short term and in the long term. Inter-interview (between-rater) reliability (as applicable): The COWAT has high interrater reliability (.99) (Ross, 2003). High consistency between letters (F v A v S r=.83, Tombaugh et al., 1999; C v F v L r=.83, Ruff et al., 1996) Statistical methods used to assess reliability: Pearson correlations and intraclass correlation coefficients (inter-rater reliability). Validity: Content validity: Differences between different versions (e.g., FAS and BHR) appear negligible (Delis et al., 2001).Construct validity: Correlates well with tests of verbal IQ (r=.44 to .87) (Henry and Crawford, 2004). Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): In published cross-sectional (Stout et al., 2011) and internal analyses (PREDICT-HD), the test reveals performance differences between premanifest HD and healthy controls, especially in individuals who are closer to an expected diagnosis. Unpublished internal analyses of 7-year longitudinal data (PREDICT) show differences in longitudinal rate of change in premanifest HD as compared to controls. Known Relationships to Other Variables (e.g. gender, education, age, etc): Performance improves with years of education and declines slightly with aging (Crossley et al., 1997). Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups): N/A
Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)	Strengths: Task has been tested at sites in the United States, Canada, United Kingdom, Australia, Germany, and Spain. Task is easy to administer. Weaknesses: The choice of letters is dependent on word frequency within a language. Selection of equivalent word-frequency letters for cross-linguistic comparison poses a challenge. Education accounts for a significant amount of variance in scores. Availability (copyright): Public domain, though published forms are also available.

	Special Requirements for administration: Stop watch. Administration Time: 5-7 minutes Translations available (e.g. Spanish, French, Other languages): Versions of the task have been used in most common languages. However, see weaknesses above.
References:	Key Reference: Strauss E, Sherman EMS, Spreen O. A compendium of neuropsychological tests: administration, norms, and commentary, 3rd ed. New York: Oxford University Press, 2006, p. 502. Benton, A., & Hamsher, K (1989). Multilingual Aphasia Examination. Iowa City: AJA Associates. Micelli, G., Caltagirone, C., Gainotti, G., et al (1981). Neuropsycholgical correlates of localized cerebral lesions in nonaphasic brain-damaged patients. J of Clin Neuropsychology, 3, 53-63. Ross TP. The reliability of cluster and switch scores for the Controlled Oral Word Association Test. Arch Clin Neuropsychol 2003; 18: 153-64. Troyer, AK, Moscovitch, M, Winocur, G et al (1998). Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease. J of the Intl Neuropsychological Soc, 4, 137-143. Crossley, M., D'Arcy, C., & Rawson, N. S. (1997). Letter and category fluency in community-dwelling Canadian seniors: a comparison of normal participants to those with dementia of the Alzheimer or vascular type. *J.Clin.Exp.Neuropsychol., 19,* 52-62. Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). Delis-Kaplan Executive Function System. San Antonio, TX: The Psychological Corporation. Henry, J. D. & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology., 18,* 284-295. Ruff, R. M., Light, R. H., Parker, S. B., & Levin, H. S. (1996). Benton Controlled Oral Word Association Test: reliability and updated norms. *Arch.Clin.Neuropsychol., 11,* 329-338.

	Tombaugh, T.N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. Archives of Clinical Neuropsychology, 14, 167-177. Stout, J. C., Paulsen, J. S., Queller, S., Solomon, A. C., Whitlock, K. B., Campbell, J. C. et al. (2011). Neurocognitive signs in prodromal Huntington disease. *Neuropsychology., 25,* 1-14.

Description of Self-Paced Tapping for HD Common Data Elements

Instrument Name:	Self-Paced Tapping
Classification:	Core
Short Description of Instrument:	Summary/ Overview of Instrument: For the self-paced tapping task the participant uses either the dominant index finger or alternating thumbs to produce the target pace. An external input device (e.g., computer mouse) connected to a computer is used. Patients are asked to listen to a pacing tone, then begin tapping the specified buttons in time with the tone and to continue tapping at that same rate after the tone stops. The majority of the evidence for using this task in HD populations was collected with paces of 1.8 or 3.0 Hz. Construct measured: Cognitive and motor timing. Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): Assessment of cognitive function in HD cross-sectional and longitudinal studies Suitable for measuring cross-sectional and longitudinal changes in paced tapping timing abilities. Means of administration (paper and pencil, computerized): Computerized Location of administration (clinic, home, telephone, etc): Clinic Intended respondent (patient, caregiver, etc.): Patient # of items: N/A # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): The precision of all taps taken together is directly estimated. Timing precision is calculated as the reciprocal of the standard deviation (SD) of the intertap interval. The reciprocal of SD is preferred because it more closely satisfies statistical modeling assumptions of linear relationships to covariates of interest, approximate normality, and maintains constant variance of the differences between observed and predicted values. Standardization of scores to a reference population (z scores, T scores, etc): N/A If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc). N/A

Description of Self-Paced Tapping for HD Common Data Elements

Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups. 1/SD begins to show floor effects in Stage II HD, particularly in the 3.0 Hz condition
Psychometric Properties	Reliability: Test-retest or intra-interview (within rater) reliability (as applicable): N/A Inter-interview (between-rater) reliability (as applicable): N/A Internal consistency: N/A Statistical methods used to assess reliability: Test-retest correlations Reliability data from the CAB study will be available by end of 2012 for 100 control, 100 premanifest, and 50 early HD subjects. Validity: Content validity: N/A Construct validity: N/A Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): In TRACK-HD (unpublished), for both 1.8 and 3.0 Hz conditions, cross-sectional differences from controls were detected in premanifest HD and early HD. PREDICT-HD also found cross-sectional differences from controls in the 1.8 Hz condition in premanifest HD (Stout et al., 2011). In TRACK-HD (Stout et al., under review), early HD longitudinal annualized change over 24 months for the 3.0 Hz condition differed from change in controls. However, this was not the case for premanifest HD compared to controls. PREDICT-HD (Rowe et al., 2010) detected annualized longitudinal change over time (7 years) in the 1.8 Hz condition in premanifest HD. However, in TRACK-HD, the 1.8 Hz condition did not show different rates of change for either premanifest HD or early HD as compared to controls (Stout et al., in submission). Known Relationships to Other Variables (e.g. gender, education, age, etc): In PREDICT-HD, performance was related to education, age, and gender. In TRACK-HD, performance was related to education but not age or gender; change in performance was related to age but not to gender or education. Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups):
Rationale/ Justification (include any information on language and countries/	Strengths: The task is quick, easy to administer, and among the most sensitive cognitive tasks for use in premanifest and early HD Weaknesses: Aging may play a significant role in a person's performance of the self-paced timing task and must be matched across groups or considered in interpretation of findings.

