WILEY MOVEMENT DISORDERS CLINICAL PRACTICE

Rating Scales and Performance-based Measures For Functional Ability In Huntington Disease

Journal:	*Movement Disorders Clinical Practice*
Manuscript ID	MDCP-17-0262.R1
Wiley - Manuscript type:	Movement Disorders Society Paper
Date Submitted by the Author:	n/a
Keywords:	Huntington's disease, rating scales, performance measures, physical function, functional ability

Complete List of Authors:
Mestre, Tiago; University of Ottawa, The Ottawa Hospital, Parkinson's Disease and Movement Disorders Clinic;
Busse, Monica; Cardiff University, Centre for Trials Research;
Davis, Aileen; University of Toronto, Krembil Research Institute, University Health Network and Institute of Health Policy, Management and Evaluation and Rehabilitation Institute
Quinn, Lori; Teachers College Columbia University, Department of Biobehavioral Sciences
Rodrigues, Filipe; Faculty of Medicine, University of Lisbon, Laboratory of Clinical Pharmacology and Therapeutics; Instituto de Medicina Molecular, Clinical Pharmacology Unit; University College London, Huntington's Disease Centre, Institute of Neurology
Burgunder, Jean-Marc; University of Bern, Neurology and Clinical Research
Carlozzi, Noelle; University of Michigan, Department of Physical Medicine and Rehabilitation
Walker, Francis; Wake Forest University School of Medicine, Department of Neurology
Ho, Aileen; University of Reading,
Sampaio, Cristina; CHDI management/ChDi foundation, ; Laboratory of Clinical Pharmacology and Therapeutics, Faculty of Medicine
Goetz, Christopher; Rush University Medical Center, Department of Neurological Sciences
Cubo, Esther; Hospital Universitario Hermanos Yague, Neurology
Martinez-Martin, Pablo; National Center of Epidemiology and CIBERNED, Carlos III Institute of Health
Stebbins, Glenn; Rush University Medical Center, Neurological Sciences

Abstract:
Limitation of functional ability is a major feature of Huntington's disease (HD). The International Parkinson and Movement Disorder Society (MDS) commissioned the appraisal of the use and clinimetric properties of clinical measures of functional ability that have been applied in HD studies and trials to date, to make recommendations regarding their use based on standardized criteria. After a systematic literature search, we included a total of 29 clinical measures grouped into two categories: 1) performance-based measures (e.g., balance, walking, reaching/grasping), and 2) rating

	scales. Three performance-based measures are rated as "recommended": the Tinetti Mobility Test for screening of fall risk and for severity assessment of mobility in patients with manifest HD (up to stage III); the Berg Balance Scale for severity of balance impairment; and the Six-Minute Walk Test for assessment of walking endurance (severity) in HD subjects with preserved ambulation. No rating scale targeting functional ability reached a "recommended" status, either for screening or severity measurement. The main challenges identified in this review include applying widely accepted conceptual frameworks to the identified measures, the lack of validation of clinical measures to detect change over time, and absence of validated measures for upper limb function. Furthermore, measures of capacity or ability to perform activities of daily living had ceiling effects in people with early and pre-manifest HD. We recommend that the MDS prioritize the development of new scales that capture small but meaningful changes in function over time for outcome assessment in clinical trials, particularly in earlier stages of HD.

Rating Scales and Performance-based Measures For Assessment of Functional Ability In Huntington's Disease: Critique And Recommendations

Tiago A. Mestre MD MSc,[1*] Monica Busse BSc. BSc (Med) Hons. MSc (Med) PhD,[2] Aileen M. Davis PhD,[3] Lori Quinn, EdD, PT,[4] Filipe B. Rodrigues, MD, [5] Jean-Marc Burgunder, [6] Noelle E. Carlozzi PhD,[7] Francis Walker MD,[8] Aileen K. Ho PhD,[9] Cristina Sampaio MD PhD,[10] Christopher G. Goetz MD,[11] Esther Cubo MD,[12] Pablo Martinez-Martin PhD,[13] Glenn T. Stebbins PhD,[11] and the Members of the MDS Committee on Rating Scales Development

1 Parkinson's disease and Movement Disorders Center, Division of Neurology, Department of Medicine, The Ottawa Hospital Research Institute, University of Ottawa Brain and Mind Institute, Canada.

2 Centre for Trials Research, Cardiff University, Wales, UK.

3 Krembil Research Institute, University Health Network and Institute of Health Policy, Management and Evaluation and Rehabilitation Institute, University of Toronto, Canada.

4 Department of Biobehavioral Sciences, Teachers College, Columbia University, USA.

5 Huntington's Disease Centre, Institute of Neurology, University College London, UK
Clinical Pharmacology Unit, Instituto de Medicina Molecular, Portugal
Laboratory of Clinical Pharmacology and Therapeutics, Faculty of Medicine, University of Lisbon, Portugal

6 Swiss HD Center, NeuroZentrumSiloah and Department of Neurology, University of Bern, Switzerland

7 Department of Physical Medicine and Rehabilitation, University of Michigan, USA.

8 Department of Neurology, Wake Forest School of Medicine, USA

9 School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK.

10 CHDI Foundation/CHDI management, USA

11 Department of Neurological Sciences, Rush University Medical Center, Chicago, USA

12 Department of Neurology, Hospital Universitário HermanosYagüe, Burgos, Spain

13 National Center of Epidemiology and CIBERNED, Carlos III Institute of Health, Madrid, Spain

*** Corresponding author:**

Tiago A. Mestre

Parkinson's disease and Movement Disorders Center

Division of Neurology, Department of Medicine

University of Ottawa

1053 Carling Avenue, Ottawa ON K1Y 4E9, Canada

Telephone: +1 613 979 1513; E-mail: tmestre@toh.on.ca

Potential conflict of interest: Nothing to report

Word count of abstract: 267

Word count of main text: 4348

Number of tables: 3

Key words: Huntington's disease, physical function, performance measures, rating scales

Abstract

Limitation of functional ability is a major feature of Huntington's disease (HD). The International Parkinson and Movement Disorder Society (MDS) commissioned the appraisal of the use and clinimetric properties of clinical measures of functional ability that have been applied in HD studies and trials to date, to make recommendations regarding their use based on standardized criteria. After a systematic literature search, we included a total of 29 clinical measures grouped into two categories: 1) performance-based measures (e.g., balance, walking, reaching/grasping), and 2) rating scales. Three performance-based measures are rated as "recommended": the Tinetti Mobility Test for screening of fall risk and for severity assessment of mobility in patients with manifest HD (up to stage III); the Berg Balance Scale for severity of balance impairment; and the Six-Minute Walk Test for assessment of walking endurance (severity) in HD subjects with preserved ambulation. No rating scale targeting functional ability reached a "recommended" status, either for screening or severity measurement.

The main challenges identified in this review include applying widely accepted conceptual frameworks to the identified measures, the lack of validation of clinical measures to detect change over time, and absence of validated measures for upper limb function. Furthermore, measures of capacity or ability to perform activities of daily living had ceiling effects in people with early and pre-manifest HD. We recommend that the MDS prioritize the development of new scales that capture small but meaningful changes in function over time for outcome assessment in clinical trials, particularly in earlier stages of HD.

Introduction

The ability to perform daily life activities depends on the integration of motor, cognitive and behavioral functioning. These domains are progressively impaired in Huntington's disease (HD). A measure of functional ability based on key life activities is thus an attractive outcome in clinical studies, namely for treatment trials. A single measure pertinent to patient overall function would be useful to capture changes occurring simultaneously in the different symptom domains in HD. Further, functional ability measures are valued as an outcome for drug development by regulatory agencies.[1]

There is a need to identify and critically appraise the measurement properties of clinical measures currently used to capture functional ability in people with HD to inform optimal application in clinical research. The scope of this review is directed towards physical function and included a wide spectrum of clinical measures from those capturing motor tasks such as walking and balance ability, to those assessing the ability to perform activities of daily living (ADL).

The current review aims to provide recommendations and identify gaps in the use and validation of these functional measures that have been used in HD studies and trials to date. Such information will inform the field, identifying where additional testing of measurement properties or development of new measures may be required.

METHODS

We followed the methodology proposed by the MDS Committee on Rating Scales Development described elsewhere,[2] and includes i) Organization and Critique Process, ii) Selection of Scales, iii) Inclusion/Exclusion for Review, iv) Criteria for Rating Scales

Recommendation (Table 1). For selection of measures, the keywords selected for this review were "Huntington*" OR "Westphal variant" OR "juvenile Huntington*", and the terms "scale" OR "questionnaire" OR "index" OR "measure" as well as keywords: "function", "activit* daily li*", "capacity", "*ability", "impairment". Manuscripts published before October 17, 2016 were retrieved using the above search strategy and thoroughly screened by the chair of the sub-committee (T.A.M.) to ascertain which clinical measure had been used in each study. To aid our categorization of clinical measures in this review, we applied a widely accepted classification of the health components of functioning and disability: The International Classification of Functioning, Disability and Health (ICF).[3] The ICF defines: 1) impairments or problems in body function or structure such as a significant deviation or loss, 2) activity or the execution of a task, 3) and participation or involvement in a life situation.[3] By consensus, we included clinical measures in this review that captured a) activity or the execution of a task or tasks, and b) participation or involvement in a life situation.

Identified Clinical Measures and Their Utilization in Clinical Research

A total of 47 potentially relevant clinical measures were identified. After screening for exclusion criteria with abstract screening and in-depth review, a total of 29 measures were included and divided in performance-based measures defined as functional assessments based on the live performance of a task (e.g., balance, walking, reaching/grasping) (n=17) and rating scales (n=12) capturing the assessment of various aspects of functional ability based on recall. (See *Supplementary material* for more details)

Critique of Measures of Functional Ability

We provide a summary description of the performance-based measures and rating scales classified as "recommended" or "suggested". See *Supplementary material* for a full description of all clinical measures included for full review, including those that were included in the "suggested *with caveats*" or "listed" categories.

1) Performance-based measures

"RECOMMENDED"

Tinetti Mobility Test (TMT)

The TMT is a 16-item clinician-administered performance measure, which consists of balance and gait subscales that measure static and dynamic balance. It was originally developed to measure balance and screen for risk of falls in the elderly,[4] but has been used in other patient populations.[4] During the 10-15 minute test, patients perform a series of balance and walking tasks and are rated on a 0-2 scale based on qualitative assessment of performance.[4] The TMT has been used in several studies in HD and demonstrates good test-retest reliability in early-, mid-, and late stage HD (ICC = 0.8-0.9).[5, 6] Higher scores in the TMT correlated positively with spatio-temporal measures of gait (e.g., velocity r=0.68; stride length r=0.74), with higher scores of the UHDRS-FAS (r=0.44) and UHDRS-TFC (r=0.42) and lower scores of the UHDRS-Total Motor Score (TMS) (r=-0.59).[5, 7, 8] The TMT has demonstrated responsiveness in the context of interventional studies, including an intensive rehabilitation intervention program in patients with HD stages I–III (pre= 15.97, post=20.79, $p<0.001$), [9] and off- (17.09 ± 4.04) and on-tetrabenazine (19.91 ± 3.53, $p<0.02$) study of manifest HD patients. [10] However, there was no significant change in the TMT following a video-based balance training program.[11] A cut-off score of 21 has 74% sensitivity and 60% specificity in identifying fallers in HD.[5]

Recommendation: The TMT is "recommended" for assessment of mobility in patients with manifest HD (up to stage III) and "recommended" for screening for risk of falls .

The Berg Balance Scale (BBS)

The BBS is a performance measure consisting of 14 subtests of various activities related to balance that takes 10 to 15 minutes to complete. These activities include static postures (e.g., sitting, standing), transitions (e.g., sitting to standing, transferring between chairs), and challenging positions (e.g., standing with eyes closed). Quality of performance for each item is scored using a 4-point scale, with higher scores indicating better balance, and a possible maximum score of 56. Although originally developed to measure balance in older people, the BBS has been widely used in HD, although it has limited applicability in non-ambulatory HD due to the nature of the activities.[6, 12-19] The available clinimetric data show that it has good test-retest reliability in both pre-manifest (ICC=0.86) and manifest HD (ICC=0.96).[6] A minimal detectable change (MDC) of 5 in people with manifest HD has been reported.[6] Convergent validity has been reported between the BBS and the HD-ADL (r= -0.47), UHDRS TFC (r=0.60[19] and r=0.43[7]), UHDRS-FAS (r=0.48)[7], and UHDRS-TMS (r= -0.55).[7] Sensitivity to change following treatment withdrawal (tetrabenazine) was reported in a small open-label cohort.[14] A cut-off score of 40 was used as a cut-off to predict being a "faller" for a plotted probability of 60%.[86]

Recommendation: The BBS is "recommended" for assessing severity of balance impairment in ambulatory HD. The BBS is "suggested" for screening for fall risk, as no sensitivity or specificity data for falls have been reported.

The Six-Minute Walk Test

The Six-Minute Walk test measures how many meters an individual can walk in 6 minutes. [20, 21] Two practice tests are recommended, but not always carried out.[22, 23] It has been applied as a measure of endurance in neurological conditions, in contrast to shorter walk tests that generally measure velocity of walking speed.[6] It has been used in patients with pre-manifest and manifest HD, although it cannot be used for those who are non-ambulatory. Excellent test-retest reliability data have been reported in pre-manifest (ICC = 0.98) and manifest HD (IC=0.94; early and late HD = 0.97, and mid-stage HD=0.86).[6, 24] It is unclear how values discriminate among pre- and manifest HD severity levels as there is an overlap of the 95% confidence interval (CI) around mean values in both groups. On the other hand, values may separate pre- and early manifest HD from mid- to late stage HD.[6] Low correlations have been reported between the Six-Minute Walk Test and the UHDRS-FAS,[7] but higher correlations are not expected due to the limited overlap of the measure constructs. The MDC has been reported to be 39.2 meters for pre-manifest HD and 86.6 meters for manifest HD (range: 56.6 to 126.1 meters).[6]

Recommendation: The Six-Minute Walk test is "recommended" for the assessment of walking endurance (severity) across HD severity.

"SUGGESTED"

Timed 'up and go' Test (TUG)

The TUG is a simple and quick (<3 minutes) to use test that assesses mobility, balance and risk of falls. Although not specifically developed for use in HD, it has been used in pre-manifest and manifest HD to measure severity and screen for risk of falls.[13, 25] The TUG

measures the time it takes for a patient to rise from a chair, walk three meters, turn around, walk back to the chair, and sit down. One practice test is recommended before scoring the test.[25] Mean scores for patients with manifest HD range from 9-17 seconds [6, 19] and a cut-off score of 14 seconds has been reported to predict being a "faller" for a plotted probability of 60%.[13] Test-retest reliability in HD has been shown to be excellent (ICC = 0.93 [pre-manifest HD], 0.96 [manifest HD]) and the MDC has been reported to be 1.34 seconds in pre-manifest HD and 2.98 seconds in manifest HD.[6] The TUG was not statistically significantly correlated with the UHDRS-TMS or the UHDRS-TFC and correlated weakly with the UHDRS-FAS ($r= -0.33$, $p<0.01$).[7] Pre-post scores improved by an average of 1.3 seconds following training in a non-controlled study, that follow within the MDC.[26] The TUG can be used in early to mid-stages of HD, but not in pre-manifest or late stage HD, and it appears to be sensitive to disease progression, but does not discriminate between disease subtypes.[6, 19, 27]

Recommendation: The TUG is "suggested" for assessing severity of balance and mobility, and "suggested" for screening for fall risk. There is no sensitivity or specificity data for the reported cut-off point. Construct validity needs further assessment.

The Ten-Meter Walk Test

The Ten-Meter Walk test is a quick and easy performance-based measure that assesses walking speed. The score is based on the mean of two tests. The test has been used in pre-manifest and manifest HD with varying walking speeds: self-paced[6, 7, 24] and fast-paced.[6, 17, 24] Test-retest reliability has been shown to be good in both pre-manifest and manifest HD for the self-paced version.[6] For the self-paced version there was no correlation with the UHDRS-TMS, a weak correlation was reported with the UHDRS-FAS ($r=0.35$, $p<0.01$) and none with the UHDRS-TFC.[7] The fast-paced version of the test has been shown to be sensitive to

change following a rehabilitation program intervention in mild to moderate manifest HD (improvement of 0.27 m/s).[17] Following a 12-week community-based exercise program there was no significant change for either the self- or fast-paced versions.[24]

Recommendation: The Ten-Meter walk test is "suggested" for assessing walking speed in manifest HD. The vast majority of the clinimetric data sustaining this recommendation was obtained using the self-paced version.

Four Square Step Test (FSST)

The FSST is a 5-10 minute test of dynamic balance. The FSST clinically assesses a patient's ability to step over canes positioned in a cross shape in three directions in a set sequence: forward, sideways, and backwards. The test was not specifically developed for use in HD, but has been used in three studies in HD, and some clinimetric data are available in pre- and manifest HD.[6, 8, 11] Test-retest reliability has been reported to be excellent in pre-manifest HD (ICC=0.91), and good in manifest HD (ICC=0.78).[6] The MDC is higher in manifest HD (15.2) than in pre-manifest HD (1.9).[6] Moderate to high correlation has been shown between the FSST and the ABC (Pearson correlations: –0.57; $p<0.05$); the Tinetti Mobility Test (Pearson correlations: –0.67, $p<0.01$), and gait velocity (Pearson correlations: –0.69, $p<0.01$).[8] The FSST has not been shown to be sensitive to change in one exercise study.[11]

Recommendation: The FSST is "suggested" for assessing dynamic balance in HD

Mini Balance Evaluation Systems Test (Mini-BESTest)

The Mini-BESTest is a 14-item measure of dynamic balance. Derived from the Balance Evaluation Systems Test (BESTest), factor analysis was used for item reduction to include

dynamic balance only, and to improve clinical utilization.[28] Administered in 10-15 minutes, the Mini-BESTest evaluates domains of postural control. Each question is rated from normal to severe and scored between 0 and 2, for a maximum total score of 28 points. The test was not specifically developed for HD, and has not been assessed comprehensively across stages of HD. The test is not applicable to non-ambulatory patients.[29] Convergent validity has been shown between the Mini-BESTest and the ABC (r^2=0.45), UHDRS-TFC (r^2=0.75) and UHDRS-TMS (r^2=0.68).[29]

Recommendation: The Mini-BESTest is "suggested" for assessing severity of balance impairment in HD, as it has been used in only one study with a very small sample size across HD severity with a partial clinimetric assessment.

Physical Performance Test (PPT)

The PPT is a ten-minute test, which assesses multiple domains of physical function using observed performance of tasks that simulate activities of daily living (ADL) of various degrees of difficulty (writing, eating, dressing, walking, and climbing stairs).[30] Each activity is timed and rated from 0-4, a higher score indicating better physical performance. The test was not specifically developed for use in HD, but some of its clinimetric properties have been assessed in both pre- and manifest HD. Good test-retest reliability has been recorded in pre-manifest HD (ICC = 0.76) and excellent reliability in manifest HD (ICC=0.95). The MDC was 3 points for pre-manifest HD and 5 points for manifest HD respectively.[6] Convergent validity has been reported in manifest HD between the PPT and the UHDRS-TMS (r = -0.41 n=63, $p<0.01$), the UHDRS-FAS (r = 0.59, $p<0.01$); and the UHDRS-TFC (r= 0.48, $p<0.05$).[7]A ceiling effect has been reported in pre-manifest HD.[6] It has also been shown to be valid in patients with cognitive impairment.[31]

Recommendation: The PPT is "suggested" for assessing severity of impairment of physical function in performance of tasks that simulate activities of daily living.

Six-condition Romberg test

The six-condition Romberg test is a 5-minute easy to administer performance-based measure of balance developed in the context of myelopathies and neuropathies with an associated sensory dysfunction. The amount of time the patient maintains the position without loss of balance for 6 standard conditions is recorded, for a maximum score of 180 seconds. Higher scores indicate better balance. The test has been used in some HD studies [6, 10] and the clinimetric data available document good test-retest reliability in both pre-manifest (ICC=0.73) and manifest HD (ICC=0.89).[6] The six-condition Romberg test is a valid tool that can be used across all stages of HD provided that the patient is ambulatory as it is likely to have floor effects in non-ambulatory patients.[6] It has not been shown to be sensitive to change in treatment.[10] People with pre-manifest HD (158.8±22.2) have higher scores (better performance) than those with manifest HD (70.0±41.1).[6]

Recommendation: The Six-Condition Romberg Test is "suggested" for assessing severity of balance impairment in HD

2) Rating Scales

"SUGGESTED":

The Unified Huntington's Disease Rating Scale (UHDRS) - Total Functional Capacity (TFC)

The UHDRS-TFC is part of a multi-component rating scale originally designed to prospectively evaluate all patients with HD and individuals at risk for HD.[34] It assesses capacity as opposed to actual performance, and consists of a 5-item interview between a clinician, and the patient and a person familiar with the patient's functioning. It takes < 5 minutes to complete and covers basic activities of living: occupation, handling finances, domestic responsibilities, ADLs such as eating, dressing, bathing, and level of care (home or facility). A higher score indicates better functional capacity. The UHDRS-TFC has been used in pre-manifest and manifest HD populations in multiple observational studies and randomized controlled trials.[34-51] The TFC total score can be categorized into Shoulson and Fahn HD stages.[35] There is evidence of excellent inter-rater reliability, but only for a modified version of the UHDRS-TFC that is filled by patient or the caregiver (ICC = 0.96, 95% CI: 0.92, 0.98).[52] Data from multiple studies suggest good convergent validity with other components of the UHDRS assessing the functional domain and quality of life, and good divergent validity with motor disability, cognitive deficits and behavioral problems.[19, 29, 34, 53-60] Extensive data from multiple observational studies and clinical trials suggest sensitivity to change over time.[34-51, 61-70] There appears to be a ceiling effect for early stage HD and a floor effect for late stage HD.[41]

Recommendation: The UHDRS-TFC is "suggested" for assessing severity of limitation in functional capacity in HD, because it lacks core clinimetric data, namely, test-retest reliability and internal consistency to reach a "recommended" status.

The UHDRS - Functional Assessment Scale (FAS)

The UHDRS-FAS is an extensively-used checklist that is also part of the UHDRS. It is a clinician-administered questionnaire with 25 items which screen an individual's capacity to complete specific tasks, enables the clinician to assess severity, and make longitudinal assessments. The questionnaire takes 5-10 minutes to complete. It is considered an extension of the TFC and is more detailed in certain tasks.[34] A total score is obtained by giving 1 point to all "yes" replies, and a higher score indicates better functioning.[34] It has been used in multiple observational studies and randomized controlled trials in manifest HD populations.[34, 39, 43, 48, 49, 61, 62, 64, 68, 70-72] The UHDRS-FAS has been shown to have high internal consistency (Cronbach's $\alpha = 0.95$).[34] There are no available data on test-retest reliability or inter-rater reliability. Good convergent validity with other components of the UHDRS has been shown, as well as with motor disability, cognitive and behavioral deficits.[34, 54, 58, 73, 74] The UHDRS-FAS has been shown to be sensitive to change over time in several studies.[34, 39, 42, 43, 48, 49, 61, 62, 64, 68, 70, 71, 75]

Recommendation: The UHDRS-FAS is "suggested" for assessing severity of limitation in functional capacity in HD, because it lacks core clinimetric data, namely, test-retest or inter-rater reliability data.

The UHDRS-Independence Scale (IS)

The UHDRS-IS is a clinician-rated tool which assesses the actual reduction of functional ability.[76] It is rated from 100 (no special care needed) to 0 (tube-fed, total bed care) and takes approximately 5 minutes to complete. It has been used in many observational and randomized controlled trials in manifest HD populations.[34, 41-44, 46, 48-50, 61, 62, 64, 68, 70] The clinimetric data available show that the UHDRS-IS has moderate inter-rater reliability but in a modified

version that compares caregiver report with patient self-report (ICC = 0.71, 95% CI: 0.48, 0.85). [59] Good correlation with other components of the UHDRS, as well as motor disability, cognitive and behavioral deficits has been shown in various studies.[34, 54, 58, 59, 73, 76-79] Data from clinical trials suggest sensitivity of the UHDRS-IS to change over time and across disease stages.[35, 41]

Recommendation: The UHDRS-IS is "suggested" for assessing severity of limitation in functional ability in HD, because reliability data are missing, including test-retest, inter-rater (for clinicians) and internal consistency.

HD Activities of Daily Living (HD-ADL) 17-item

The HD-ADL Scale, which was developed to be used specifically in HD, was modeled after the Scale for Instrumental Activities of Daily Living.[80] It is a 17-item informant-completed instrument on which the informant rates the HD patient's ability to perform specific activities, covering the domains of personal care, household care, work and money, social relationships, and communication. For each item, the patient is rated on a 4-point scale, from normal to severely limited. The total score of the HD-ADL scale ranges from 0 (normal) to 51 (maximal limitation). [53] With exception of one study,[19] the scale has not been used outside the John Hopkins group who developed it. Clinimetric testing show that the HD-ADL has good internal consistency (α = 0.91-0.96).[53] Principal Component Analysis showed that four factors account for 72-74% of the total variance.[53] Convergent validity has been shown between the total score of the HD-ADL and the UHDRS-TFC ($r = -0.89$, $p < 0.001$), as well as all factors except for the domain "family relationships".[53] Multiple correlations have been reported with measures of cognitive impairment or disease duration.[53, 81, 82] The HD-ADL failed to show differences in treatment compared to placebo.[83, 84]

Recommendation: The HD-ADL is "suggested" for assessing severity of limitation in ADL, because studies of the scale's clinimetric properties are lacking, namely for any type of reliability.

Activity-Specific Balance Scale (ABC)

The ABC is a patient-completed scale that measures balance confidence and fear of falling. The ABC can take anywhere between 6 and 30 minutes to complete depending on the patient. Although it is a self-administered scale, a face-to-face interview is recommended.[85] Patients rate their balance confidence on a visual analogue scale ranging from 0 to 100 for each of 16 tasks, with higher scores indicating greater confidence and lower fall risk. The ABC has been widely used in HD,[8, 17, 29] including a modified ABC-UK version adapted for British culture,[86] but normative cut-off scores have not been established. The clinimetric data available show that the ABC has good test-retest reliability (ICC = 0.74 95% CI: 0.58, 1.0),[8] the MDC has been reported to be 27.33.[8] There is good convergent validity with the Mini-BESTest,[29] and the modified ABC-UK can distinguish between non-fallers and fallers in HD (mean score: 77.5 vs. 47.9).[86] While the ABC has been shown to be sensitive to change in one study (after a 9-month multidisciplinary rehabilitation program),[87] no change was reported in two other studies.[8, 17]

Recommendation: The ABC is "suggested" for assessing level of self-reported balance confidence in HD. The use of the ABC is challenged since the lack of insight is a feature of HD.

Rivermead Mobility Index (RMI)

The RMI is an extension of the Rivermead Motor Assessment Gross Function Scale that assesses functional mobility and was initially developed for stroke. The RMI consists of 14 questions about a patient's ability to perform a wide range of activities, from turning over in bed to running, and one observation (standing for 10 seconds without any aid). Questions are answered as "able" (1 point) or "unable" (0 points) and summed to produce a total score, with a higher score reflecting better mobility.[88] Test-retest reliability has been reported in HD (ICC in pre-manifest HD= 0.81; ICC in manifest HD = 0.94).[6] A MDC of 2 points has been reported in manifest HD; ceiling effects are present in pre-manifest HD.[6] There are no cut-off scores established in HD, which limits its use as a screening tool in HD.

Recommendation: The RMI is "suggested" for the assessment of severity of restriction of mobility

Discussion

The current critique focuses on performance-based measures and rating scales assessing functional ability in HD. In the process of developing the protocol for the review, we found a variety of scale constructs and other instruments that could be associated with various aspects of function ability. We used the ICF[3] as a conceptual framework related with function to guide us in the inclusion or exclusion of rating scales based on the adequacy of their constructs. Nevertheless, we realize that the measures included in this review represent a wide variety of concepts that apply across the components of the ICF. Many of these measures included multiple ICF components, raising challenges for conceptual clarity and subsequent evaluation of validity. For example, balance can be seen as a sheer impairment but it can overlap with activity/function, depending how it is captured in a given clinical measure.

Considering these aspects, we decided to be inclusive and included balance measures in this review. Ultimately, there is a need for clear definitions for future measures to better enable validation and application in HD populations.

We identified and included a range of performance-based measures. We provide a "recommended" level of recommendation for both screening purposes related to balance, gait and/or risk of falling, and measurement of severity of impairment of specific motor tasks. There were however no "recommended" performance-based measures covering upper limb function. It is also important to emphasize that the majority of these performance measures were only used in ambulatory HD populations.

We did not identify a rating scale that met the criteria for "recommended". If further testing of the measurement properties is conducted, we agreed that UHDRS sub-scales related with function (TFC, FAS and IS) are in a good position to reach the higher level of recommendation in the future due to their widespread use, specific development in HD and known initial clinimetric development. For each one of these scales, important shortcomings in terms of clinimetric development were identified, namely incomplete reliability testing, which precluded a "recommended" level of recommendation. In addition, these scales have limiting ceiling effects that make them unattractive for use in earlier stages of HD. For example, the use of these UHDRS subscales in a clinical trial conducted with the purpose of capturing a disease-modifying effect in an ideal HD subgroup of individuals with a high level of functional ability would be performed at the cost of a prohibitively long trial duration to capture a meaningful change. Rating scales such as Functional Rating Scale Task force for pre-Huntington Disease 2 (FuRST-pHD)[89, 90] are currently being developed and are expected to fill this gap in the future.

The assessment of functional ability as a clinical outcome is deemed essential for therapeutic approval by regulatory agencies such as the FDA.[1] In this regulatory context, it is important to

emphasize that there was no recommendation for the purpose of measuring change over time in individuals or groups of subjects in either a pure observational study or in an interventional context. In fact, formal testing for responsiveness was missing in all the included rating scales, and important measures of reliability such as test-retest had not been evaluated in many cases. Along the same lines, there is a need to assess the validity of each rating scale in different subgroups of patients with HD, as these data are presently lacking for most of the measures. The knowledge about responsiveness and its variation in important patient subgroups can determine sample size requirements and help with the interpretation of clinical trial results, respectively.[1]

Looking towards the future, the committee concludes that there are well-validated performance-based measures that capture motor tasks such as walking or balance, but further clinimetric development is required for performance-based measures that capture other aspects of physical function such as upper limb function. For rating scales, including those evaluating activities of daily living, we are cannot endorse an existing scale at a "recommended" level and encourage the MDS to prioritize the development of such instruments for clinical care and research purposes. Further validation of HD-specific scales such as the UHDRS-TFC are warranted, as is the development of new scales designed to have greater sensitivity in capturing function in HD subgroups who have a relatively well preserved functional ability as measured by currently available rating scales.

Author Roles:

Tiago A. Mestre

1. Research project: A. Conception, B. Organization, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: A. Writing of the first draft; B. Review and Critique;

Monica Busse

1. Research project: A. Conception, B. Organization, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

Aileen M. Davis

1. Research project: A. Conception, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

Lori Quinn

1. Research project: A. Conception, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

Filipe B.Rodrigues

1. Research project: A. Conception, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

467

468 Jean-Marc Burgunder

469 1. Research project: A. Conception, C. Execution;

470 2. Statistical Analysis: not applicable;

471 3. Manuscript Preparation: B. Review and Critique;

472

473 Noelle Carlozzi

474 1. Research project: A. Conception, C. Execution;

475 2. Statistical Analysis: not applicable;

476 3. Manuscript Preparation: B. Review and Critique;

477

478 Francis Walker

479 1. Research project: A. Conception, C. Execution;

480 2. Statistical Analysis: not applicable;

481 3. Manuscript Preparation: B. Review and Critique;

482

483 Aileen Ho

484 1. Research project: A. Conception, C. Execution;

485 2. Statistical Analysis: not applicable;

486 3. Manuscript Preparation: B. Review and Critique;

487

488

489 Cristina Sampaio

490 1. Research project: A. Conception, C. Execution;

491 2. Statistical Analysis: not applicable;

3. Manuscript Preparation: B. Review and Critique;

Christopher G. Goetz

1. Research project: A. Conception;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

Esther Cubo

1. Research Project: B. Organization
2. Statistical Analysis: not applicable;
3. Manuscript Prep: B. Review and critique.

Pablo Martinez-Martin

1. Research Project: A. Conception
2. Statistical Analysis: not applicable;
3. Manuscript Prep: B. Review and critique.

Glenn T. Stebbins

4. Research Project: A. Conception
5. Statistical Analysis: not applicable;
6. Manuscript Prep: B. Review and critique.

Financial Disclosures:

Tiago A. Mestre

Financial disclosure related to research covered in this article: Consulting for CHDI Foundation/Management

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: Abbvie, CHDI Foundation/Management

Grants and Research: University of Ottawa Medical Associates, PSG/PDF, Parkinson Canada, Parkinson Research Consortium

Honoraria: International Parkinson and Movement Disorders Society, American Academy of Neurology, U of Ottawa, Abbvie

Intellectual Property Rights: none

Ownership interests: none

Royalties: none

Expert Testimony: none

Salary: University of Ottawa Medical Associates

Monica Busse

Financial disclosure related to research covered in this article: none

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: None

Grants and Research: European Framework funding, Health and Care Research Wales (HCRW), Wellcome Trust, Medical Research Council UK, Gossweiler Foundation, National Institute of Health Research (NIHR)

Honoraria: none

Intellectual Property Rights: none

Ownership interests: none

Royalties: none

Expert Testimony: none

Salary: Cardiff University

Aileen M. Davis

Financial disclosure related to research covered in this article: none

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: Flexion Therapeutics Inc

Grants and Research: none

Honoraria: Associate Editor of Osteoarthritis and Cartilage

Intellectual Property Rights: none

Ownership interests: none

Royalties: none

Expert Testimony: none

Salary: U of Toronto.