cultures/ ethnic groups where tested)	Availability (copyright): This task was adapted for computerized presentation by Julie Stout's (julie.stout@monash.edu) lab as part of the TRACK-HD study. Special Requirements for administration: Administration Time: One condition (e.g., alternate thumbs at 1.8 Hz) takes about 3 minutes. Translations available (e.g. Spanish, French, Other languages): Dutch, French and English
References:	Key Reference: Rowe KC, et al. Self-paced timing detects and tracks change in prodromal Huntington disease. Neuropsychology 2010;24(4):435-442. Other References: Stout, J. C., Paulsen, J. S., Queller, S., Solomon, A. C., Whitlock, K. B., Campbell, J. C. et al. (2011). Neurocognitive signs in prodromal Huntington disease. Neuropsychology., 25, 1-14. Tabrizi SJ, et al. Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: the 12-month longitudinal analysis. Lancet Neurology 2011; 10: 31-42. Stout JC et al. Evaluation of longitudinal 12- and 24-month cognitive outcomes in premanifest and early Huntington's disease. In submission 2011.

Description of Simple and Two-Choice Reaction Time for HD Common Data Elements

Instrument Name:	Simple and Two-Choice Reaction Time
Classification:	Supplemental
Short Description of Instrument:	Summary/ Overview of Instrument: The task can be administered in several different ways. In the simple reaction time (RT) condition, the participant responds to the same stimulus on each trial (i.e., measures simple psychomotor speed). In the 2-choice RT condition, the participant makes one of two responses to one of two stimuli. Longer RTs are observed for choice compared with simple RT. Construct measured: Response selection and psychomotor speed Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): Means of administration (paper and pencil, computerized): Computerized Location of administration (clinic, home, telephone, etc): Clinic Intended respondent (patient, caregiver, etc.): Patient # of items: N/A # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): The main measure is reaction time, which is defined as the amount of time elapsed between the presentation of a stimulus and the subsequent response Standardization of scores to a reference population (z scores, T scores, etc): The task has not been standardized If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc).
Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups. No ceiling or floor effects
Psychometric Properties	Reliability: N/A Validity: Construct validity: Longer RT for the 2-choice than the simple RT condition

	suggest that the experimental manipulation influences response selection. Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): In published cross-sectional (Stout et al., 2011) and internal analyses (PREDICT-HD), simple and choice RT are sensitive to changes in prodromal HD, especially in individuals who are closer to an expected diagnosis. Unpublished internal analyses of 7-year longitudinal data (PREDICT) show changes in prodromal HD for simple and choice RT. Choice RT is slightly more sensitive in cross-sectional and longitudinal analyses. Known Relationships to Other Variables (e.g. gender, education, age, etc): N/A Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups): N/A
Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)	Strengths: Tasks is highly sensitive to changes in prodromal HD, both cross-sectionally and longitudinally. Task has been tested at sites in the United States, Canada, United Kingdom, Australia, Germany, and Spain. Task is easy to administer. Weaknesses: Touch screen response devices are not as reliable as button press devices. Availability (copyright): This task was adapted for computerized presentation by Julie Stout's lab (julie.stout@monash.edu) as part of the PREDICT-HD study. http://hdresearch.ucl.ac.uk/completed-studies/track-hd/ Special Requirements for administration: Computer and a touch screen or an external button-press device. Administration Time: 10 minutes Translations available (e.g. Spanish, French, Other languages): The task can be administered in any language
References:	Key Reference: Stout JC, Paulsen JS, Queller S, Solomon AC, Whitlock KB, Campbell JC, Carlozzi N, Duff K, Beglinger LJ, Langbehn DR, Johnson SA, Biglan KM, Aylward EH. Neurocognitive signs in prodromal Huntington disease. *Neuropsychology* 2011;25:1-14.

Description of Speeded Tapping Test for HD Common Data Elements

Instrument Name:	Speeded Tapping Test
Classification:	Core
Short Description of Instrument:	Summary/ Overview of Instrument: The participant is required to tap a key with their index finger as quickly as possible for a 10-second period. The task is repeated five times for each hand, with a brief rest period between trials. Several commercial versions of the task are available, using specialized equipment. In addition, numerous bespoke computerized versions of the task have been developed, using a computer keyboard or external hardware button. Construct measured: Psychomotor speed. Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool: The speeded tapping test may be used as a longitudinal marker of disease severity in manifest/premanifest HD or as a cross-sectional measure of impairment across disease stages or between manifest/premanifest HD and healty controls. Means of administration: Mechanical or computerized Location of administration: Clinic Intended respondent : Patient # of items: 5 administrations of 10 second trials # of subscales and names of sub-scales: N/A
Scoring	Scoring: The score is the mean number of taps produced across 5 trials for the dominant and non-dominant hands. Standardization of scores to a reference population (z scores, T scores, etc): If the task is conducted using procedures from the Halstead-Reitan Battery (Reitan & Wolfson, 1985), raw score can be converted to t-scores. If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc). General population
Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups. No ceiling or floor effects. However, severely advanced HD patients may not have sufficient manual dexterity to perform the task.

Description of Speeded Tapping Test for HD Common Data Elements

Psychometric Properties	Reliability: Test re-rest reliability: Test re-test reliability (assessed by correlational analysis and regression coefficients) varies across studies but is consistently in the high range (Lezak, Hoeison & Loring, 2004). Inter-interview (between-rater) reliability (as applicable): N/A Internal consistency: N/A Statistical methods used to assess reliability: Test-retest correlations Reliability data from the CAB study will be available by end of 2012 for 100 control, 100 premanifest, and 50 early HD subjects. Validity: Content validity: N/A Construct validity: N/A Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): Cross-sectional differences in tapping speed and the within-subject standard deviation of tapping speed were detected between controls and premanifest HD (PREDICT-HD and TRACK-HD) and early HD (TRACK-HD). In PREDICT-HD, but not TRACK-HD, these measures were sensitive to longitudinal change in premanifest HD, especially in individuals closer to an expected diagnosis. Known Relationships to Other Variables: Tapping speed declines with age, particularly from the fifth decade of life. In addition, there is an effect of gender, with males consistently tapping faster than females (Mitrushina et al., 2005). Diagnostic Sensitivity and Specificity, if applicable: Reduced tapping speed is seen in a variety of conditions, some not involving brain dysfunction, and ultimate interpretation has to be based upon the context provided by other tests.
Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)	Strengths: The task is highly sensitive to cross-sectional and longitudinal change in premanifest HD. The speeded finger tapping test is quick, easy to administer and well tolerated amongst patient groups. Education effects are small (Lezak et al., 2004). Weaknesses: The task may not be suitable for use in patients with severe motor impairment. Availability (copyright): Stand-alone tapping devices are available from Reitan Labs and Western Psychological Services. A computerized version of the task is also commercially available. However, the use of bespoke software and equipment is also common. http://www.reitanlabs.com/catalog/default.php?osCsid=42c018ae57cfd30f738942c41559550b&cPath=48

Description of Speeded Tapping Test for HD Common Data Elements

	http://portal.wpspublish.com/portal/page?_pageid=53,70151&_dad=portal&_schema=PORTAL http://www.cogtest.com/tests/cognitive_int/ts.html This task was adapted for computerized presentation by Julie Stout's lab (julie.stout@monash.edu) as part of the PREDICT-HD study. Special Requirements for administration: Computer with appropriate software / equipment or a stand-alone mechanical or electronic tapper required Administration Time: Less than 5 minutes. Translations available: N/A
References:	Key Reference: Reitan, R.M. (1979) Manual for administration of neuropsychological test batteries for adults and children. Tucson, AZ: Reitan Neuropsychology Laboratories, Inc. Other References: Reitan, R.M., & Wolfson, D. (1985) The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation. Tucson: Neuropsychology Mitrushina, M.M., Boone, K.B., Razani, J., & D'Elia, L.F., (2005). Handbook of normative data for neuropsychological assessment (2nd ed.). New York: Oxford University Press. Lezak, MD, Howieson, D.B., & Loring, D.W. (2004). Neuropsychological Assessment (4th ed.). New York: Oxford University Press.

Description of Spot the Change for HD Common Data Elements

Instrument Name:	Spot the Change
Classification:	Supplemental
Short Description of Instrument:	Summary/ Overview of Instrument: Spot the Change is a visual working memory task in which subjects are presented briefly with a computer screen showing 5 colored squares. The screen then displays the same squares in the same position, but with 1 square circled. Subjects were asked to indicate, within a specified time, whether the color of the circled square had changed. Construct measured: Visual working memory Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): Assessment of cognitive function in HD cross-sectional and longitudinal studies Means of administration (paper and pencil, computerized): Computerized Location of administration (clinic, home, telephone, etc): Clinic Intended respondent (patient, caregiver, etc.): Patient # of items: N/A # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): The number correct was adjusted for guessing using k as computed per Cowan et al (2000): $k = (H + CR - 1)N$ where H = # hits, CR = # correct rejections, N = # items displayed = 5 Standardization of scores to a reference population (z scores, T scores, etc): N/A If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc). N/A

Description of Spot the Change for HD Common Data Elements

Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups. The set size 5 we recommend is relatively immune to floor and ceiling effects in healthy controls, premanifest HD and early HD. However, set size 3 results approached ceiling in controls and premanifest HD. Set size to 7 is not recommended for use in early HD because of task difficulty and frustration in this population.
Psychometric Properties	Reliability: Test-retest or intra-interview (within rater) reliability (as applicable): 24-month test-retest correlations ranged from .34 to .60 in controls, late premanifest HD, and early (stage 1) HD in the TRACK-HD study. Inter-interview (between-rater) reliability (as applicable): N/A Internal consistency: N/A Statistical methods used to assess reliability: test-retest correlation Reliability data from the CAB study will be available by end of 2012 for 100 control, 100 premanifest, and 50 early HD subjects. Validity: Content validity: N/A Construct validity: N/A Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): In TRACK-HD, cross-sectional differences from controls were detected in premanifest HD and early HD (Tabrizi et al, 2009); longitudinal rate of change (annualized over 24 months) in early HD but not premanifest HD differed from rate of change in controls (Tabrizi et al., 2011; Stout et al., in submission). Known Relationships to Other Variables (e.g. gender, education, age, etc): In TRACK-HD, performance was related to age and education but not gender; change in performance was not related to age, gender or education. Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups):
Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)	Strengths: This task is nonverbal and is feasible for use across languages. The response does not rely on motor speed or fine motor control as task performance requires only an untimed button press. Weaknesses: Ideally, multiple set sizes would be used but this would increase test time. Track-HD found that set size 5 was neither at ceiling nor at floor for controls, premanifest, and early HD. However, set size 3 approached ceiling in controls and premanifest subjects far from onset whereas set size 7 began to approach floor in Early HD subjects.