Lori Quinn

Financial disclosure related to research covered in this article: reimbursement for travel to meetings by Movement Disorders Society

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: None

Grants and Research: Huntington Study Group; Jacques and Gloria Gossweiller Foundation

Honoraria: None

Intellectual Property Rights: None

Ownership interests: None

Royalties: Elsevier Publishers for textbook Documentation for Rehabilitation: A guide to clinical decision making in physical therapy

Expert Testimony: None

Salary: None

Filipe B. Rodrigues

Financial disclosure related to research covered in this article: None

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: None

Grants and Research: CHDI Foundation Clinical Research Fellowship Award (Aug 2015 to Aug 2016)

Honoraria: None

Intellectual Property Rights: None

Ownership interests: None

Royalties: None

Expert Testimony: None

Salary: University College London

Jean-Marc Burgunder

Financial disclosure related to research covered in this article: None

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: Chair of the EHDN Executive Committee

Grants and Research: no conflicts

Honoraria: no conflicts

Intellectual Property Rights: none

Ownership interests: none

Royalties: none

Expert Testimony: none

Salary: no conflict

Noelle Carlozzi

Financial disclosure related to research covered in this article: None.

Full financial disclosure for the last 12 months:

Consulting and Advisory Board Membership with honoraria: Teva Pharmaceuticals; Boston Medical Center

Grants and Research: National Institute for Neurological Disorders and Stroke; National Institute of Nursing Research; National Institute on Aging; CHDI Foundation; Frankel Cardiovascular Center Micro Grant Award;

Honoraria: None

Intellectual Property Rights: None.

Ownership interests: None

Royalties: None.

Expert Testimony: None

Salary: University of Michigan

Francis Walker

Financial disclosure related to research covered in this article: none

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: none

Grants and Research: Grant support by Pfizer, Vaccinex, and Teva; interest free instrument loans from Monarch Medical, Terason, Natus.

Honoraria: Grifols

Intellectual Property Rights: none

Ownership interests: None

Royalties: Elsevier, Up To Date

Expert Testimony: none

Salary: Wake Forest School of Medicine

Aileen Ho

Financial disclosure related to research covered in this article: none

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: Pfizer, National Institute of Health and Care Excellence UK.

Grants and Research: National Institute of Health Research (NIHR), European Huntington's Disease Network.

Honoraria: none

Intellectual Property Rights: none

Ownership interests: none

Royalties: none

Expert Testimony: none

Salary: University of Reading

Cristina Sampaio

Financial disclosure related to research covered in this article: Salary: CHDI Management

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: I received honoraria from Nestle, vTv Therapeutics, Neurotrope Stealth.

Grants and Research: none

Honoraria: International Parkinson and Movement Disorders Society

Intellectual Property Rights: none

Ownership interests: none

Royalties: none

Expert Testimony: none

Salary: CHDI Management

Christopher G. Goetz

Financial disclosure related to research covered in this article: None

Full financial disclosure for the last 12 months:

Consulting or Advisory Board Membership with honoraria: Addex, Avanir, Boston Scientific, CHDI Foundation/CHDI management, Clevexel, Kanter Health, Oxford Biomedica, Pfizer, WebMD.

Grants/Research: Funding to Rush University Medical Center from NIH, Michael J. Fox Foundation for research conducted by Dr. Goetz. Dr. Goetz directs the Rush Parkinson's Disease Research Center that receives support from the Parkinson's Disease Foundation and some of these funds support Dr. Goetz's salary as well as his research efforts. He directs the translation program for the MDS-UPDRS and UDysRS and receives funds directed to Rush

University Medical Center from the International Parkinson and Movement Disorder Society (IPMDS) for this effort.

Honoraria: American Academy of Neurology, Captain James A Lovell Federal Health Care Center, University of Pennsylvania, University of Rochester

Intellectual Property Rights: none

Ownership interests: none

Royalties: Elsevier Publishers, Oxford University Press, Wolters Kluwer,

Salary: Rush University Medical Center

Esther Cubo

Financial disclosure related to research covered in this article: none.

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: Abbvie, Allergan

Grants and Research: Junta de Castilla y León, International Parkinson disease and Movement Disorder Society

Honoraria: none

Intellectual Property Rights: none

Ownership interests: none

Royalties: none

Expert Testimony: none

Salary: Hospital Universitário Burgos, Spain**.**

Pablo Martinez-Martin

Financial disclosure related to research covered in this article: None

690

691 **Full financial disclosure For the last 12 months:**

692 Consulting and Advisory Board Membership with honoraria: Abbvie.

693 Grants and Research: International Parkinson and Movement Disorder Society for the Pilot
694 Study of the MDS-Non-Motor Symptoms Scale.

695 Honoraria: Editorial Viguera and Movement Disorder Society; AbbVie

696 Intellectual Property Rights: None

697 Ownership interests: None

698 Royalties: None

699 Expert Testimony: None

700 Salary: Carlos III Institute of Health

701

702 **Glenn T. Stebbins**

703 **Financial disclosure related to research covered in this article:** None

704 **Full financial disclosure For the last 12 months:**

705 Consulting and Advisory Board Membership with honoraria: Acadia, Pharmaceuticals,
706 Adamas Pharmaceuticals, Inc., Ceregene, Inc., CHDI Management, Inc., Ingenix
707 Pharmaceutical Services (i3 Research), Neurocrine Biosciences, Inc., Pfizer, Inc., Ultragenyx
708 Pharmaceutical.

709 Grants and Research: National Institutes of Health, Michael J. Fox Foundation for
710 Parkinson's Research, Dystonia Coalition, CHDI, International Parkinson and Movement
711 Disorder Society, CBD Solutions.

712 Honoraria: International Parkinson and Movement Disorder Society, American Academy of
713 Neurology, Michael J. Fox Foundation for Parkinson's Research, Food and Drug
714 Administration.

715 Intellectual Property Rights: none

716

717 Ownership interests: none

718 Royalties: none

719 Expert Testimony: none

720 Salary: Rush University Medical Center

721

722

Acknowledgments

We would like to thank Anne-Marie Williams for the editorial support, and Theresa Bolton for the assistance in literature search of the current review.

726 **Ethical Compliance Statement:** The authors confirm that the approval of an institutional
727 review board was not required for this work. We confirm that we have read the Journal's
728 position on issues involved in ethical publication and affirm that this work is consistent with
729 those guidelines.
730
731

References

1. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;4:79.
2. Schrag A, Barone P, Brown RG, et al. Depression rating scales in Parkinson's disease: critique and recommendations. Mov Disord 2007;22:1077-1092.
3. Towards a Common Language for Functioning, Disability and Health: ICF, The International Classification of Functioning, Disability and Health. Geneva; 2002.
4. Tinetti ME, Williams TF, Mayewski R. Fall risk index for elderly patients based on number of chronic disabilities. Am J Med 1986;80:429-434.
5. Kloos AD, Kegelmeyer DA, Young GS, Kostyk SK. Fall risk assessment using the Tinetti mobility test in individuals with Huntington's disease. Mov Disord 2010;25:2838-2844.
6. Quinn L, Khalil H, Dawes H, et al. Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. Phys Ther 2013;93:942-956.
7. Busse M, Quinn L, Khalil H, McEwan K. Optimising mobility outcome measures in Huntington's disease. J Huntingtons Dis 2014;3:175-188.
8. Kloos AD, Fritz NE, Kostyk SK, Young GS, Kegelmeyer DA. Clinimetric properties of the Tinetti Mobility Test, Four Square Step Test, Activities-specific Balance Confidence Scale, and spatiotemporal gait measures in individuals with Huntington's disease. Gait Posture 2014;40:647-651.
9. Zinzi P, Salmaso D, De Grandis R, et al. Effects of an intensive rehabilitation programme on patients with Huntington's disease: a pilot study. Clin Rehabil 2007;21:603-613.
10. Kegelmeyer DA, Kloos AD, Fritz NE, Fiumedora MM, White SE, Kostyk SK. Impact of tetrabenazine on gait and functional mobility in individuals with Huntington's disease. J Neurol Sci 2014;347:219-223.
11. Kloos AD, Fritz NE, Kostyk SK, Young GS, Kegelmeyer DA. Video game play (Dance Dance Revolution) as a potential exercise therapy in Huntington's disease: a controlled clinical trial. Clinical rehabilitation 2013;27:972-982.
12. Bohlen S, Ekwall C, Hellstrom K, et al. Physical therapy in Huntington's disease--toward objective assessments? Eur J Neurol 2013;20:389-393.
13. Busse ME, Khalil H, Quinn L, Rosser AE. Physical therapy intervention for people with Huntington disease. Phys Ther 2008;88:820-831.
14. Fekete R, Davidson A, Jankovic J. Clinical assessment of the effect of tetrabenazine on functional scales in huntington disease: a pilot open label study. Tremor Other Hyperkinet Mov (N Y) 2012;2.
15. Ferrara JM, Mostile G, Hunter C, Adam OR, Jankovic J. Effect of tetrabenazine on motor function in patients with huntington disease. Neurol Ther 2012;1:5.
16. Khalil H, Quinn L, van Deursen R, et al. What effect does a structured home-based exercise programme have on people with Huntington's disease? A randomized, controlled pilot study. Clinical rehabilitation 2013;27:646-658.
17. Piira A, van Walsem MR, Mikalsen G, Nilsen KH, Knutsen S, Frich JC. Effects of a One Year Intensive Multidisciplinary Rehabilitation Program for Patients with Huntington's Disease: a Prospective Intervention Study. PLoS Curr 2013;5.
18. Quinn L, Debono K, Dawes H, et al. Task-specific training in Huntington disease: a randomized controlled feasibility trial. Phys Ther 2014;94:1555-1568.

19. Rao AK, Muratori L, Louis ED, Moskowitz CB, Marder KS. Clinical measurement of mobility and balance impairments in Huntington's disease: validity and responsiveness. Gait Posture 2009;29:433-436.
20. Balke B. A Simple Field Test for the Assessment of Physical Fitness. Rep 63-6. Rep Civ Aeromed Res Inst US 1963:1-8.
21. Butland RJ, Pang J, Gross ER, Woodcock AA, Geddes DM. Two-, six-, and 12-minute walking tests in respiratory disease. Br Med J (Clin Res Ed) 1982;284:1607-1608.
22. Guyatt GH, Thompson PJ, Berman LB, et al. How should we measure function in patients with chronic heart and lung disease? J Chronic Dis 1985;38:517-524.
23. Guyatt GH, Pugsley SO, Sullivan MJ, et al. Effect of encouragement on walking test performance. Thorax 1984;39:818-822.
24. Busse M, Quinn L, Debono K, et al. A randomized feasibility study of a 12-week community-based exercise program for people with Huntington's disease. J Neurol Phys Ther 2013;37:149-158.
25. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991;39:142-148.
26. Quinn L, Debono K, Dawes H, et al. Task-specific training in Huntington's disease: A randomised, controlled feasibility trial. In; 2014. p. A66-a67.
27. Rao AK, Louis ED, Marder KS. Clinical assessment of mobility and balance impairments in pre-symptomatic Huntington's disease. Gait Posture 2009;30:391-393.
28. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. J Rehabil Med 2010;42:323-331.
29. Jacobs JV, Boyd JT, Hogarth P, Horak FB. Domains and correlates of clinical balance impairment associated with Huntington's disease. Gait Posture 2015;41:867-870.
30. Reuben DB, Siu AL. An objective measure of physical function of elderly outpatients. The Physical Performance Test. J Am Geriatr Soc 1990;38:1105-1112.
31. Farrell MK, Rutt RA, Lusardi MM, Williams AK. Reliability of the Physical Performance Test in People with Dementia. Physical & Occupational Therapy In Geriatrics 2010;28:144-153.
32. Sharpened Romberg. 2013; http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1160 Last accessed: 6 February 2017.
33. Romberg Test. 2013; http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1173 Last accessed: 6 February 2017.
34. Huntington Study Group. Unified Huntington's Disease Rating Scale: reliability and consistency. Mov Disord 1996;11:136-142.
35. Shoulson I. Huntington disease: functional capacities in patients treated with neuroleptic and antidepressant drugs. Neurology 1981;31:1333-1335.
36. Shoulson I, Odoroff C, Oakes D, et al. A controlled clinical trial of baclofen as protective therapy in early Huntington's disease. Annals of neurology 1989;25:252-259.
37. Feigin A, Kieburtz K, Bordwell K, et al. Functional decline in Huntington's disease. Mov Disord 1995;10:211-214.
38. Como PG, Rubin AJ, O'Brien CF, et al. A controlled trial of fluoxetine in nondepressed patients with Huntington's disease. Movement disorders 1997;12:397-401.
39. Siesling S, van Vugt JP, Zwinderman KA, Kieburtz K, Roos RA. Unified Huntington's disease rating scale: a follow up. Mov Disord 1998;13:915-919.
40. Kremer B, Clark CM, Almqvist EW, et al. Influence of lamotrigine on progression of early Huntington disease: a randomized clinical trial. In; 1999. p. 1000-1011.

831 41. Marder K, Zhao H, Myers RH, et al. Rate of functional decline in Huntington's
832 disease. Huntington Study Group. Neurology 2000;54:452-458.
833 42. Huntington Study Group. A randomized, placebo-controlled trial of coenzyme Q10
834 and remacemide in Huntington's disease. Neurology 2001;57:397-404.
835 43. Huntington Study Group. Dosage effects of riluzole in Huntington's disease: A
836 multicenter placebo-controlled study. Neurology 2003;61:1551-1556.
837 44. Bonelli RM, Hodl AK, Hofmann P, Kapfhammer HP. Neuroprotection in Huntington's
838 disease: A 2-year study on minocycline. International Clinical Psychopharmacology
839 2004;19:337-342.
840 45. Tommaso M, Specchio N, Sciruicchio V, Difruscolo O, Specchio LM. Effects of
841 rivastigmine on motor and cognitive impairment in Huntington's disease. In; 2004. p. 1516-
842 1518.
843 46. Huntington Study G. Minocycline safety and tolerability in Huntington disease.
844 Neurology 2004;63:547-549.
845 47. de Tommaso M, Di Fruscolo O, Sciruicchio V, et al. Efficacy of levetiracetam in
846 Huntington disease. Clin Neuropharmacol 2005;28:280-284.
847 48. Puri BK, Leavitt BR, Hayden MR, et al. Ethyl-EPA in Huntington disease: a double-
848 blind, randomized, placebo-controlled trial. In; 2005. p. 286-292.
849 49. Cubo E, Shannon KM, Tracy D, et al. Effect of donepezil on motor and cognitive
850 function in Huntington disease. Neurology 2006;67:1268-1271.
851 50. Landwehrmeyer GB, Dubois B, De Yebenes JG, et al. Riluzole in Huntington's
852 disease: A 3-year, randomized controlled study. Annals of Neurology 2007;62:262-272.
853 51. Tommaso M, Difruscolo O, Sciruicchio V, Specchio N, Livrea P. Two years' follow-
854 up of rivastigmine treatment in Huntington disease. In; 2007. p. 43-46.
855 52. Carlozzi NE, Victorson D, Sung V, et al. HD-PRO-TRIAD Validation: A Patient-
856 reported Instrument for the Symptom Triad of Huntington's Disease. Tremor Other
857 Hyperkinet Mov (N Y) 2014;4:223.
858 53. Bylsma. Assessment of Adaptive Functioning in Huntington's Disease. Mov Disord
859 1993;8:183-190.
860 54. Siesling S, Zwinderman AH, van Vugt JP, Kieburtz K, Roos RA. A shortened version
861 of the motor section of the Unified Huntington's Disease Rating Scale. Mov Disord
862 1997;12:229-234.
863 55. Thompson JC, Snowden JS, Craufurd D, Neary D. Behavior in Huntington's disease:
864 dissociating cognition-based and mood-based changes. J Neuropsychiatry Clin Neurosci
865 2002;14:37-43.
866 56. Ready RE, Mathews M, Leserman A, Paulsen JS. Patient and caregiver quality of life
867 in Huntington's disease. Mov Disord 2008;23:721-726.
868 57. Ho AK, Gilbert AS, Mason SL, Goodman AO, Barker RA. Health-related quality of
869 life in Huntington's disease: Which factors matter most? Mov Disord 2009;24:574-578.
870 58. Youssov K, Dolbeau G, Maison P, et al. Unified Huntington's disease rating scale for
871 advanced patients: validation and follow-up study. Mov Disord 2013;28:1717-1723.
872 59. Carlozzi NE, Tulsky DS, Chiaravalloti ND, et al. NIH Toolbox Cognitive Battery
873 (NIHTB-CB): the NIHTB Pattern Comparison Processing Speed Test. J Int Neuropsychol Soc
874 2014;20:630-641.
875 60. Klempir J, Klempirova O, Spackova N, Zidovska J, Roth J. Unified Huntington's
876 disease rating scale: clinical practice and a critical approach. Funct Neurol 2006;21:217-221.
877 61. Ravina B, Romer M, Constantinescu R, et al. The relationship between CAG repeat
878 length and clinical progression in Huntington's disease. Mov Disord 2008;23:1223-1227.
879 62. Huntington Study Group DI. A futility study of minocycline in Huntington's disease.
880 Movement Disorders 2010;25:2219-2224.

63. Huntington Study Group. A randomized, placebo-controlled trial of coenzyme Q10 and remacemide in Huntington's disease. Neurology 2001;57:397-404.
64. Kieburtz K, McDermott MP, Voss TS, et al. A randomized, placebo-controlled trial of latrepirdine in Huntington disease. In; 2010. p. 154-160.
65. Tabrizi SJ, Scahill RI, Durr A, et al. Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: the 12-month longitudinal analysis. The Lancet Neurology 2011;10:31-42.
66. Tabrizi SJ, Reilmann R, Roos RAC, et al. Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data. The Lancet Neurology 2012;11:42-53.
67. Tabrizi SJ, Scahill RI, Owen G, et al. Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. The Lancet Neurology 2013;12:637-649.
68. Verbessem P, Lemiere J, Eijnde BO, et al. Creatine supplementation in Huntington's disease: a placebo-controlled pilot trial. In; 2003. p. 925-930.
69. Beglinger LJ, Adams WH, Langbehn D, et al. Results of the citalopram to enhance cognition in Huntington disease trial. Mov Disord 2014;29:401-405.
70. Sussmuth SD, Haider S, Landwehrmeyer GB, et al. An exploratory double-blind, randomized clinical trial with selisistat, a SirT1 inhibitor, in patients with Huntington's disease. British journal of clinical pharmacology 2015;79:465-476.
71. Vaddadi KS, Soosai E, Chiu E, Dingjan P. A randomised, placebo-controlled, double blind study of treatment of Huntington's disease with unsaturated fatty acids. In; 2002. p. 29-33.
72. Mahant N, McCusker EA, Byth K, Graham S. Huntington's disease: clinical correlates of disability and progression. Neurology 2003;61:1085-1092.
73. Tumas V, Camargos ST, Jalali PS, Galesso Ade P, Marques Jr W. Internal consistency of a Brazilian version of the unified Huntington's disease rating scale. Arq Neuropsiquiatr 2004;62:977-982.
74. Banaszkiewicz K, Sitek EJ, Rudzinska M, Soltan W, Slawek J, Szczudlik A. Huntington's disease from the patient, caregiver and physician's perspectives: three sides of the same coin? J Neural Transm 2012;119:1361-1365.
75. Huntington Study G. Tetrabenazine as antichorea therapy in Huntington disease: a randomized controlled trial. Neurology 2006;66:366-372.
76. Myers RH, Sax DS, Schoenfeld M, et al. Late onset of Huntington's disease. J Neurol Neurosurg Psychiatry 1985;48:530-534.
77. Myers RH, Sax DS, Koroshetz WJ, et al. Factors associated with slow progression in Huntington's disease. Arch Neurol 1991;48:800-804.
78. Ho AK, Robbins AO, Walters SJ, Kaptoge S, Sahakian BJ, Barker RA. Health-related quality of life in Huntington's disease: a comparison of two generic instruments, SF-36 and SIP. Mov Disord 2004;19:1341-1348.
79. Ho AK, Robbins AO, Barker RA. Huntington's disease patients have selective problems with insight. Mov Disord 2006;21:385-389.
80. Lawton MP. The functional assessment of elderly people. J Am Geriatr Soc 1971;19:465-481.
81. Brandt J, Strauss ME, Larus J, Jensen B, Folstein SE, Folstein MF. Clinical correlates of dementia and disability in Huntington's disease. J Clin Neuropsychol 1984;6:401-412.
82. Rothlind JC, Brandt J. A brief assessment of frontal and subcortical functions in dementia. J Neuropsychiatry Clin Neurosci 1993;5:73-77.

930 83. Peyser CE, Folstein M, Chase GA, et al. Trial of d-alpha-tocopherol in Huntington's
931 disease. American journal of psychiatry 1995;152:1771-1775.
932 84. Ranen NG, Peyser CE, Coyle JT, et al. A controlled trial of idebenone in Huntington's
933 disease. Mov Disord 1996;11:549-554.
934 85. Powell LE, Myers AM. The Activities-specific Balance Confidence (ABC) Scale. J
935 Gerontol A Biol Sci Med Sci 1995;50A:M28-34.
936 86. Busse ME, Wiles CM, Rosser AE. Mobility and falls in people with Huntington's
937 disease. J Neurol Neurosurg Psychiatry 2009;80:88-90.
938 87. Thompson JA, Cruickshank TM, Penailillo LE, et al. The effects of multidisciplinary
939 rehabilitation in patients with early-to-middle-stage Huntington's disease: a pilot study. Eur J
940 Neurol 2013;20:1325-1329.
941 88. Collen FM, Wade DT, Robb GF, Bradshaw CM. The Rivermead Mobility Index: a
942 further development of the Rivermead Motor Assessment. Int Disabil Stud 1991;13:50-54.
943 89. FuRST 2.0: Cognitive Pre-Testing Study for a New Functional Rating Scale for Use in
944 Huntington's Disease. 2017; https://clinicaltrials.gov/ct2/show/NCT02881931?term=Furst-
945 &rank=1 Last accessed: 01/24/2017.
946 90. Vaccarino AL, Sills T, Anderson KE, et al. Assessment of Day-to-Day Functioning in
947 Prodromal and Early Huntington Disease. PLoS Curr 2011;3:RRN1262.
948

Table 1: Classification System For Scale Recommendation

Category	Criteria
"Recommended"	(1) Scale has been used in HD populations. (2) Use in HD by groups other than the original developers and data on its use were available. * (3) The available clinimetric/psychometric data in HD supports the goals of screening (e.g., evaluation of sensitivity/specificity, score cut-points, and reliability) or measurement of severity (e.g., evaluation of reliability, construct validity, and score discrimination across levels of symptom severity), or measurement of a change in severity (e.g. responsiveness or sensitivity to change)
"Suggested"	(1) Scale has been used in HD populations. (2) Only one other criteria (2) or (3) from the above recommended category applies.
"Listed"	(1) Scale has been applied to HD populations, but no further criterion met.

HD=Huntington's Disease, * For rating scales not originally developed for use in HD, criterion 2 was fulfilled if used in at least one group in HD that reported any kind of clinimetric/psychometric data in HD.

Table 2: Summary of all included scales or instruments in HD

Scale/Questionnaire	Developed for use in HD	Scale has been applied to HD populations	Used by other groups beyond the original developing group	Appropriate clinimetric testing in HD	Recommendation level	COMMENTS
PERFORMANCE-BASED MEASURES						
Tinetti mobility test	No	Yes	Yes	Yes	Recommended for assessment of gait and balance problems in patients with manifest HD (up to stage III) Recommended for screening for risk of falls	
The Berg Balance Scale	No	Yes	Yes	Yes[1]/No[2]	[1] Recommended for assessing severity of balance impairment in HD with preserved ambulation [2] Suggested for screening risk of falls	
6-Minute Walk Test	No	Yes	Yes	Yes	Recommended for assessing walking endurance (severity) in HD with preserved ambulation	
Timed 'up and go' Test	No	Yes	Yes	No	Suggested for assessing balance and mobility (severity) Suggested for screening for risk of falls	

10 Meter walk Test	No	Yes	Yes	No	Suggested for assessing walking speed in manifest HD	
4 Square step test (FSST)	No	Yes	Yes	No	Suggested for assessing dynamic balance in HD	
Mini-BESTest	No	Yes	Yes	No	Suggested for assessing severity of balance impairment in HD	
Physical Performance Test (PPT)	No	Yes	Yes	No	Suggested for assessing severity of impairment of physical function (activities of daily living)	
Six-condition Romberg test	No	Yes	Yes	No	Suggested for assessing severity of balance impairment in HD	
Functional reach test	No	Yes	Yes	No	Suggested *with caveats*	Very limited data by a single group in HD
5 Times Sit to Stand Test	No	Yes	Yes	No	Suggested *with caveats*	Very limited data in a single trial in HD
30 Second Chair Stand	No	Yes	Yes	No	Suggested *with caveats*	Very limited data in a single trial in HD
Dynamic gait index	No	Yes	*Yes*	No	Suggested *with caveats*	Very limited data in a single trial in HD
Walking while talking test	No	Yes	*Yes*	No	Suggested *with caveats*	Very limited data in a single study in HD
Timed 25 Foot Walk Test	No	Yes	No	No	Listed	

12 meter walking, hand tapping in 30s, and time to drink 120 ml	No	Yes	No	No	Listed	
Jebsen-Taylor Hand Function Test	No	Yes	No	No	Listed	
RATING SCALES						
The Unified Huntington's Disease Rating Scale (UHDRS) Total Functional Capacity	Yes	Yes	Yes	No	Suggested for assessing severity of limitation in functional capacity in HD	
UHDRS - Functional Assessment Scale	Yes	Yes	Yes	No	Suggested for assessing severity of limitation in functional capacity in HD	
UHDRS - Independence Scale	Yes	Yes	Yes	No	Suggested for assessing severity of limitation in functional ability in HD	
HD Activities of Daily Living	Yes	Yes	Yes	No	Suggested for assessing severity of limitation in ADLs in HD	
Activity-specific balance scale	No	Yes	Yes	No	Suggested for assessing level of self-reported balance confidence in HD	Questionable use, since lack of insight is a feature in HD
Rivermead Mobility Index	No	Yes	Yes	No	Suggested for assessing severity of mobility restriction (as a generic measure)	
Barthel Index of ADL	No	Yes	*Yes*	No	Suggested *with caveats*	Very limited clinimetric data

Modified Self-Assessment PD Disability Scale	No	Yes	*Yes*	No	Suggested *with caveats*	Very limited clinimetric data
Self-report HD Work function	Yes	Yes	No	No	Listed	
Behavior Observation Scale Huntington - ADL subscale	Yes	Yes	No	No	Listed	
Alzheimer's Disease Cooperative Study Activities of Daily Living Scale	No	Yes	No	No	Listed	
Quick DASH	No	Yes	No	No	Listed	

Table 3. Summary of clinimetric data of all instruments used in HD with a recommendation level of "suggested" or "recommended"

SCALE	INTERNAL CONSISTENCY	TEST-RETEST RELIABILITY	INTER-RATER RELIABILITY	CONSTRUCT VALIDITY	DISCRIMINATION ACROSS DISEASE STAGES/SEVERITY	RESPONSIVENESS	CEILING /FLOOR EFFECT	SENSITIVITY/ SPECIFICITY (E/S)
PERFORMANCE-BASED MEASURES								
Tinetti Mobility test	NR	+	NR	+	+/-	+/- (in non-RCTs)	ceiling and floor effects	+
The Berg Balance Scale	NR	+	NR	+	+ (Stage I *vs.* II/III)	+/- (in non-RCT trials)	ceiling and floor effects	+/- (no E/S data for cut-off for risk of falls)
Six Minute Walk Test	N/A	+	NR	+	+/-	- (data from RCTs)	NR	NR
Timed 'up and go' Test	N/A	+	NR	+/- (no correlation with UHDRS TMS)	+/-	+/- (data from trials)	ceiling and floor effects	+/- (no E/S data for cut-off for risk falls)
Ten Meter Walk Test	N/A	+	NR	+	+/- (non-linear with disease stages)	+/- (in rehabilitation trials)	None	NR
Four square step test	N/A	+	NR	+	- (poor discrimination)	NR	NR	NR
Mini-BESTest	NR	NR	NR	+	+/-	NR	floor effect	NR

Physical Performance Test	NR	+	NR	+	+/- (separates pre/mild vs/ middle/late)	+/-	ceiling effect	NR
Six-condition Romberg test	NR	+	NR	NR	+/- (premanifest vs. manifest)	- (data from 1 trial)	NR	NR
RATING SCALES								
UHDRS - Total Functional Capacity	NR	NR	+/-	+	+	+	ceiling and floor effects	NR
UHDRS - Functional Assessment Scale	+	NR	NR	+	+	+	ceiling effects	NR
UHDRS - Independence Scale	NR	NR	+/-	+	+	+	ceiling effects	NR
HD Activities of Daily Living	+	NR	NR	+/-	+/-	+/- (data from RCTs)	Floor effect	NR
Activity-specific balance scale	NR	+	NR	+/-	NR	+/- (data from 1 trial)	NR	+/- (no E/S data for cut-off for fallers)
Rivermead Mobility Index	NR	+	NR	NR	+/-	NR	Ceiling effect	NR

N/A - not applicable; NR - not reported; UHDRS - The Unified Huntington's Disease Rating Scale, RCTs – randomized controlled trial, HD – Huntington's disease. (+) - good performance, (+/-) contradictory data or very limited data (-) poor performance. NOTE: data regarding Minimally Clinically Important Difference were only assessed for Tinetti Mobility Test.

Supplemental Material. Full description of all clinical measures included for full review, including those that were included in the “suggested with caveats” or “listed” categories

Supplemental table 1: EXCLUDED SCALES:

- **Inadequate construct (n=3)**
  - Rehabilitation Evaluation (Hall and Baker)[1]
  - The Leisure Time Exercise Questionnaire by Godin and Shephard [2]
  - World Health Organization Disability Assessment Schedule (*included in review on Health-related Quality-of-life rating scales*)
- **Precursors of the Unified Huntington's Disease Rating Scale (UHDRS) scales measuring function (n=3)**
  - Physical Disability Scale and the Independence Scale[3]
  - Shoulson-Fahn Disability Scale for HD[4]
  - HD functional capacity scale[3]
- **Not used in HD studies (n=7)**
  - Work limitation questionnaire[5]
  - Scales for Outcomes in Parkinson's Disease–Psychosocial (SCOPA-PS)[6]
  - Parkinson's Problem Schedule (PPS)[7]
  - Global Assessment of Functioning (GAF)[8]
  - Endicott Work Productivity Scale (EWPS)[9]
  - Belastungsfragebogen Parkinson kurzversion (BELA-P-k)[10]
  - The World Health Organization Health and Work Performance Questionnaire (HPQ)[11]
- **Study on device and not on a rating scale or performance measure (n=4)**
  - Posturography using a force plate (FP)[12]
  - GAITRite mat[13]
  - Step Watch Step Activity Monitor (SAM)[14]
  - Sensory Organization Test [15]
- **Incomplete scale (n=1)**
  - Functional Rating Scale Taskforce for pre-Huntington Disease (FuRST-pHD)[16]

Supplemental references 1

1. Baker R, Hall JN. Users manual : for rehabilitation evaluation, Hall and Baker. Aberdeen: Vine, 1983.
2. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. *Can J Appl Sport Sci* 1985;10:141-6.
3. Unified Huntington's disease rating scale: Reliability and consistency. *Movement Disorders* 1996;11.
4. Shoulson I, Fahn S. Huntington disease: clinical care and evaluation. *Neurology* 1979;29:1-3.
5. Lerner D, Amick BC, 3rd, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. *Med Care* 2001;39:72-85.
6. Marinus J, Visser M, Martinez-Martin P, van Hilten JJ, Stiggelbout AM. A short psychosocial questionnaire for patients with Parkinson's disease: the SCOPA-PS. *J Clin Epidemiol* 2003;56:61-7.
7. Brod M, Mendelsohn GA, Roberts B. Patients' experiences of Parkinson's disease. *J Gerontol B Psychol Sci Soc Sci* 1998;53:P213-22.
8. Diagnostic and statistical manual of mental disorders : DSM-IV. 4th ed. ed. Washington, D.C.: American Psychiatric Association, 1994.
9. Endicott J, Nee J. Endicott Work Productivity Scale (EWPS): a new measure to assess treatment effects. *Psychopharmacol Bull* 1997;33:13-6.
10. Spliethoff-Kamminga NG, Zwinderman AH, Springer MP, Roos RA. Psychosocial problems in Parkinson's disease: evaluation of a disease-specific questionnaire. *Mov Disord* 2003;18:503-9.
11. Kessler RC, Barber C, Beck A, Berglund P, Cleary PD, McKenas D*, et al.* The World Health Organization Health and Work Performance Questionnaire (HPQ). *J Occup Environ Med* 2003;45:156-74.
12. Blaszczyk JW. The use of force-plate posturography in the assessment of postural instability. *Gait Posture* 2016;44:1-6.
13. Menz HB, Latt MD, Tiedemann A, Mun San Kwan M, Lord SR. Reliability of the GAITRite walkway system for the quantification of temporo-spatial parameters of gait in young and older people. *Gait Posture* 2004;20:20-5.
14. Bowden MG, Behrman AL. Step Activity Monitor: accuracy and test-retest reliability in persons with incomplete spinal cord injury. *J Rehabil Res Dev* 2007;44:355-62.
15. Cohen H, Heaton LG, Congdon SL, Jenkins HA. Changes in sensory organization test scores with age. *Age Ageing* 1996;25:39-44.
16. Evans K, Anderson K, Borowsky B, Giuliano J, Guttman M, Ho A*, et al.* F08 The functional rating taskforce for pre-huntington's disease: results so far. *Journal of Neurology, Neurosurgery & Psychiatry* 2010;81:A25.2-A25.