Description of Spot the Change for HD Common Data Elements

	Availability (copyright): Adapted by Julie Stout's lab as part of the TRACK-HD study based on Cowan et al. (2005). http://hdresearch.ucl.ac.uk/completed-studies/track-hd/ **Special Requirements for administration**: Computer and mouse **Administration Time**: 8 minutes **Translations available** (e.g. Spanish, French, Other languages): French, Dutch, and English
References:	Key Reference: Cowan, N., The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behavioral and Brain Research 2000; 24: 87-185 (see page 166). Cowan, N., Elliott, EM, Saults, JS, Morey, CC, Mattox, S, Hismjatullina, A, Conway, ARA. On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. Cognitive Psychology 2005; 51: 42-100. Other References: Tabrizi SJ, et al. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. Lancet Neurology 2009; 8: 791-801 Tabrizi SJ, et al. Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: the 12-month longitudinal analysis. Lancet Neurology 2011; 10: 31-42. Stout JC et al. Evaluation of longitudinal 12- and 24-month cognitive outcomes in premanifest and early Huntington's disease. In submission 2011.

Description of Stroop Tests for HD Common Data Elements

Instrument Name:	Stroop Tests: Naming, Interference, Reading
Classification:	Naming and Reading- Core (select 1); Interference- Supplemental
Short Description of Instrument:	Summary/ Overview of Instrument: The Stroop test involves three trials. In the WORD trial, the subject reads words of color names (e.g., red, blue) printed in black ink. In the COLOR trial, the subject identifies colors (e.g., rectangles printed in red or blue). Finally, in the COLOR-WORD response inhibition trial, the subject must name the color in which a word is presented, while ignoring the printed word. Thus, incongruence between the word's color and identity (e.g., the word "blue" presented in red) requires inhibition and response selection. Multiple versions of the Stroop test are available (e.g. Victoria, Golden, D-KEFS, and Trenerry versions). The UHDRS version of the Stroop task has been most commonly used in HD research. To date, no one version of the Stroop Tests has been shown to be clearly superior to others. Construct measured: Cognitive flexibility and processing speed Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): Assessment of cognitive function in HD cross-sectional and longitudinal studies. Means of administration (paper and pencil, computerized): Paper and Computerized Location of administration (clinic, home, telephone, etc): Clinical Settings Intended respondent (patient, caregiver, etc): Patient # of items: N/A # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): Scoring for each trial type is based on the number of correct responses in a fixed amount of time, typically within 45 seconds. Higher scores indicate better cognitive performance. Standardization of scores to a reference population (z scores, T scores, etc): Raw scores can be converted to t scores for different ranges of age and years of education, depending on norms used. Studies reporting raw scores should control for age and education. If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc). General population (5-90 years of age; education levels of 2 to 20 years).

Description of Stroop Tests for HD Common Data Elements

Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous. If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups. No floor effects. Ceiling effects can be avoided if any subjects who reach the end of the page before the allotted time has elapsed are redirected to the top row and continue working until the end of the alotted time period. Individuals with advanced disease may struggle with the interference trial.
Psychometric Properties	Reliability: High reliability across different versions. Test-retest or intra-interview (within rater) reliability (as applicable): Test-retest reliabilities covers periods of 1 minute to 10 days. Reliablities for Word, Color, and Color-Word are respectively .88, .79 and .71 (Jensen, 1965) and .89, .84., and .73 (Golden, 1975). Inter-interview (between-rater) reliability (as applicable): Internal consistency: Correlations among the subtests are moderate to high (.71 to .84) (Chafetz and Mathew, 2004). Statistical methods used to assess reliability: Intraclass correlations Reliability data from the CAB study will be available for the Stroop Word condition of this task by end of 2012 for 100 control, 100 premanifest, and 50 early HD subjects. Validity: Content validity: Construct validity: The interference score correlates well with measures of attention and prepotent response inhibition (May and Hasler, 1998) Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): In published cross-sectional (Stout et al., 2011) and internal analyses (PREDICT-HD), the test is sensitive to changes in premanifest HD, especially in individuals who are closer to an expected diagnosis. Unpublished internal analyses of 7-year longitudinal data (PREDICT) also shows changes in rates of change over time in premanifest HD on all subtests, especially color and word naming. Cross sectionally, in Stroop WORD, the TRACK-HD study found that healthy controls performed significantly better than both the early HD and the premanifest HD groups. Longitudinally, the TRACK-HD study found significant differences in rates of change for early HD compared to controls, but did not find significant differences in rates of change for premanifest HD compared to controls. In Stroop WORD, the TRACK-HD premanifest participants may be less likely to show cognitive effects than the PREDICT-HD Premanifest participants because 1) they are further from estimated onset based on CAG repeat length and age (Langbehn et al., 2004) and 2) they are potentially less progressed in actuality because the TRACK-HD study excluded premanifest subjects based on UHDRS motor scores >= 5. Generally speaking, cognitive tests will be more effective metrics in studies of premanifest HD when the focus is on subjects that are close to onset. Meta-analysis of HD observational studies published 1993-2007 reveals both cross sectional performance differences compared to healthy controls and longitudinal change within HD groups over time for Stroop Reading and Stroop Color that is evident in both premanifest

and Early HD. The Stroop Interference findings are less impressive, with smaller cross sectional effect sizes and nonsignificant longitudinal effects (see table below).

	Cross-Sectional sensitivity in PreHD (Group: EffectSize, Pvalue, # of studies/ total # of HD participants across studies)	Cross-Sectional sensitivity in HD (Group: EffectSize, Pvalue, # of studies/total # of HD participants across studies)	Longitudinal sensitivity within subjects (Group: EffectSize, Pvalue, # of studies/ total # of HD participant across studies)
Stroop Reading	All Pre: -0.44, 0.001, 13/242; Near Pre: -0.65, 0.001, 4/152	Early: -1.29, <0.001, 10/220	Dx: -0.65, 0.022, 4/115; Near Pre: -0.61, <0.001, 2/160; All Pre: -0.47, <.003, 4/180
Stroop Colour	All Pre: -0.44, 0.002, 14/260; Near Pre: -0.87, 0.001, 4/152	Early: -1.35, <0.001, 9/207	Dx: -0.79, 0.008, 3/102; Near Pre: -0.44, 0.001, 2/160; All Pre: -0.34, 0.001, 4/180
Stroop Interference	All Pre: -0.24, 0.065, 18/332; Near Pre: -0.64, 0.004, 5/158	Early: -1.09, <0.001, 10/184	Dx: -0.15, 0.108, 4/115; Near Pre: -0.3, 0.215, 2/159; All Pre: 0, .999, 5/212

Known Relationships to Other Variables (e.g. gender, education, age, etc): May not be valid in color-blind individuals. The color-word interference score is vulnerable to aging (Mitrushina et al., 2005). Age and education should be controlled if reporting raw scores.

Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups):

Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)

Strengths: The color and word subtest are particularly sensitive in cross-sectional and longitudinal studies of premanifest and early manifest HD. Task has been tested at sites in the United States, Canada, United Kingdom, Australia, Germany, and Spain. Task is easy to administer.

Weaknesses: N/A

Availability (copyright): The majority of the evidence supporting use of the Stroop test in HD relies upon the version of the test that is administered along with the UHDRS http://www.huntington-study-group.org/Resources/UHDRS/tabid/67/Default.aspx.This version of the Stroop is believed to be in the public domain. Other versions of the Stroop are expected to show similar effects and various versions of the test are available commercially.

Special Requirements for administration: A stopwatch is required.

Administration Time: Assessment takes approximately 2 minutes for each of the three trial types.

	Translations available (e.g. Spanish, French, Other languages): Spanish (Golden Version), Cantonese (Victoria Version). The UHDRS version is available in a large number of European languages including Czech, Danish, Dutch, Finnish, French, German, Italian, Norwegian, Polish, Portuguese, Spanish and Swedish.
References:	Key Reference: Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology: General, 18, 643-662. Golden & Greshwater (2002). The Stroop Color and Word Test: A Manual for Clinical and Experimental Uses. Wood Dale, IL: Stoelting Co. Other References: Chafetz, M. D. & Matthews, L. H. (2004). A new interference score for the Stroop test. *Arch.Clin.Neuropsychol., 19,* 555-567. Jensen, A.R. (1965). Scoring the Stroop test. Acta Psychologia, 24, 298-408. Koga H, Takashima Y, Murakawa R, Uchino A, Yuzuriha T, Yao H. (2009). Cognitive consequences of multiple lacunes and leukoaraiosis as vascular cognitive impairment in community-dwelling elderly individuals. J Stroke Cerebrovasc Dis. 2009 Jan;18(1):32-7. Golden, C.J. (1975). The measurement of creativity by the Stroop color and word test. Journal of Personality Assessment, 39, 502-506. May, C. P. & Hasher, L. (1998). Synchrony effects in inhibitory control over thought and action. *J.Exp.Psychol.Hum.Percept.Perform., 24,* 363-379. Mitrushina, M.M., Boone, K.B., Razani, J., & D'Elia, L.F., (2005). Handbook of normative data for neuropsychological assessment (2nd ed.). New York: Oxford University Press. Murphy CF, Gunning-Dixon FM, Hoptman MJ, Lim KO, Ardekani B, Shields JK, Hrabe J, Kanellopoulos D, Shanmugham BR, Alexopoulos GS (2007). Biological Psychiatry, 61, 1007-10 Strauss E, Sherman EMS, Spreen O. A compendium of neuropsychological tests: administration, norms, and commentary, 3rd ed. New York: Oxford University Press; 2006. Stout, J. C., Paulsen, J. S., Queller, S., Solomon, A. C., Whitlock, K. B., Campbell, J. C. et al. (2011). Neurocognitive signs in premanifest Huntington disease. *Neuropsychology., 25,* 1-14.

Description of Symbol Digit Modality Test for HD Common Data Elements

Instrument Name:	Symbol Digit Modality Test
Classification:	Core
Short Description of Instrument:	Summary/ Overview of Instrument: The SDMT measures the time to pair abstract symbols with specific numbers. The test requires elements of attention, visuoperceptual processing, working memory, and psychomotor speed. Construct measured: Psychomotor speed, attention/integration Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): This test has been shown to predict group membership defined by processing speed deficits, such as brain-injured versus control samples and has been used as a sensitive outcome in studies identifying predictors of longitudinal decline in elders. It is currently a part of the UHDRS. Means of administration (paper and pencil, computerized): Written (oral form also available) Location of administration (clinic, home, telephone, etc): Clinical Setting Intended respondent (patient, caregiver, etc): Patient # of items: N/A # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): The score is the number of correctly coded items from 0-110 in 90 seconds Standardization of scores to a reference population (z scores, T scores, etc): Manual norms have been criticized because they are based on a sample of convenience and were collected in the 1970's. Other published norms are available based on age, education, and sex. If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc). General population
Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups. No ceiling or floor effects. Individuals with advanced HD may struggle to write legibly due to motor disability.

Psychometric Properties

Reliability:

Test-retest or intra-interview (within rater) reliability (as applicable): The test demonstrates strong reliability and validity coefficients (6-year interval) . Test-retest reliability in other studies ranges between 29 days to 2 years (r=.70 to .91) (e.g., Smith et al., 1991).

Inter-interview (between-rater) reliability (as applicable): N/A

Internal consistency: NA

Statistical methods used to assess reliability: Reliability coefficient.

Reliability data from the CAB study will be available by end of 2012 for 100 control, 100 premanifest, and 50 early HD subjects.

Validity:

Content validity: SDMT correlates with oral versions (r=.78) (Smith et al., 1991).

Construct validity: SDMT correlates well with the Wechsler Digit Symbol subtest (r=.62 to .91) (Hinton-Bayre et al., 1999).

Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): In published cross-sectional (Stout et al., 2011) and internal analyses (PREDICT-HD), the test is sensitive to impairments in premanifest HD, especially in individuals who are closer to expected diagnosis. Unpublished internal analyses of 7-year longitudinal data (PREDICT) show significant difference in rates of change over time in premanifest HD compared to gene negative controls.