PERFORMANCE-BASED MEASURES

Supplemental Table 2

Tinetti mobility test (TMT) ***or Tinetti Performance Oriented Mobility Assessment (POMA)***	
I. Scale description	
Are there several versions of the scale?	Yes. Various versions exist, with variations for both the name of the test and scoring method, e.g., 16-item, 28-point version of the POMA (see http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1039)
If you replied YES, which was been assessed?	The 16-item version.
Scale construct/ overall structure	**Construct:** Balance and gait maneuvers used during daily activities. The TMT was developed as a measure to screen older adults for balance and gait impairments and to be easy to use, reliable and sensitive to significant changes. The TMT includes balance and gait subscales (9 items for balance - POMA-B; 7 items for gait - POMA-G) that measure static and dynamic balance. The test quantitatively ranks gait deviations. Participants are asked to perform a series of functional tasks. The total maximum score is 28 points, where higher scores indicate better performance. The maximum scores is 16 for the gait subscale, and 12 for the balance subscale. (See http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1039)
a. Question items	
Items of presence of symptom/sign?	N/A.
Items of severity of symptom/sign?	NOTE, the scores are made according to descriptors of the performance of tasks, and reflect different degrees of severity (personal judgment). Each item of the TMT is scored using a scale of 0 to 1 or 2; 0 – better performance, 1 or 2- worse performance. (See http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1039)
b. Response scale	

Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Discrete steps (0, 1, 2). (See http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1039)
c. Is the scale easy to score?	
Approx. time to score patient	10-15 minutes. (See http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1039)
d. Raters	
Patient, caregiver, or clinician	Clinician. (See http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1039)
If clinician-rated, is training for application required?	No.
e. Access to scale	
Copyright or public domain?	Public domain.
How can the scale be obtained (address or website)?	http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1039
Has the scale been published in other languages?	Not formally.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	No process for item generation and/or reduction identified.[1]
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Yes, gait and balance items.[1]
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	No, the scale is specific for aspects of balance and gait.[1]
Does it score current state or is it based on the patient/caregiver recall?	Current state.
What is the time frame (e.g. "during the past week")?	N/A.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	Severity of balance impairment, screen for risk of falls.
Is there a cut-off score? (for HD, for non-HD)	Yes.

	HD population: Cut-off for falls (definition: *unintentionally coming to rest on the ground or other surface*): fallers vs. vs. non-fallers: cut-off = 21, with sensitivity of 74% and specificity of 60%.[2]
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	No (personal judgment).
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Applicable for early and mid-stage but not for non-ambulatory patients. Questionable sensitivity in very early or pre-manifest HD (personal judgment).
e. Has this scale been specifically developed for use in HD (yes/no)?	No.[1]
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	N/A.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.
a. Reliability	
Assessed – not assessed	Assessed in HD.
Internal consistency	Not assessed in HD.
Test-retest reliability	**Assessed in HD**: n = 20, TFC Stages 1–3, ICC = 0.83; 95% CI (0.7; 1.0). [3] n = 11 (Pre-manifest HD) and n=64 (Manifest HD) [2, 4] 1. Pre-manifest HD: ICC=0.92 2. Manifest HD: ICC=0.91: • early stage (TFC=11–13) - ICC= 0.98 • middle stage (TFC=7–9) - ICC= 0.96 • late stage (TFC<=6) - ICC= 0.80

Inter-rater reliability	Not assessed in HD.
b. Validity in HD	
Assessed vs. not assessed	Assessed in HD.
Criterion validity (any comparison with gold-standard)	No gold standard available.
Construct validity	
Factor analysis	Not assessed in HD
Convergent validity	1) n=20 (Manifest HD): a. Correlated with spatio-temporal measures of gait and other clinical measures:[3] Pearson correlations: **Spatio-temporal measures:** Forward walking: Velocity, 0.68; Stride length, 0.74; Swing percent, 0.34; Double support percent, –0.54; Base of support, –0.58; CV step time, –0.83; CV stride length, –0.88; CV swing time, –0.82.[3] Backward walking: Velocity, 0.68; Stride length, 0.74; Swing percent, 0.52; Double support percent, –0.64; Base of support, –0.40; CV step time, –0.67; CV stride length, –0.73; CV swing time, –0.41. **Activities-specific Balance Confidence Scale:** r =0.50. **Four Square Step Test:** r = - 0.67. 2) n=11 (Pre-manifest HD) and n=64 (Manifest HD): TMT *vs.* UHDRS-FAS (r=0.44), TMT *vs.* UHDRS-TFC (r=0.42). TMT *vs.* UHDRS-Total Motor Score (TMS): r=-0.59.[5] 3) n=78, manifest HD: lower scores of the TMT correlated with higher scores of the UHDRS-TMS (r= -0.75, p< 0.0001);[2, 3] **Comment:** Correlation signs are correct given direction of measures.
Divergent validity	–
Overall impression: good – not good (based on references preferably, personal judgment can be	–

stated)	
Generalizability	
Shown to be valid at any stage of HD?	Usually not for late stage HD as where many patients are non-ambulatory (personal judgment). Validation studies[2-4] were completed in pre-manifest, manifest HD including late stage HD (worse reliability).[4]
Shown to be valid in any population with dementia or significant cognitive impairment?	No.
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Yes. N=40, manifest HD, non-randomized study, with no control group, of an inpatient rehabilitation program. After a three-week period of treatment there was a significant average improvement in the TMT (4.7, $p<0.001$).[6] N=11, manifest HD, Open label study of on- and off-tetrabenazine, significant average improvement in the TMT ($t = 4.20$, $p = 0.002$).[7] N=18, manifest HD, not shown to detect change after a supervised video game balance training program that improved some spatio-temporal measures of gait.[8]
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No. NOTE: Minimum Detectable Change (MDC) has been calculated ranging from 1 (premanifest) to 5 (late stage, TFC≤6).[3, 4]
Floor and ceiling effects	Floor effects (not suitable for administration in non-ambulatory individuals; Ceiling effects, not sensitive to differences in pre-manifest and healthy controls. (personal judgment)
Score distributions	n=11 (Pre-manifest HD) and n=64 (Manifest HD), mean (SD):[2, 4] o Pre-manifest HD: 28 (0.7) o Manifest HD: 22 (5) • early stage (TFC=11–13) - 24 (5) • middle stage (TFC=7–9) - 22 (4)

	• late stage (TFC<=6) - 21 (4) Fallers (n = 34) vs. Non-fallers (n = 60): 17.24 +/- 5.61 *vs.* 21.37 +/- 4.85. [2]
IV. Overall impression	
Advantages	Quick and easy to administer, includes balance and gait sections, fairly well validated in HD, requires some subjective rating of movement performance.
Disadvantages	Questionable sensitivity to detect change due to low quality of study design.

V. Recommendation	**Recommended for screening for risk of falls.** **Recommended for assessment of gait and balance problems in patients with manifest HD, up to HD stage III.**

Supplemental references 2

1. Tinetti ME, Williams TF, Mayewski R. Fall risk index for elderly patients based on number of chronic disabilities. *Am J Med* 1986;80:429-34.
2. Kloos AD, Kegelmeyer DA, Young GS, Kostyk SK. Fall risk assessment using the Tinetti mobility test in individuals with Huntington's disease. *Mov Disord* 2010;25:2838-44.
3. Kloos AD, Fritz NE, Kostyk SK, Young GS, Kegelmeyer DA. Clinimetric properties of the Tinetti Mobility Test, Four Square Step Test, Activities-specific Balance Confidence Scale, and spatiotemporal gait measures in individuals with Huntington's disease. *Gait Posture* 2014;40:647-51.
4. Quinn L, Khalil H, Dawes H, Fritz NE, Kegelmeyer D, Kloos AD*, et al.* Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. *Phys Ther* 2013;93:942-56.
5. Busse M, Quinn L, Khalil H, McEwan K. Optimising mobility outcome measures in Huntington's disease. *J Huntingtons Dis* 2014;3:175-88.
6. Zinzi P, Salmaso D, De Grandis R, Graziani G, Maceroni S, Bentivoglio A*, et al.* Effects of an intensive rehabilitation programme on patients with Huntington's disease: a pilot study. *Clin Rehabil* 2007;21:603-13.
7. Kegelmeyer DA, Kloos AD, Fritz NE, Fiumedora MM, White SE, Kostyk SK. Impact of tetrabenazine on gait and functional mobility in individuals with Huntington's disease. *J Neurol Sci* 2014;347:219-23.

8. Kloos AD, Fritz NE, Kostyk SK, Young GS, Kegelmeyer DA. Video game play (Dance Dance Revolution) as a potential exercise therapy in Huntington's disease: a controlled clinical trial. *Clinical rehabilitation* 2013;27:972-82.

TINETTI BALANCE ASSESSMENT TOOL

Tinetti ME, Williams TF, Mayewski R, Fall Risk Index for elderly patients based on number of chronic disabilities. Am J Med 1986:80:429-434

PATIENTS NAME _____ D.o.b. _____ Ward _____

BALANCE SECTION

Patient is seated in hard, armless chair;

		Date		
Sitting Balance	Leans or slides in chair Steady, safe	= 0 = 1		
Rises from chair	Unable to without help Able, uses arms to help Able without use of arms	= 0 = 1 = 2		
Attempts to rise	Unable to without help Able, requires > 1 attempt Able to rise, 1 attempt	= 0 = 1 = 2		
Immediate standing Balance (first 5 seconds)	Unsteady (staggers, moves feet, trunk sway) Steady but uses walker or other support Steady without walker or other support	= 0 = 1 = 2		
Standing balance	Unsteady Steady but wide stance and uses support Narrow stance without support	= 0 = 1 = 2		
Nudged	Begins to fall Staggers, grabs, catches self Steady	= 0 = 1 = 2		
Eyes closed	Unsteady Steady	= 0 = 1		
Turning 360 degrees	Discontinuous steps Continuous	= 0 = 1		
	Unsteady (grabs, staggers) Steady	= 0 = 1		
Sitting down	Unsafe (misjudged distance, falls into chair) Uses arms or not a smooth motion Safe, smooth motion	= 0 = 1 = 2		
	Balance score		/16	/16

P.T.O.

TINETTI BALANCE ASSESSMENT TOOL

GAIT SECTION

Patient stands with therapist, walks across room (+/- aids), first at usual pace, then at rapid pace.

		Date		
Indication of gait (Immediately after told to 'go'.)	Any hesitancy or multiple attempts No hesitancy	= 0 = 1		
Step length and height	Step to Step through R Step through L	= 0 = 1 = 1		
Foot clearance	Foot drop L foot clears floor R foot clears floor	= 0 = 1 = 1		
Step symmetry	Right and left step length not equal Right and left step length appear equal	= 0 = 1		
Step continuity	Stopping or discontinuity between steps Steps appear continuous	= 0 = 1		
Path	Marked deviation Mild/moderate deviation or uses w. aid Straight without w. aid	= 0 = 1 = 2		
Trunk	Marked sway or uses w. aid No sway but flex. knees or back or uses arms for stability No sway, flex., use of arms or w. aid	= 0 = 1 = 2		
Walking time	Heels apart Heels almost touching while walking	= 0 = 1		
	Gait score		/12	/12
	Balance score carried forward		/16	/16
	Total Score = Balance + Gait score		/28	/28

Risk Indicators:

Tinetti Tool Score	Risk of Falls
≤18	High
19-23	Moderate
≥24	Low

Supplemental table 3

Berg Balance Scale (BBS)	
I. Scale description	
Are there several versions of the scale?	No.
If you replied YES, which was been assessed?	Not applicable.
Scale construct/ overall structure	The BBS was designed to assess static balance and fall risk in adult populations. The BBS entails 14 subtests of various activities related to balance control. Subtests include static postures (e.g., sitting, standing), transitions (e.g., sitting to standing, transferring between chairs), and challenging positions (e.g., standing with eyes closed). Quality of performance for each item is scored using a 4-point scale, with higher scores indicating better balance. The maximum possible score is 56. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=888
a. Question items	
Items of presence of symptom/sign?	N/A
Items of severity of symptom/sign?	N/A
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Discrete steps, 4-point scale.
c. Is the scale easy to score?	
Approx. time to score patient	10-15 minutes (personal judgment). 15-20 minutes *in* http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=888
d. Raters	
Patient, caregiver, or clinician	Clinician.

	http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=888
If clinician-rated, is training for application required?	Not required. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=888
e. Access to scale	
Copyright or public domain?	Public domain. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=888
How can the scale be obtained (address or website)?	Multiple sources. For example: www.aahf.info/pdf/Berg_Balance_Scale.pdf
Has the scale been published in other languages?	No.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	–
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Yes, the scale covers items across the balance domain.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	No (personal judgment).
Does it score current state or is it based on the patient/caregiver recall?	Current state.
What is the time frame (e.g. "during the past week")?	Current state.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	Severity.
Is there a cut-off score? (for HD, for non-HD)	Yes. For cut-off of BBS≤40 / Predicted probability for being a "faller" was virtually 60%. "Fallers" ≥1 fall in the previous 12 months.[1]
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	Generally, no. Some of the qualifiers for the ratings (safely, easily) can be considered

	inherently subjective (personal judgment).
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Not applicable for non-ambulatory HD (personal judgment).
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	N/A
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.
a. Reliability	
Assessed – not assessed	Assessed in HD.[2]
Internal consistency	-
Test-retest reliability	n = 11 (Pre-manifest HD) and n=64 (Manifest HD) [2] 1. Pre-manifest HD: ICC=0.86 2. Manifest HD: ICC=0.96: • early stage (TFC=11–13) - ICC= 0.90. • middle stage (TFC=7–9) - ICC= 0.91. • late stage (TFC<=6) - ICC= 0.97.
Inter-rater reliability	-
b. Validity in HD	
Assessed vs. not assessed	Assessed in HD.[1-3]
Overall impression: good – not good	-
Criterion validity (any comparison with gold-standard)	
Construct validity	
Factor analysis	Not assessed.

Convergent validity	1) n=64 (Manifest HD), BBS *vs.* UHDRS-TFC (r = 0.60, p<0.01).[3]The BBS and the "Timed UP and GO" have been reported to have high correlations between them. [4] 2) n=11 (Pre-manifest HD) and n=64 (Manifest HD): BBS *vs.* UHDRS-FAS (r=0.48), BBS *vs.* UHDRS-TFC (r=0.43). BBS *vs.* UHDRS-Total Motor Score (TMS): r=-0.55. [5] 1) n=64 (Manifest HD): BBS *vs.* measures of quantitative gait: falls (r=-0.48, p<0.01) and fall risk (coefficients of variation for stride length (n.s), step time (r=-0.47, p<0.05), various balance measures (n.s).[3] **Comment:** Correlation signs are correct given direction of measures.
Divergent validity	1) n=64 (Manifest HD): BBS *vs.* HD-ADL (r = - 0.48, p<0.01).
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Good test-retest reliability. Reasonable construct validity testing.
Generalizability	
Shown to be valid at any stage of HD?	Not suitable for non-ambulatory HD (personal judgment).
Shown to be valid in any population with dementia or significant cognitive impairment?	No (personal judgment).
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Withdrawal of tetrabenazine resulted in significant reduction of BBS scores in a manifest HD cohort (n=10), with no change in cognitive or behavioral measures.[6] Manifest HD, mean group change in response to a 1-year rehabilitation multidisciplinary Program intervention = +1.0 (p<0.03).[7] Randomized Trial of structured home-based exercise *vs.* usual care: early to moderate HD with walking or balance difficulties, n=25: Mean difference: 5.4 (95% CI: 1.0, 9.9,

	p=0.01). [8] Manifest HD with chorea, n=11. [9] Non-significant change in response to tetrabenazine: mean (SD) off: 48.8 ± 6 and on drug 49.8 ± 7.5 (n.s.). Same trial n.s. change for Timed 25 Foot Walk Test.
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No. NOTE: MDC = 5.[2]
Floor and ceiling effects	Susceptible to ceiling effect[2] and floor effect (not able to administer in non-ambulatory individuals) (personal judgment).
Score distributions	n = 11 (Pre-manifest HD) and n=64 (Manifest HD) Mean (SD): [2] 3. Pre-manifest HD: 55 (1) 4. Manifest HD: 47 (8) • early stage (TFC=11–13) - 51 (4). • middle stage (TFC=7–9) - 47 (6). • late stage (TFC<=6) - 45 (12).
IV. Overall impression	
Advantages	–
Disadvantages	–
V. Recommendation	**Recommended for assessing severity of balance impairment in HD with preserved ambulation** **Suggested for screening for risk of falls.**

Supplemental references 3

1. Busse ME, Wiles CM, Rosser AE. Mobility and falls in people with Huntington's disease. *J Neurol Neurosurg Psychiatry* 2009;80:88-90.
2. Quinn L, Khalil H, Dawes H, Fritz NE, Kegelmeyer D, Kloos AD*, et al.* Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. *Phys Ther* 2013;93:942-56.
3. Rao AK, Muratori L, Louis ED, Moskowitz CB, Marder KS. Clinical measurement of mobility and balance impairments in Huntington's disease: validity and responsiveness. *Gait Posture* 2009;29:433-6.