Meta-analysis of HD observational studies published 1993-2007 reveals cross sectional performance differences between early HD and healthy controls and marginally significant cross sectional differences between premanifest HD and controls. The meta-analysis indicates that decline over time is not statistically significant in either premanifest HD or early HD. It is important to note, however, that the meta-analysis focused on decline over time within each HD subgroup. A comparison of longitudinal rate of change in HD v. rate of change in healthy controls will produce significant effects (as seen in TRACK-HD and PREDICT-HD) that are not revealed by the meta-analysis, due to practice effects in healthy controls that are less apparent in the HD subgroups.

	Cross-Sectional sensitivity in PreHD (Group: EffectSize, Pvalue, # of studies/ total # of HD participants across studies)	Cross-Sectional sensitivity in HD (Group: EffectSize, Pvalue, # of studies/total # of HD participants across studies)	Longitudinal sensitivity within subjects (Group: EffectSize, Pvalue, # of studies/ total # of HD participants across studies)
SDMT	All Pre: -0.33, 0.064, 13/268; Near Pre: -0.66, 0.18, 3/134	Early: -1.69, <0.001, 8/98	Dx: -0.22, 0.090, 3/93; Near Pre: -0.27, 0.006, 2/160; All Pre: -0.08, .630, 4/205

	In TRACK-HD, cross-sectional differences from controls were detected in premanifest HD and early HD (Tabrizi et al, 2009); longitudinal rate of change (annualized over 24 months) in early HD but not premanifest HD differed from rate of change in controls (Tabrizi et al., 2011; Stout et al., in submission). The test tracks the rate of progression in HD over a median period of 2.7 years (Mahant, 2003). The TRACK-HD premanifest participants may be less likely to show cognitive effects than the PREDICT-HD Premanifest participants because 1) they are further from estimated onset based on CAG repeat length and age (Langbehn et al., 2004) and 2) they are potentially less progressed in actuality because the TRACK-HD study excluded premanifest subjects based on UHDRS motor scores >= 5. Generally speaking, cognitive tests will be more effective metrics in studies of premanifest HD when the focus is on subjects that are close to onset. Known Relationships to Other Variables (e.g. gender, education, age, etc): Performance improves with IQ (Nielsen, 1989) and declines with age (Selnes, et al., 1991). Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups): N/A
Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)	Strengths: Sensitive to changes in premanifest HD in cross-sectional and longitudinal studies. Easy to administer and score. Multiple forms available. Weaknesses: More severe motor impairment may influence results. An oral form of the test is also available, although much less is known about this version in HD. Availability (copyright): Copyright belongs to Western Psychological Services http://portal.wpspublish.com/portal/page?_pageid=53,69289&_dad=portal&_schema=PORTAL Special Requirements for administration: Stopwatch/clock. Administration Time: less than 5 minutes Translations available (e.g. Spanish, French, Other languages): Involves only geometric figures and numbers, and therefore can be administered to people who do not speak English.
References:	Key Reference: Smith A. Symbol digit modalities test: Manual. Los Angeles: Western Psychological Services; 1982. Other References: DeMonte, VE, Geffen, GM, May, CR, & MacFarland, K. (2009). Improved sensitivity of the rapid screen of mild traumatic brain injury. J Clin Exp Neuropsychology, 6, 1-11.

Hinton-Bayre, A. D., Geffen, G. M., Geffen, L. B., McFarland, K. A., & Friis, P. (1999). Concussion in contact sports: reliable change indices of impairment and recovery. *J.Clin.Exp.Neuropsychol., 21,* 70-86.

Mahant, N., McCusker, E. A., Byth, K., & Graham, S. (2003). Huntington's disease: clinical correlates of disability and progression. *Neurology, 61,* 1085-1092.

Selnes, O. A., Jacobson, L., Machado, A. M., Becker, J. T., Wesch, J., Miller, E. N. et al. (1991). Normative data for a brief neuropsychological screening battery. Multicenter AIDS Cohort Study. *Percept.Mot.Skills, 73,* 539-550.

Stout, J. C., Paulsen, J. S., Queller, S., Solomon, A. C., Whitlock, K. B., Campbell, J. C. et al. (2011). Neurocognitive signs in prodromal Huntington disease. *Neuropsychology., 25,* 1-14.

Knopman, DS, Mosley, TH, Catellier, DJ, Coker, LH, Atherosclerosis risk in communities study brain MRI study (2009). Fourteen-year longitudinal study of vascular risk factors, APOE genotype, and cognition: the ARIC MRI study. Alzheimer's & Dementia: the Journal of the Alzheimer's Association, 5, 207-214.

Wechsler D. (1997). Wechsler adult intelligence scale-III. New York: Psychological Corporation.

Description of Trail Making Test for HD Common Data Elements

Instrument Name:	Trail Making Test (Parts A and B)
Classification:	Core
Short Description of Instrument:	Summary/ Overview of Instrument: The first part of this test, Trails A, requires the subject to rapidly sequence numbers from 1 through 25, with the score being the time to complete the task. The second part, Trails B, is a more difficult cognitive flexibility task requiring the subject to follow a sequential pattern while shifting cognitive sets, sequencing from 1 to 13 while switching between numbers and letters (i.e., 1-A-2-B, etc), with the score being the time to complete the task. The utility and psychometric properties of the Trails B are so well accepted that it is one of the few measures that it is used across neurologic and psychiatric clinical and research patient populations. Construct measured: Psychomotor speed and executive functions Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): The Trail Making Test is a measure of psychomotor speed, visual scanning, and executive ability. Means of administration (paper and pencil, computerized): Written Location of administration (clinic, home, telephone, etc): Clinical setting Intended respondent (patient, caregiver, etc.): Patient # of items: N/A # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): Scoring of A and B are reported as the number of seconds required to complete the task. Higher scores indicate greater impairment. Performance varies by age and education, and thus normative standards are used to classify patient performance. Errors affect the patient's score only in that the correction of errors is included in the completion time for the task. If a patient has not completed both parts after five minutes, it is unnecessary to continue the test. Parts A & B must be completed together and in the correct order for test administration to be valid. Standardization of scores to a reference population (z scores, T scores, etc): Raw scores (time to complete) are converted to scaled scores (0-19). Scale score is converted to t score by sex, education, age, and ethnicity. Norms are from the Halstead-Reitan Battery.

	If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc). General population
Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): Continuous. If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups. No ceiling effects. More cognitively impaired individuals may not be able to complete Trails B.
Psychometric Properties	Reliability: Test-retest or intra-interview (within rater) reliability (as applicable): For intervals of 3 weeks to 1 year, test-retest reliability is moderate to high for Part A (r=.36 to .79) and Part B (r=.44 to .89) (Bornstein et al., 1987, Matarazzon et al., 1974, Dikmen et al., 1999) Inter-interview (between-rater) reliability (as applicable): Interrater reliability has been found to be high for both Part A r=(.94) and Part B (r=.90). Internal consistency: Statistical methods used to assess reliability: Correlational analyses and reliability coefficients. Reliability data from the CAB study will be available by end of 2012 for 100 control, 100 premanifest, and 50 early HD subjects. Validity: Content validity: Part A and B correlate moderately (r=.31) (Heilbronner et al., 1991). Construct validity: Subtests correlate with visual search tasks (r-.37 to .93) (Ehrenstein et al., 1982). Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): In published cross-sectional (Stout et al., 2011) and internal analyses (PREDICT-HD), Parts A and B are sensitive to impairments in premanifest HD, especially Part B in individuals who are closer to an expected diagnosis. Unpublished internal analyses of 7-year longitudinal data (PREDICT) show differences in rates of longitudinal change in premanifest HD on both subtests, but especially Part A, compared to gene negatives. Cross sectionally, the TRACK-HD study found that healthy controls performed significantly better than both the early HD and the premanifest HD groups. Longitudinally, the TRACK-HD study found significant differences in rates of change for early HD compared to controls, but did not find significant differences in rates of change for premanifest HD compared to controls. The TRACK-HD premanifest participants may be less likely to show cognitive effects than the PREDICT-HD Premanifest participants because 1) they are further from estimated onset based on CAG repeat length and age (Langbehn et

	al., 2004) and 2) they are potentially less progressed in actuality because the TRACK-HD study excluded premanifest subjects based on UHDRS motor scores >= 5. Generally speaking, cognitive tests will be more effective metrics in studies of premanifest HD when the focus is on subjects that are close to onset. Known Relationships to Other Variables (e.g. gender, education, age, etc): Performance declines with IQ and educational level (Diaz-Asper et al., 2004; Clark et al, 2004; Hester et al., 2005). Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups): Well-established in multiple disease groups- see Strauss, Sherman, & Spreen (2006) or Mitrushina et al. (2005) for details. O'Rourke et al. (2011) showed sensitivity to detect differences between prodromal HD and gene negative controls.
Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)	Strengths: Parts A and B are sensitive in cross-sectional and longitudinal studies of prodromal HD. Task has been tested at sites in the United States, Canada, United Kingdom, Australia, Germany, and Spain. Weaknesses: The examiner must carefully monitor a participant's performance to accurately score errors. The reliability of test administration can vary by examiner's reaction time in noticing errors and pointing them out, which introduces imprecision. More severe motor impairment may influence results. Participants who are very cognitively impaired may not be able to complete the task, which must be dealt with statistically (e.g., set a maximum time for noncompleters). • Availability (copyright): Public domain and may be photocopied, though there are versions available for purchase. http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8091-108&Mode=summary. Special Requirements for administration: Stopwatch. Administration Time: 10 minutes. Translations available (e.g. Spanish, French, Other languages):
References:	Key Reference: Chen P, Ratcliff G, Belle S, al. e. Cognitive test that best discriminate between presymptomatic AD and those who remain nondemented. Neurology 2000; 55:1847-1853. Fals-Stewart W. An Interrater Reliability Study of the Trail Making Test (Parts A and B). Perceptual and Motor Skills, 1992, 74: 39-42. Heaton, R. K., Grant, I., & Matthews, C. G. (1991). Comprehensive norms for an

expanded Halstead–Reitan Battery: Demographic corrections, research findings, and clinical applications. Odessa, FL: Psychological Assessment Resources.

Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*, 271–276.

Reitan RM. (1992). Trail making test. Tucson, AZ: Reitan Neuropsychology Laboratory. Heilbronner, R.L., Kinsella, G.J., Ong, B., & McGregor, J., (1991). Lateralized brain damage and performance on Trail Making A and B, Digit Span Forward and Backward, and TPT memory and location. Archives of Clinical Neuropsychology, 6, 251-258.

Bornstein, R. A., Baker, G. B., & Douglass, A. B. (1987). Short-term retest reliability of the Halstead-Reitan Battery in a normal sample. *J.Nerv.Ment.Dis., 175,* 229-232.

Clark, M. S., Dennerstein, L., Elkadi, S., Guthrie, J. R., Bowden, S. C., & Henderson, V. W. (2004). Normative verbal and non-verbal memory test scores for Australian women aged 56-67. *Aust.N.Z.J.Psychiatry, 38,* 532-540.

Diaz-Asper, C. M., Schretlen, D. J., & Pearlson, G. D. (2004). How well does IQ predict neuropsychological test performance in normal adults? *J.Int.Neuropsychol.Soc., 10,* 82-90.

Dikmen, S. S., Heaton, R. K., Grant, I., & Temkin, N. R. (1999). Test-retest reliability and practice effects of expanded Halstead-Reitan Neuropsychological Test Battery. *J.Int.Neuropsychol.Soc., 5,* 346-356.

Ehrenstein, W. H., Heister, G., & Cohen, R. (1982). Trail Making Test and visual search. *Arch.Psychiatr.Nervenkr., 231,* 333-338.