4. Busse ME, Khalil H, Quinn L, Rosser AE. Physical therapy intervention for people with Huntington disease. *Phys Ther* 2008;88:820-31.
5. Busse M, Quinn L, Khalil H, McEwan K. Optimising mobility outcome measures in Huntington's disease. *J Huntingtons Dis* 2014;3:175-88.
6. Fekete R, Davidson A, Jankovic J. Clinical assessment of the effect of tetrabenazine on functional scales in huntington disease: a pilot open label study. *Tremor Other Hyperkinet Mov (N Y)* 2012;2.
7. Piira A, van Walsem MR, Mikalsen G, Nilsen KH, Knutsen S, Frich JC. Effects of a One Year Intensive Multidisciplinary Rehabilitation Program for Patients with Huntington's Disease: a Prospective Intervention Study. *PLoS Curr* 2013;5.
8. Khalil H, Quinn L, van Deursen R, Dawes H, Playle R, Rosser A*, et al.* What effect does a structured home-based exercise programme have on people with Huntington's disease? A randomized, controlled pilot study. *Clinical rehabilitation* 2013;27:646-58.
9. Ferrara JM, Mostile G, Hunter C, Adam OR, Jankovic J. Effect of tetrabenazine on motor function in patients with huntington disease. *Neurol Ther* 2012;1:5.

Berg Balance Scale

The Berg Balance Scale (BBS) was developed to measure balance among older people with impairment in balance function by assessing the performance of functional tasks. It is a valid instrument used for evaluation of the effectiveness of interventions and for quantitative descriptions of function in clinical practice and research. The BBS has been evaluated in several reliability studies. *A recent study of the BBS, which was completed in Finland, indicates that a change of eight (8) BBS points is required to reveal a genuine change in function between two assessments among older people who are dependent in ADL and living in residential care facilities.*

Description:
14-item scale designed to measure balance of the older adult in a clinical setting.

Equipment needed: Ruler, two standard chairs (one with arm rests, one without), footstool or step, stopwatch or wristwatch, 15 ft walkway

Completion:
Time: 15-20 minutes
Scoring: A five-point scale, ranging from 0-4. "0" indicates the lowest level of function and "4" the highest level of function. Total Score = 56

Interpretation: 41-56 = low fall risk
21-40 = medium fall risk
0 –20 = high fall risk

A change of 8 points is required to reveal a genuine change in function between 2 assessments.

Berg Balance Scale

Name: _____ Date: _____

Location: _____ Rater: _____

ITEM DESCRIPTION	SCORE (0-4)
Sitting to standing	_____
Standing unsupported	_____
Sitting unsupported	_____
Standing to sitting	_____
Transfers	_____
Standing with eyes closed	_____
Standing with feet together	_____
Reaching forward with outstretched arm	_____
Retrieving object from floor	_____
Turning to look behind	_____
Turning 360 degrees	_____
Placing alternate foot on stool	_____
Standing with one foot in front	_____
Standing on one foot	_____

Total _____

GENERAL INSTRUCTIONS
Please document each task and/or give instructions as written. When scoring, please record the lowest response category that applies for each item.

In most items, the subject is asked to maintain a given position for a specific time. Progressively more points are deducted if:

- the time or distance requirements are not met
- the subject's performance warrants supervision
- the subject touches an external support or receives assistance from the examiner

Subject should understand that they must maintain their balance while attempting the tasks. The choices of which leg to stand on or how far to reach are left to the subject. Poor judgment will adversely influence the performance and the scoring.

Equipment required for testing is a stopwatch or watch with a second hand, and a ruler or other indicator of 2, 5, and 10 inches. Chairs used during testing should be a reasonable height. Either a step or a stool of average step height may be used for item # 12.

Berg Balance Scale

SITTING TO STANDING
INSTRUCTIONS: Please stand up. Try not to use your hand for support.
( ) 4 able to stand without using hands and stabilize independently
( ) 3 able to stand independently using hands
( ) 2 able to stand using hands after several tries
( ) 1 needs minimal aid to stand or stabilize
( ) 0 needs moderate or maximal assist to stand

STANDING UNSUPPORTED
INSTRUCTIONS: Please stand for two minutes without holding on.
( ) 4 able to stand safely for 2 minutes
( ) 3 able to stand 2 minutes with supervision
( ) 2 able to stand 30 seconds unsupported
( ) 1 needs several tries to stand 30 seconds unsupported
( ) 0 unable to stand 30 seconds unsupported

If a subject is able to stand 2 minutes unsupported, score full points for sitting unsupported. Proceed to item #4.

SITTING WITH BACK UNSUPPORTED BUT FEET SUPPORTED ON FLOOR OR ON A STOOL
INSTRUCTIONS: Please sit with arms folded for 2 minutes.
( ) 4 able to sit safely and securely for 2 minutes
( ) 3 able to sit 2 minutes under supervision
( ) 2 able to able to sit 30 seconds
( ) 1 able to sit 10 seconds
( ) 0 unable to sit without support 10 seconds

STANDING TO SITTING
INSTRUCTIONS: Please sit down.
( ) 4 sits safely with minimal use of hands
( ) 3 controls descent by using hands
( ) 2 uses back of legs against chair to control descent
( ) 1 sits independently but has uncontrolled descent
( ) 0 needs assist to sit

TRANSFERS
INSTRUCTIONS: Arrange chair(s) for pivot transfer. Ask subject to transfer one way toward a seat with armrests and one way toward a seat without armrests. You may use two chairs (one with and one without armrests) or a bed and a chair.
( ) 4 able to transfer safely with minor use of hands
( ) 3 able to transfer safely definite need of hands
( ) 2 able to transfer with verbal cuing and/or supervision
( ) 1 needs one person to assist
( ) 0 needs two people to assist or supervise to be safe

STANDING UNSUPPORTED WITH EYES CLOSED
INSTRUCTIONS: Please close your eyes and stand still for 10 seconds.
( ) 4 able to stand 10 seconds safely
( ) 3 able to stand 10 seconds with supervision
( ) 2 able to stand 3 seconds
( ) 1 unable to keep eyes closed 3 seconds but stays safely
( ) 0 needs help to keep from falling

STANDING UNSUPPORTED WITH FEET TOGETHER
INSTRUCTIONS: Place your feet together and stand without holding on.
( ) 4 able to place feet together independently and stand 1 minute safely
( ) 3 able to place feet together independently and stand 1 minute with supervision
( ) 2 able to place feet together independently but unable to hold for 30 seconds
( ) 1 needs help to attain position but able to stand 15 seconds feet together
( ) 0 needs help to attain position and unable to hold for 15 seconds

Berg Balance Scale continued…

REACHING FORWARD WITH OUTSTRETCHED ARM WHILE STANDING
INSTRUCTIONS: Lift arm to 90 degrees. Stretch out your fingers and reach forward as far as you can. (Examiner places a ruler at the end of fingertips when arm is at 90 degrees. Fingers should not touch the ruler while reaching forward. The recorded measure is the distance forward that the fingers reach while the subject is in the most forward lean position. When possible, ask subject to use both arms when reaching to avoid rotation of the trunk.)
( ) 4 can reach forward confidently 25 cm (10 inches)
( ) 3 can reach forward 12 cm (5 inches)
( ) 2 can reach forward 5 cm (2 inches)
( ) 1 reaches forward but needs supervision
( ) 0 loses balance while trying/requires external support

PICK UP OBJECT FROM THE FLOOR FROM A STANDING POSITION
INSTRUCTIONS: Pick up the shoe/slipper, which is in front of your feet.
( ) 4 able to pick up slipper safely and easily
( ) 3 able to pick up slipper but needs supervision
( ) 2 unable to pick up but reaches 2-5 cm(1-2 inches) from slipper and keeps balance independently
( ) 1 unable to pick up and needs supervision while trying
( ) 0 unable to try/needs assist to keep from losing balance or falling

TURNING TO LOOK BEHIND OVER LEFT AND RIGHT SHOULDERS WHILE STANDING
INSTRUCTIONS: Turn to look directly behind you over toward the left shoulder. Repeat to the right. (Examiner may pick an object to look at directly behind the subject to encourage a better twist turn.)
( ) 4 looks behind from both sides and weight shifts well
( ) 3 looks behind one side only other side shows less weight shift
( ) 2 turns sideways only but maintains balance
( ) 1 needs supervision when turning
( ) 0 needs assist to keep from losing balance or falling

TURN 360 DEGREES
INSTRUCTIONS: Turn completely around in a full circle. Pause. Then turn a full circle in the other direction.
( ) 4 able to turn 360 degrees safely in 4 seconds or less
( ) 3 able to turn 360 degrees safely one side only 4 seconds or less
( ) 2 able to turn 360 degrees safely but slowly
( ) 1 needs close supervision or verbal cuing
( ) 0 needs assistance while turning

PLACE ALTERNATE FOOT ON STEP OR STOOL WHILE STANDING UNSUPPORTED
INSTRUCTIONS: Place each foot alternately on the step/stool. Continue until each foot has touched the step/stool four times.
( ) 4 able to stand independently and safely and complete 8 steps in 20 seconds
( ) 3 able to stand independently and complete 8 steps in > 20 seconds
( ) 2 able to complete 4 steps without aid with supervision
( ) 1 able to complete > 2 steps needs minimal assist
( ) 0 needs assistance to keep from falling/unable to try

STANDING UNSUPPORTED ONE FOOT IN FRONT
INSTRUCTIONS: (DEMONSTRATE TO SUBJECT) Place one foot directly in front of the other. If you feel that you cannot place your foot directly in front, try to step far enough ahead that the heel of your forward foot is ahead of the toes of the other foot. (To score 3 points, the length of the step should exceed the length of the other foot and the width of the stance should approximate the subject's normal stride width.)
( ) 4 able to place foot tandem independently and hold 30 seconds
( ) 3 able to place foot ahead independently and hold 30 seconds
( ) 2 able to take small step independently and hold 30 seconds
( ) 1 needs help to step but can hold 15 seconds
( ) 0 loses balance while stepping or standing

STANDING ON ONE LEG
INSTRUCTIONS: Stand on one leg as long as you can without holding on.
( ) 4 able to lift leg independently and hold > 10 seconds
( ) 3 able to lift leg independently and hold 5-10 seconds
( ) 2 able to lift leg independently and hold ≥ 3 seconds
( ) 1 tries to lift leg unable to hold 3 seconds but remains standing independently.
( ) 0 unable to try of needs assist to prevent fall

( ) TOTAL SCORE (Maximum = 56)

Supplemental table 4

6-Minute Walk Test (6MWT)	
I. Scale description	
Are there several versions of the scale?	The original test consisted of a 12-minute walk that was shortened to 6 minutes.[1, 2] A shortened 2-minute walk version has been tested in geriatric and other populations. Variations in walking distance impact responsiveness so changing from 6 minutes is not recommended (personal judgment).
If you replied YES, which was been assessed?	The 6-minute walk test.
Scale construct/ overall structure	The test measures how many meters an individual is able to walk in 1, 3, and 6 minutes. It was originally developed as a measure of pulmonary function in athletes, then in patients with various diseases in respiratory diseases like COPD and related disorders that affect oxygen consumption, but has been applied as a test of walking distance/"endurance"; others have applied the test in various populations without performing the testing needed for standardization. In the context of neurological conditions, it is used as a measure of endurance (i.e., different from shorter walking tests that measure velocity of walking speed).[3]
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	Severity.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Severity is measured in continuous values of meters.
c. Is the scale easy to score?	
Approx. time to score patient	The standardized approach includes two practice tests, lasting up to 30 minutes. The practice test is not always used, which is problematic for consistency of the obtained results.[4, 5]
d. Raters	
Patient, caregiver, or clinician	Clinician.
If clinician-rated, is training for application	No, but it is necessary to follow and specify a standardized protocol.

required?	
e. Access to scale	
Copyright or public domain?	Public domain.
How can the scale be obtained (address or website)?	For example: http://www.cscc.unc.edu/spir/public/UNLICOMMSMWSixMinuteWalkTestFormQxQ08252011.pdf
Has the scale been published in other languages?	Not applicable.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	No.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	No. The items are not intended to do so.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	The distance walked in a specified time.
Does it score current state or is it based on the patient/caregiver recall?	Current state.
What is the time frame (e.g. "during the past week")?	Current time.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To measure walking endurance (severity).
Is there a cut-off score? (for HD, for non-HD)	**Not for HD.** Normative data - needs to be age-matched as distances reduce with age. Summary in *http://www.Rehabmeasures.org*
d. Acceptability	
Is the length of the scale appropriate?	It is time consuming compared to other walking tests, but it measures endurance, while other tests measure speed/velocity (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No.
Are there ambiguities in rating anchors?	Not applicable.

Are the questions appropriate for use in an HD population?	Not applicable.
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Used in pre-manifest and across mild to severe manifest HD. [3, 6] It cannot be used with those who need physical assistance to walk (personal judgment). **NOTE:** If people cannot walk six minutes they can rest and resume, and the distance walked in the total six minutes (including the rest time) is recorded. If they cannot resume, the distance walked is recorded.
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Not applicable.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.
a. Reliability	
Assessed – not assessed	Assessed in HD.[3]
Internal consistency	Not applicable.
Test-retest reliability	n = 11 (Pre-manifest HD) and n=64 (Manifest HD) [3] 1. Pre-manifest HD: ICC=0.98. 2. Manifest HD: ICC=0.94: • early stage (TFC=11–13) - ICC= 0.97. • middle stage (TFC=7–9) - ICC= 0.86. • late stage (TFC<=6) - ICC= 0.97.
Inter-rater reliability	Not assessed in HD.
b. Validity in HD	
Assessed vs. not assessed	Not assessed in HD.
Criterion validity (any comparison with gold-standard)	
Construct validity	
Factor analysis	Not assessed in HD.
Convergent validity	Manifest HD: [6]

	6MWT *vs.* UHDRS-TMS (n=62): r= -0.25 (at 6 minutes) but n.s. at 1 and 3 minutes; 6MWT *vs.* UHDRS Functional Assessment (n=61): r= 0.37, 0.38 and 0.41 at 1, 3, 6 minutes (all significant); 6MWT *vs.* UHDRS-TFC (n=62): r = 0.25, 0.29 and 0.29 at 1, 3, 6 minutes (all significant). **Personal Comment:** Correlation signs are correct given direction of measures.
Divergent validity	Not assessed in HD.
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	–
Generalizability	
Shown to be valid at any stage of HD?	Limited use for non-ambulatory subjects (personal judgment).
Shown to be valid in any population with dementia or significant cognitive impairment?	No.
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Manifest HD, n=37. Change in response to a 1-year rehabilitation multidisciplinary program intervention = 68.71 meters (group).[7] **Personal Comment:** less than MDC_{95} of Quinn 2013[3] (individual change needed), so we conclude that it was not meaningful. Manifest HD, n=31. A 12-week Community-Based Exercise Program vs. usual care RCT. Treatment effect estimate: [95 % CI]: 27.2 [-2.8 to 57.2], p=0.08 [8]
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No. **NOTE:** MDC_{95} (formula based on SEM): pre-manifest HD, 39.22 meters; Manifest HD, overall 86.57 meters; early HD 56.60, mid HD 126.14 and late HD 70.65 meters. Only 6-minute results seem to be reported, and only one walk was conducted.
Floor and ceiling effects	Not assessed in HD.
Score distributions	It is unclear how values discriminate among pre- and manifest HD severity levels according to confidence intervals (CI) on means overlap; it seems that values may separate pre- and mild from mid/severe levels based on CIs.[3]

	Moderate and severe HD also have similar values. (i.e., two groups).
IV. Overall impression	
Advantages	It is reasonably short although other walking tests are much shorter. High reliability.
Disadvantages	It has not been used in HD according to the original protocol, which would make it longer. There is a conceptual issue due to the original intent being to assess pulmonary function and not a measure of walking distance.
V. Recommendation	**Recommended for assessing walking endurance (severity) in HD with preserved ambulation**

Supplemental references 4

1. Balke B. A Simple Field Test for the Assessment of Physical Fitness. Rep 63-6. *Rep Civ Aeromed Res Inst US* 1963:1-8.
2. Butland RJ, Pang J, Gross ER, Woodcock AA, Geddes DM. Two-, six-, and 12-minute walking tests in respiratory disease. *Br Med J (Clin Res Ed)* 1982;284:1607-8.
3. Quinn L, Khalil H, Dawes H, Fritz NE, Kegelmeyer D, Kloos AD*, et al.* Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. *Phys Ther* 2013;93:942-56.
4. Guyatt GH, Thompson PJ, Berman LB, Sullivan MJ, Townsend M, Jones NL*, et al.* How should we measure function in patients with chronic heart and lung disease? *J Chronic Dis* 1985;38:517-24.
5. Guyatt GH, Pugsley SO, Sullivan MJ, Thompson PJ, Berman L, Jones NL*, et al.* Effect of encouragement on walking test performance. *Thorax* 1984;39:818-22.
6. Busse M, Quinn L, Khalil H, McEwan K. Optimising mobility outcome measures in Huntington's disease. *J Huntingtons Dis* 2014;3:175-88.
7. Piira A, van Walsem MR, Mikalsen G, Nilsen KH, Knutsen S, Frich JC. Effects of a One Year Intensive Multidisciplinary Rehabilitation Program for Patients with Huntington's Disease: a Prospective Intervention Study. *PLoS Curr* 2013;5.
8. Busse M, Quinn L, Debono K, Jones K, Collett J, Playle R*, et al.* A randomized feasibility study of a 12-week community-based exercise program for people with Huntington's disease. *J Neurol Phys Ther* 2013;37:149-58.

SPIROMICS

INSTRUCTIONS FOR SIX MINUTE WALK TEST FORM
SMW, VERSION 1.0 (QxQ)

I. GENERAL INSTRUCTIONS

The Six Minute Walk Test Form is filled out by the study clinician conducting the test. Using a paper copy of the form to record the data while the test is in progress is recommended.

The Six Minute Walk, an assessment of lung function is the Flexible Block A procedure. Usually the walk should follow shortly after spirometry since it is performed after bronchodilation (for participants with COPD and/or asthma).

The testing area must be a 30m (100 ft.) segment of straight, unimpeded hallway.

Prepare the area by applying markers for the endpoints and 3m intervals to the baseboard on one side of the hall, with special attention to avoid doorways, etc.

Use the provided 30m metric tape measure. If a pre-existing 100 ft. (30.48m) course with 10ft. markers has been previously laid out, it may be used.

If available, place the traffic cones at the center of the proximal and distal turn points. Place the turn signs at the proximal and distal turn points of the course.

Have ready the following materials: *stopwatch/timer, worksheet for counting laps, oximeter, Borg breathlessness and exertion scales, a chair that can be easily moved along the walking course, emergency equipment (according to local policy): telephone, sphygmomanometer, oxygen source.*

A "warm-up" period before the test should not be performed.
Participants should use their usual walking aids during the test (cane, walker, etc.) and be dressed in comfortable clothing and walking shoes.

In general, it is preferable to use room air. If the participant is on long-term oxygen therapy with a resting saturation off oxygen of less than 88%, supplemental oxygen may be used during the test. Future yearly tests should be done at the same amount of supplemental oxygen if possible.

The University of Utah will use 1.5L/min by continuous nasal canula for all subjects to simulate sea level inspired pO_2 unless the participant is receiving a high flow rate of long-term oxygen therapy ***and*** *desaturates to less than 88% on 1.5L/min at rest (see above). All other sites should use room air as noted above.*
See the SPIROMICS MOP 2, Section 2.14 for further details on oxygen use.

Prior to the test, the participant should sit in a chair, located near the starting position for at least 10 minutes before assessing pulse and SpO2 (and Blood Pressure if not taken and recorded within 4 hours prior to test).

If systolic BP is > 200mmHg or < 60mmHg, or diastolic blood pressure > 110mmHg discontinue the test.

If resting heart rate is > 120 or < 50 beats per minute discontinue the test.

If resting SpO2 is < 88% the participant is not eligible to continue the test *(exception noted above for participants on long-term oxygen therapy).*
Reasons for immediately stopping the test include:

- *if SpO2 falls below 80%*
- *the participant asks to stop the test*
- *if the participant experiences chest pain*
- *intolerable dyspnea*
- *leg cramps*
- *staggering*
- *diaphoresis*
- *pale or ashen appearance*

II. DETAILED INSTRUCTIONS FOR CHALLENGE

Explain the use of the modified Borg scale (0-10) for assessing breathlessness.

Explain the use of the Borg rating of perceived exertion scale (6-20) for rating perceived exertion.

Read the following instructions to the participant:

"The object of this test is to walk as far as possible for 6 minutes. You will walk back and forth in this hallway. Six minutes is a long time to walk, so you will be exerting yourself. You will probably get out of breath or become exhausted. You are permitted to slow down, to stop, and to rest as necessary. You may lean against the wall while resting, but resume walking as soon as you are able. You will be walking back and forth around the cones. You should pivot briskly around the cones and continue back the other way without hesitation. Now I'm going to show you. Please watch the way I turn without hesitation."

Demonstrate by walking one lap yourself. Walk and pivot around a cone briskly.

Record completed and partial laps on the lap count worksheet.

Say to the participant:

"Are you ready to do that? I am going to use this counter to keep track of the number of laps you complete. I will click it each time you turn around at this starting line. Remember that the object is to walk AS FAR AS POSSIBLE for 6 minutes, but don't run or jog. Start now, or whenever you are ready."

Standardized Encouragement read in a steady voice:

After the 1st minute: *"You are doing well. You have 5 minutes to go."*

When the timer shows 4 minutes remaining: *"Keep up the good work. You have 4 minutes to go."*

When the timer shows 3 minutes remaining: *"You are doing well. You are halfway done.*

When the timer shows 2 minutes remaining: *"Keep up the good work. You have only 2 minutes left.*

When the timer shows 1 minute remaining: *"You are doing well. You only have 1 minute to go.*

With 15 seconds to go: *"In a moment I'm going to tell you to stop. When I do, just stop right where you are and I will come to you."*

At 6 minutes: *"Stop"*

If the participant stops at any time prior, you can say: *"You can lean against the wall if you would like; then continue walking whenever you feel able."*

Do not use other words of encouragement (or body language) to influence the patient's walking speed. Accompany the participant along the walking course, but keep just behind them. Do not lead them.

If available record the distance at which the oxygen saturation drops < 88%.

III GENERAL INSTRUCTIONS FOR THE FORM

Header Information: The header information consists of key fields which uniquely identify each recorded instance of a form.

FORM DATE: Record date this is being completed. Select the date from the pop up calendar or type in the date in the space provided. Dates should be entered in the mm/dd/yyyy format.

INITIALS: Record the staff code of the person entering the data on this form. This code is assigned to each person at each site by the GIC. If you do not have a staff code and are collecting SPIROMICS data please contact the GIC in order to receive your own individual staff code.

III. DETAILED INSTRUCTIONS FOR EACH ITEM

Item 1. Medications taken **since** post-bronchodilator spirometry: Record 'Y' for Yes or 'N' for No. If No, go to Item 2. If Yes, complete 1a-c.

Item1a-c. Record medication name, dose and time taken for up to 3 medications. Record time in hours and minutes. Choose AM or PM.

Item 2. Blood pressure more than 4 hours prior to 6MW: Record 'Y' for Yes or 'N' for No. If No, go to Item 3. If Yes, complete Item2a-b.

Item2a. Record systolic pressure

Item2b. Record diastolic pressure

Item 3. Supplemental Oxygen **during** test : Record 'Y' for Yes or 'N' for No. If No, go to Item 4. If Yes, complete Item 3a-b.

Item3a. Oxygen Flow rate: Record in Liters per minute.

Item3b. Oxygen type: Record 1 for continuous flow nasal canula or 2 for Pulsed delivery system (conserver).

Item4a. SpO2 at rest prior to 6MW: Record as percentage.

Item4b. Pulse: Record beats per minute.

Item 5. Continuous oximetry recorded: Record 'Y' for Yes, or 'N' for No.

Item 6. Start of 6-minute walk: Record time in hours and minutes. Choose AM or PM.

Item 7. Immediately following 6MW: Record the following:

Item7a. SpO2: Record as percentage.

Item7b. Pulse: Record beats per minute.

Item7c. Breathlessness: Record participant's response from 0-10 on the Modified Borg Scale *(0=no breathlessness, nothing at all, 0.5=very, very slight, 1=very slight, 2=slight breathlessness, 3=moderate, 4=somewhat severe, 5=severe breathlessness,6=is between severe breathlessness and very severe breathlessness, 7=very severe breathlessness, 8=between very severe breathlessness and very,very severe breathlessness, 9=very, very severe breathlessness, 10=maximum breathlessness.)*

Item7d. Exertion: Record participant's response from 6-20 on the Borg Scale of Perceived Exertion *(6=none, 7-8=very,very light, 9-10=very light, 11-12=fairly light, 13-14=somewhat hard, 15-16=hard, 17-18=very hard, 19-20=very,very hard.*

Item 8a. Type of course used: Select the type of course used. Record 1 for 30 meters x 2 lengths, 2 for 100 feet x 2 lengths, or 3 for other. If Other, specify in the space provided.

Item 8b. Record the number of completed laps

Item 8c. Record the distance walked the final partial lap in meters if 8a is in meters or in feet if 8b is in feet.

Item 9. Stopped before 6 minutes: Record Y for Yes or N for No. If No skip out of form. If Yes answer 9a and 10.

Item9a. Duration: Record in minutes and seconds.

Item 10. Reason for stopping: Record one response 1-5. *(1=desaturation <80%, 2=foot, knee, hip or other orthopedic pain, 3=muscle fatigue or pain, 4=breathlessness, 5=adverse event)*

Item105. If response to Item 10=5, select all that apply. *(a=angina, b=lightheadedness, c=intolerable dyspnea, d=leg cramps, e=staggering, f=diaphoresis, g=pale or ashen appearance, h=mental confusion or headache, i=other).* If other is selected, please explain.

Supplemental table 5

Timed 'up and go' Test (TUG)	
I. Scale description	
Are there several versions of the scale?	Yes (in terms of scoring method). The initial study design by Podsiadlo and Richardson consisted of one practice and one test.[1] In HD: 1) Rao et al.[2] used the mean score of three tests. 2) Quinn et al.[3] and Busse et al.[4] used the mean of two tests.
If you replied YES, which was been assessed?	All were considered: mean of two tests, mean of three tests and one practice and one test.[2, 3, 5-9]
Scale construct/ overall structure	The test covers mobility and balance, and falls' risk.[9] The patient sits in the chair with his/her back against the back of the chair. On the command "go", the patient rises from the chair, walks 3 meters at a comfortable and safe pace, turns, walks back to the chair and sits down. Timing begins at the instruction "go" and stops when the patient is seated. Podsiadlo & Richardson[1] quantified the test by recommending timing the duration (sec) between the command "go" and the moment the buttocks touch the chair. The patient should have one practice test that is not included in the score.[1] The patient must use the same assistive device each time he/she is tested so that scores can be compared. The chair also needs to be consistent within and between patients for comparisons to be made. There is an alternative way of scoring performance from 1 to 5 based on the observer's perception of the patient's risk of falling, [10] which has not been used in HD.
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	No.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Scored in continuous values of seconds.

c. Is the scale easy to score?	
Approx. time to score patient	3 minutes total, including set-up. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=903
d. Raters	
Patient, caregiver, or clinician	Clinician.
If clinician-rated, is training for application required?	No.
e. Access to scale	
Copyright or public domain?	Public.
How can the scale be obtained (address or website)?	http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=903
Has the scale been published in other languages?	No (it is not necessary).
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Not applicable.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Yes - sit to stand, walking and turning, and stand to sit (personal judgment).
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	Content of activity reflects more mobility than balance (personal judgment).
Does it score current state or is it based on the patient/caregiver recall?	Current state.
What is the time frame (e.g. "during the past week")?	Not applicable.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To measure severity, screen for ability to go outside alone safely, and screen for risk of falls.[1, 9]
Is there a cut-off score? (for HD, for non-HD)	Yes. For cut-off of TUG>14 / Predicted probability for being a "faller" was virtually 60%. "Fallers" ≥1 fall in the previous 12 months.[9]
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).

Are there ambiguities in instructions to patient/rater (as applicable)?	Number of tests that should be performed varies (practice, test or average of 2 or 3). In some studies, it is not clear how many tests were performed.[1, 3, 8, 9]
Are there ambiguities in rating anchors?	No (personal judgment).
Are the questions appropriate for use in an HD population?	Not applicable.
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	It is not applicable in later stages, it is potentially not sensitive in early stages.[2, 3, 8]
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Not applicable.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.[3]
a. Reliability	
Assessed – not assessed	Assessed.
Internal consistency	Not applicable in HD.
Test-retest reliability	n = 11 (Pre-manifest HD) and n=64 (Manifest HD).[3] 1. Pre-manifest HD: ICC=0.93 2. Manifest HD: ICC=0.96: • early stage (TFC=11–13) - ICC= 0.94. • middle stage (TFC=7–9) - ICC= 0.95. • late stage (TFC≤6) - ICC= 0.97.
Inter-rater reliability	Not assessed in HD.
b. Validity in HD	
Assessed vs. not assessed	Not assessed in HD.
Criterion validity (any comparison with gold-standard)	-
Construct validity	
Factor analysis	Not assessed in HD.

Convergent validity	Manifest HD:[4] (Mean values of TUG as average of two tests). TUG *vs.* UHDRS-TMS (n=61): r= 0.16 (n.s.); TUG *vs.* UHDRS Functional Assessment (n=60): r= -0.33 (p<0.01); TUG *vs.* UHDRS-TFC (n=61): r = -0.25 (n.s.).
Divergent validity	Not assessed in HD.
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Good (personal judgment). Maybe useful in mild stages of the disease when speed/bradykinesia becomes more of a factor, but also becomes more variable at that point.[3]
Generalizability	
Shown to be valid at any stage of HD?	Early-mid stages of HD; not pre-manifest; it appears sensitive to disease progression.[2, 3, 8]
Shown to be valid in any population with dementia or significant cognitive impairment?	The TUG may demonstrate less reliability among patients suffering from cognitive impairment. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=903
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Yes. Manifest HD, n=30. Task-specific mobility training vs. usual care, RCT[7]: Effect size: 0.17, n.s. Manifest HD, n=37. Change in response to a 1-year rehabilitation multidisciplinary, no control. Program intervention = -1.3 sec (group). NOTE: within the MDC. [6] Manifest HD, n=30, 6-week program of intervention for posture and gait, no control. S.S. improvement in TUG but results given in graphic, no values reported. [5]
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No. **NOTE:** MDC: 1.34 seconds in pre-manifest HD and 2.98 sec in manifest HD.[3]
Floor and ceiling effects	Yes, both. Patients were at the lower (better) range of TUG scores and likely have ceiling effects for improvement (data not reported so unable to determine percent).[5]
Score distributions	In HD, ranging from 7-14 sec.[3]
IV. Overall impression	

Advantages	Timing measure - continuous variable. General measure of mobility. Incorporates turning and sit to stand which is a problem with patients.
Disadvantages	May not be sensitive to change. Speed/bradykinesia may not be the biggest problem in patients with HD. The test is not useful in pre-manifest and end stage HD.[2, 3]
V. Recommendation	**Suggested for assessing balance and mobility (severity)** **Suggested for screening for risk of falls**

Supplemental references 5

1. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 1991;39:142-8.
2. Rao AK, Louis ED, Marder KS. Clinical assessment of mobility and balance impairments in pre-symptomatic Huntington's disease. *Gait Posture* 2009;30:391-3.
3. Quinn L, Khalil H, Dawes H, Fritz NE, Kegelmeyer D, Kloos AD, *et al.* Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. *Phys Ther* 2013;93:942-56.
4. Busse M, Quinn L, Khalil H, McEwan K. Optimising mobility outcome measures in Huntington's disease. *J Huntingtons Dis* 2014;3:175-88.
5. Bohlen S, Ekwall C, Hellstrom K, Vesterlin H, Bjornefur M, Wiklund L, *et al.* Physical therapy in Huntington's disease--toward objective assessments? *Eur J Neurol* 2013;20:389-93.
6. Piira A, van Walsem MR, Mikalsen G, Nilsen KH, Knutsen S, Frich JC. Effects of a One Year Intensive Multidisciplinary Rehabilitation Program for Patients with Huntington's Disease: a Prospective Intervention Study. *PLoS Curr* 2013;5.
7. Quinn L, Debono K, Dawes H, Rosser A, Nemeth AH, Quarrell O, *et al.* Task-specific training in Huntington's disease: A randomised, controlled feasibility trial. In; 2014. p. A66-a67.
8. Rao AK, Muratori L, Louis ED, Moskowitz CB, Marder KS. Clinical measurement of mobility and balance impairments in Huntington's disease: validity and responsiveness. *Gait Posture* 2009;29:433-6.
9. Busse ME, Khalil H, Quinn L, Rosser AE. Physical therapy intervention for people with Huntington disease. *Phys Ther* 2008;88:820-31.
10. Mathias S, Nayak US, Isaacs B. Balance in elderly patients: the "get-up and go" test. *Arch Phys Med Rehabil* 1986;67:387-9.

Timed Up and Go Instructions

General Information (derived from Podsiadlo and Richardson, 1991):

- The patient should sit on a standard armchair, placing his/her back against the chair and resting his/her arms chair's arms. Any assistive device used for walking should be nearby.
- Regular footwear and customary walking aids should be used.
- The patient should walk to a line that is 3 meters (9.8 feet) away, turn around at the line, walk back to the chair, and sit down.
- The test ends when the patient's buttocks touch the seat.
- Patients should be instructed to use a comfortable and safe walking speed.
- A stopwatch should be used to time the test (in seconds).

Set-up:

- Measure and mark a 3 meter (9.8 feet) walkway
- Place a standard height chair (seat height 46cm, arm height 67cm) at the beginning of the walkway

Patient Instructions (derived from Podsiadlo and Richardson, 1991):

- Instruct the patient to sit on the chair and place his/her back against the chair and rest his/her arms chair's arms.
- The upper extremities should not be on the assistive device (if used for walking), but it should be nearby.
- Demonstrate the test to the patient.
- When the patient is ready, say "Go"
- The stopwatch should start when you say go, and should be stopped with the patient's buttocks touch the seat.