Hester, R. L., Kinsella, G. J., Ong, B., & McGregor, J. (2005). Demographic influences on baseline and derived scores from the trail making test in healthy older Australian adults. *Clin.Neuropsychol., 19,* 45-54.

Matarazzo, J. D., Wiens, A. N., Matarazzo, R. G., & Goldstein, S. G. (1974). Psychometric and clinical test-retest reliability of the Halstead impairment index in a sample of healthy, young, normal men. *J.Nerv.Ment.Dis., 158,* 37-49.

	Stout, J. C., Paulsen, J. S., Queller, S., Solomon, A. C., Whitlock, K. B., Campbell, J. C. et al. (2011). Neurocognitive signs in prodromal Huntington disease. *Neuropsychology., 25,* 1-14. Justin J. F. O'Rourke, Leigh J. Beglinger, Megan M. Smith, James Mills, David J. Moser, Kelly C. Rowe, Douglas R. Langbehn, Kevin Duff, Julie C. Stout, Deborah L. Harrington, Noelle Carlozzi, Jane S. Paulsen & the PREDICT-HD Investigators of the Huntington Study Group (2011): The Trail Making Test in prodromal Huntington disease: Contributions of disease progression to test performance, Journal of Clinical and Experimental Neuropsychology, 33:5, 567-579. Other References: Specialty Automated Systems Corporation offers an online version of the Trail Making Test. For more information, visit: http://www.trailmakingtest.com/.

Description of Verbal Fluency Tests (Category) for HD Common Data Elements

Instrument Name:	Verbal Fluency Tests (Category)
Classification:	Supplemental
Short Description of Instrument:	Summary/ Overview of Instrument: The participant is required to verbally generate items from a given semantic category. The most frequently used category is animals; however, other categories used include: fruits and vegetables, items of clothing, things found in a supermarket. Construct measured: Generation, executive function, semantic knowledge Generic vs. disease specific: Generic Intended use of instrument/ purpose of tool (cross-sectional, longitudinal, diagnostic, etc): The category fluency task is a measure of fluency in verbal generation of semantic category members. It may be used as a longitudinal marker of disease severity in manifest/premanifest HD or as a cross-sectional measure of cognitive impairment across disease stages or between manifest/premanifest HD and healthy controls. Means of administration (paper and pencil, computerized): Verbal (responses recorded on paper verbatim for later scoring if needed). Location of administration (clinic, home, telephone, etc): Clinic Intended respondent (patient, caregiver, etc.): Patient # of items: N/A # of subscales and names of sub-scales: N/A
Scoring	Scoring (include reference to detailed scoring instructions, including calculation of a total score and subscale scores, and any limitations of scale or scoring posed by item nonresponse): The score is the total number of correct items produced in one minute. Error responses (e.g. perseverations, intrusions) are often also recorded. Standardization of scores to a reference population (z scores, T scores, etc): Normative data, stratified by age and education level are available for the general population (Tombaugh et al., 1999). If scores have been standardized to a reference population, indicate frame of reference for scoring (general population, HD subjects, other disease groups, etc). Normative data are available for the general population.

Description of Verbal Fluency Tests (Category) for HD Common Data Elements

Measurements	Type of scale used to describe individual items and total/subscale scores (nominal, ordinal, or [essentially] continuous): continuous If ordinal or continuous, explain if ceiling or floor effects are to be expected if the measure is used in specific HD Subgroups. No ceiling effects expected. Floor effect performance would only be anticipated in severely advanced stage HD.
Psychometric Properties	Reliability: Test-retest or intra-interview (within rater) reliability (as applicable): Inter-interview (between-rater) reliability (as applicable): Internal consistency: Statistical methods used to assess reliability: Validity: Category fluency scores are moderately correlated with Phonemic Fluency scores (*r*=.52) (Tombaugh et al., 1999), another verbal generation task. Sensitivity to Change/ Ability to Detect Change (over time or in response to an intervention): Meta-analysis of HD observational studies published between 1993-2007 reveals longitudinal change within pre-HD and manifest HD over time. Pre-HD cross sectional: Effect size = -0.11 (95% CI= -0.34, 0.12) based on 9 studies including 126 participants. Pre-HD longitudinal: Effect size = -0.40 (95% CI= -0.99, 0.19) based on 1 study including 12 participants. Manifest HD cross sectional: Effect size = -1.34 (95% CI= -1.90, -0.78) based on 4 studies including 40 participants. Manifest HD longitudinal: Effect size = -0.50 (95% CI= -1.06, 0.06) based on 3 studies including 102 participants. Known Relationships to Other Variables (e.g. gender, education, age, etc): Performance improves with years of education and decreases with age, with education accounting for 13.6% of variance and age 23.4% (Tombaugh et al, 1999). Diagnostic Sensitivity and Specificity, if applicable (in general population, HD population- premanifest/ manifest, other disease groups): N/A

Description of Verbal Fluency Tests (Category) for HD Common Data Elements

Rationale/ Justification (include any information on language and countries/ cultures/ ethnic groups where tested)	Strengths: Task is sensitive to longitudinal change in pre-manifest and manifest HD. The available evidence suggests this test may be more sensitive than phonemic fluency. Category fluency is less sensitive to education than phonemic fluency (Tombaugh et al, 1999). The task is easily administered and well-tolerated amongst patient groups. Weaknesses: Results might vary depending upon the choice of the category used in the administration (e.g., animals vs fruits, etc.) Availability (copyright): Public Domain Special Requirements for administration: Stopwatch required Administration Time: One minute per category Translations available (e.g. Spanish, French, Other languages): N/A
References:	Key Reference: Mitrushina, M.M., Boone, K.B., Razani, J., & D'Elia, L.F., (2005). Handbook of normative data for neuropsychological assessment (2nd ed.). New York: Oxford University Press. Other References: Tombaugh, T.N., Kozak, J., Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. Archives of Clinical Neuropsychology, 14, 167-177.