Supplemental table 6

10-meter walk test (10MWT)	
I. Scale description	
Are there several versions of the scale?	Yes, in the sense that other walking tests of varying length e.g., 6 minutes, exist. The test can be performed at patient self-selected/comfortable[1-3] walking pace or at maximum pace. The 10-meter self-selected or 'comfortable pace' walking pace is the most common.
If you replied YES, which was been assessed?	N/A
Scale construct/ overall structure	The 10MWT assesses walking speed in meters per second over a short duration. Self-selected or 'comfortable pace' walking speed and number of steps taken; 10 meter walk with 2 meters at the beginning and at the end to allow for acceleration and de-acceleration; requires 10 meter floor markings and a stopwatch. Single performance measure scored in meters/seconds and number of steps taken. In HD, the score has been based on the average of two tests.[1, 2] The 10MWT score can also be calculated using the average of the three trials. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=901 **NOTE:** Piira et al.[4] used fast-paced, while Busse et al.[3] used both self-selected and fast paced.
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	Yes.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Continuous score (seconds); step count, although time in seconds or gait speed is more commonly reported. Note: timing measure is problematic, if a person is unable to perform the test.
c. Is the scale easy to score?	
Approx. time to score patient	< 5 minutes.

	http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=901
d. Raters	
Patient, caregiver, or clinician	Clinician.
If clinician-rated, is training for application required?	No, but it is necessary to follow standardized instructions from test to test with documentation of self-selected or fast pace; walking aid is permissible but needs to be documented. Personal assistance with walk is not permitted.
e. Access to scale	
Copyright or public domain?	Public domain.
How can the scale be obtained (address or website)?	http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=901
Has the scale been published in other languages?	Not applicable.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	No.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	No.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	Not applicable.
Does it score current state or is it based on the patient/caregiver recall?	Current state.
What is the time frame (e.g. "during the past week")?	Current time.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To measure severity.
Is there a cut-off score? (for HD, for non-HD)	No. NOTE: there are age- and sex-based normative data in healthy adults aged in their 20s to 70s. Males comfortable speed ranges are 1.39 m/s to 1.33, and females 1.41 to 1.27 m/s and range for fast-paces for males from 2.53 to 2.08 m/s and for females from 2.47 to 1.74

	m/s (note, n=230, age 20-70 measure over 6.62 meters with acceleration and deceleration period so compatible with 10 meter walk – (25 foot walk).[5]
d. Acceptability	
Is the length of the scale appropriate?	Yes.
Are there ambiguities in instructions to patient/rater (as applicable)?	No.
Are there ambiguities in rating anchors?	Not applicable.
Are the questions appropriate for use in an HD population?	Yes, if ambulatory and the patients do not require personal assistance.
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	No, patients need to be able to walk without personal assistance.
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Not applicable.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.
a. Reliability	
Assessed – not assessed	Assessed in HD.
Internal consistency	Not applicable
Test-retest reliability	**10MWT (self-paced)** n = 11 (Pre-manifest HD) and n=64 (Manifest HD).[1] 1. Pre-manifest HD: ICC=0.96. 2. Manifest HD: ICC=0.95: • early stage (TFC=11–13) - ICC= 0.97. • middle stage (TFC=7–9) - ICC= 0.92. • late stage (TFC<=6) - ICC= 0.96. **NOTE:** MDC ranges 0.20 to 0.46 (stage dependent and non-linear). **Piira (2013)**[4]: Fast-paced version. MDC ranged from 0.16 in late manifest HD, mid

	HD, 0.24, 0.20 early manifest HD); 0.20 overall for manifest HD.
Inter-rater reliability	Not assessed.
b. Validity in HD	
Assessed vs. not assessed	Assessed in HD.[2]
Criterion validity (any comparison with gold-standard)	No
Construct validity	
Factor analysis	N/A
Convergent validity	Manifest HD[2] (Mean values of TUG as average of two tests). **10MWT (self-paced)** is not associated with UHDRS Motor Score: correlation -0.19 and -0.15 (n=62, p>0.05) at time 1 and time 2 (one week later); it is significantly associated with the UHDRS Functional Assessment Scale 0.35 and 0.34 (times 1 and 2, respectively, n=62, p<0.05), but not with the UHDRS-TFC (0.24 and 0.22) (time 1 and 2, respectively, n=62). **Personal comment:** Higher correlations are not expected with performance measure and these measures. **NOTE:** not set up as a study of validity.
Divergent validity	No, see above.
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Very good test-retest reliability, reasonable construct validity with measures provided (as expected).
Generalizability	
Shown to be valid at any stage of HD?	Questionable validity in pre-manifest HD. For late HD: potentially not applicable as patients must be ambulatory to perform test (personal judgment).
Shown to be valid in any population with dementia or significant cognitive impairment?	It is unclear if it can be used in the severely cognitively impaired (personal judgment).
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over	**Piira (2013)[4]**: Fast-paced version. Mild to moderate manifest HD (n=37) with

time or to treatment)?	rehabilitation intervention. Statistically significant improvement of 0.27 m/s which exceeds MDC. **Busse (2013)[3]:** Community walking program pilot, RCT (n=18 total) n.s. change in self-paced or fast-paced walking test. Note, means and ES also indicate no effect so it is not simply an issue of small numbers.
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No. NOTE: MDC as per data mentioned in reliability above.
Floor and ceiling effects	No.
Score distributions	n = 11 (Pre-manifest HD) and n=64 (Manifest HD) Mean (SD): [1] 1. Pre-manifest HD: 1.31 (0.31). 2. Manifest HD: 1.20 (0.39). • early stage (TFC=11–13) - 1.34 (0.29). • middle stage (TFC=7–9) - 1.10 (0.42). • late stage (TFC<=6) - 1.15 (0.42). NOTE: scores across HD severity are not linear.[1]
IV. Overall impression	
Advantages	Very easy to administer but critical conditions (e.g., instructions/standardization are reported).
Disadvantages	
V. Recommendation	**Suggested for assessment of walking speed in manifest HD.** (10MWT self-paced has more clinimetric data)

Supplemental references 6

1. Quinn L, Khalil H, Dawes H, Fritz NE, Kegelmeyer D, Kloos AD, *et al.* Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. *Phys Ther* 2013;93:942-56.
2. Busse M, Quinn L, Khalil H, McEwan K. Optimising mobility outcome measures in Huntington's disease. *J Huntingtons Dis* 2014;3:175-88.
3. Busse M, Quinn L, Debono K, Jones K, Collett J, Playle R, *et al.* A randomized feasibility study of a 12-week community-based exercise program for people with Huntington's disease. *J Neurol Phys Ther* 2013;37:149-58.

4. Piira A, van Walsem MR, Mikalsen G, Nilsen KH, Knutsen S, Frich JC. Effects of a One Year Intensive Multidisciplinary Rehabilitation Program for Patients with Huntington's Disease: a Prospective Intervention Study. *PLoS Curr* 2013;5.
5. Bohannon RW. Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants. *Age Ageing* 1997;26:15-9.

Supplemental table 7

Four square step test (FSST)	
I. Scale description	
Are there several versions of the scale?	No.
If you replied YES, which was been assessed?	Not applicable
Scale construct/ overall structure	Dynamic balance measured in seconds.[6] Test of dynamic balance that clinically assesses a person's ability to step over objects forward, sideways, and backwards. A test procedure may be demonstrated and one practice test is allowed prior to administering the actual test: 1. Two trials are then performed, and the better time (in seconds) is taken as the score. 2. Timing starts when the right foot contacts the floor in the square. <u>Instructions:</u> "Try to complete the sequence as fast as possible without touching the sticks. Both feet must make contact with the floor in each square. If possible, face forward during the entire sequence." Repeat a test if the patient: • Fails to complete the sequence successfully. • Loses balance. • Makes contact with the cane. • Patient steps over four canes set-up like a cross on the floor with the tips of the canes facing together. At the start of the test, the patient stands on the upper left square (in Square #1, facing Square #2). • The stepping sequence is (clockwise): o 1, Square 2, Square 4, Square 3, return to Square 1 with both feet. • Then (counterclockwise): o Back to Square 3, Square 4, Square 2, and end in Square 1 with both feet.

	*Patients who are unable to face forward during the entire sequence and may turn before stepping into the next square and are timed accordingly.
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	No.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	No.
c. Is the scale easy to score?	
Approx. time to score patient	5-10 minutes (personal judgment).
d. Raters	
Patient, caregiver, or clinician	Clinician.
If clinician-rated, is training for application required?	No.
e. Access to scale	
Copyright or public domain?	Public domain.
How can the scale be obtained (address or website)?	http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=900
Has the scale been published in other languages?	No.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Not applicable.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Not applicable.
Is the scale heavily weighted towards one aspect of the domain it measures and which	Not applicable.

components of the domain are not covered?	
Does it score current state or is it based on the patient/caregiver recall?	Current state.
What is the time frame (e.g. "during the past week")?	Current time: performance based.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To measure severity of motor planning/balance.
Is there a cut-off score? (for HD, for non-HD)	No.
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	No (personal judgment).
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? **Are there HD stages in which the scale is not applicable?**	Applicable in later stages of HD.[1]
e. Has this scale been specifically developed for use in HD (yes/no)?	No.[6]
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Not applicable.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.
a. Reliability	
Assessed – not assessed	Not assessed in HD.

Internal consistency	Not applicable.
Test-retest reliability	n = 11 (Pre-manifest HD) and n=64 (Manifest HD).[1] 1. Pre-manifest HD: ICC=0.91. 2. Manifest HD: ICC=0.78: • early stage (TFC=11–13) - ICC= 0.74. • middle stage (TFC=7–9) - ICC= 0.53. • late stage (TFC<=6) - ICC= 0.91. Manifest HD (n = 20): ICC= 0.86 (0.76,1.00).[7]
Inter-rater reliability	Not assessed.
b. Validity in HD	
Assessed vs. not assessed	Not assessed.
Criterion validity (any comparison with gold-standard)	No.
Construct validity	
Factor analysis	No.
Convergent validity	Manifest HD (n = 20): FSST *vs.* ABC scale (Pearson correlation = -0.57, p<0.05); *vs.* TMT (Pearson correlation = –0.67, p<0.01); *vs.* gait velocity (Pearson correlation = - 0.69, p<0.01).[7]
Divergent validity	No.
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Limited information.
Generalizability	
Shown to be valid at any stage of HD?	No.
Shown to be valid in any population with dementia or significant cognitive impairment?	No.
Responsiveness (detect change over time in the construct)	

Demonstrated to be sensitive to change (change over time or to treatment)?	No. Adjusted mean difference in one exercise cross-over, controlled, single-blinded, six-week study: - 0.06 (-1.72, 0.60).[8]
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No. NOTE: MDC - Pre-manifest HD, - 1.9[1]; Manifest HD- 15.2[1]; and 3.25.[7]
Floor and ceiling effects	Not reported. In stroke: Floor effect: 40-62% of participants had unsuccessful test at least once during testing.[9]
Score distributions	n = 11 (Pre-manifest HD) and n=64 (Manifest HD) Mean (SD):[1] 1. Pre-manifest HD: 11.68 (2.36). 2. Manifest HD: 14.98 (11.19): • early stage (TFC=11–13) - 12.91 (6.82). • middle stage (TFC=7–9) - 17.50 (8.71). • late stage (TFC<=6) - 14.79 (15.64).
IV. Overall impression	
Advantages	Unique measure of balance and motor planning. Easy to administer. May be good sensitive measure in early stages/pre-manifest HD.
Disadvantages	Is difficult for patients in later stages. If they are unable to complete the test it is not clear how scoring should be performed.
V. Recommendation	**Suggested for assessment of dynamic balance in HD.**

Supplemental references 7

1. Quinn L, Khalil H, Dawes H, Fritz NE, Kegelmeyer D, Kloos AD, *et al.* Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. *Phys Ther* 2013;93:942-56.
2. Busse M, Quinn L, Khalil H, McEwan K. Optimising mobility outcome measures in Huntington's disease. *J Huntingtons Dis* 2014;3:175-88.

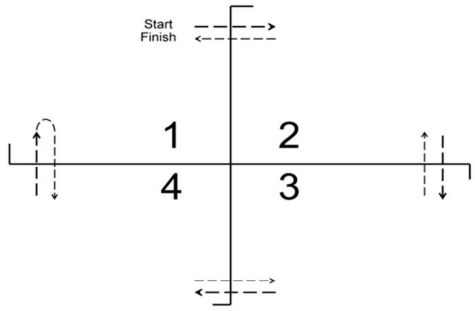
3. Busse M, Quinn L, Debono K, Jones K, Collett J, Playle R, *et al.* A randomized feasibility study of a 12-week community-based exercise program for people with Huntington's disease. *J Neurol Phys Ther* 2013;37:149-58.
4. Piira A, van Walsem MR, Mikalsen G, Nilsen KH, Knutsen S, Frich JC. Effects of a One Year Intensive Multidisciplinary Rehabilitation Program for Patients with Huntington's Disease: a Prospective Intervention Study. *PLoS Curr* 2013;5.
5. Bohannon RW. Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants. *Age Ageing* 1997;26:15-9.
6. Dite W, Temple VA. A clinical test of stepping and change of direction to identify multiple falling older adults. *Arch Phys Med Rehabil* 2002;83:1566-71.
7. Kloos AD, Fritz NE, Kostyk SK, Young GS, Kegelmeyer DA. Clinimetric properties of the Tinetti Mobility Test, Four Square Step Test, Activities-specific Balance Confidence Scale, and spatiotemporal gait measures in individuals with Huntington's disease. *Gait Posture* 2014;40:647-51.
8. Kloos AD, Fritz NE, Kostyk SK, Young GS, Kegelmeyer DA. Video game play (Dance Dance Revolution) as a potential exercise therapy in Huntington's disease: a controlled clinical trial. *Clinical rehabilitation* 2013;27:972-82.
9. Blennerhassett JM, Jayalath VM. The Four Square Step Test is a feasible and valid clinical test of dynamic standing balance for use in ambulant people poststroke. *Arch Phys Med Rehabil* 2008;89:2156-61.

Four Step Square Test Instructions

General Information:

- The patient is instructed to stand in square 1 facing square number 2 (see figure below)
- The patient is required to step as fast as possible into each square in the following sequence: 2, 3, 4, 1, 4, 3, 2, and 1
  - requires the patient to step forward, backward, and sideway to the right and left
- Equipment required for the FSST includes a stopwatch and 4 canes.

Set-up (derived from Dite and Temple 2002): A square is formed with the 4 canes by resting them flat on the floor.



Patient Instructions (derived from Dite and Temple 2002):

- "Try to complete the sequence as fast as possible without touching the sticks. Both feet must make contact with the floor in each square. If possible, face forward during the entire sequence."
- Demonstrate the sequence to the patient.
- Ask the patient to complete one practice trial to ensure the patient knows the sequence. Repeat the trial if the patient is unsuccessful at completing the sequence, loses balance, or contacts a cane during the trial.
- Two FSST are completed with the best time taken as the score.
- A score is still provided if the patient is unable to face forward during the entire sequence.

Scoring:

- the best time of two FSST is the score
- stopwatch starts when the first foot contacts the floor in square 2
- stopwatch finishes when the last foot comes back to touch the floor in square 1

Supplemental table 8

Mini-BESTest	
I. Scale description	
Are there several versions of the scale?	No. The Mini-BESTest, the original BESTest 4 level (0 - 3) scoring was revised to 3 levels (0 - 2) due to redundancy.
If you replied YES, which was been assessed?	–
Scale construct/ overall structure	The Mini BESTest is a 14-item test scored on a 3-level ordinal scale as a measure of dynamic balance. Total score = 28 points per test directions. Two items have right and left assessment in which the lower score is used within the total score (directions specify which to use). For research, many studies specify use of both left and right data, thus calculating data based on 32 (vs. 28) points.[1]
a. Question items	
Items of presence of symptom/sign?	—
Items of severity of symptom/sign?	—
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Discrete steps (3-level ordinal scale).
c. Is the scale easy to score?	
Approx. time to score patient	10-15 minutes (personal judgment).
d. Raters	
Patient, caregiver, or clinician	Clinician.
If clinician-rated, is training for application required?	No.
e. Access to scale	
Copyright or public domain?	Copyright.
How can the scale be obtained (address or website)?	—

Has the scale been published in other languages?	Not applicable.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Item generation described in original author paper.[1]
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Dynamic balance scale, evaluates domains of postural control namely anticipatory postural adjustments during voluntary postural transitions, postural responses to an externally induced loss of balance, standing balance under challenging sensory conditions and gait.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	No.
Does it score current state or is it based on the patient/caregiver recall?	Current state.
What is the time frame (e.g. "during the past week")?	Not applicable
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To measure severity.
Is there a cut-off score? (for HD, for non-HD)	No. A cut-off of 27 has been used to differentiate HD vs. non-HD (82% specificity and 78% sensitivity).[2]
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	No (personal judgment).
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not	Not applicable for non-ambulatory HD (personal judgment).

applicable?	
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Not applicable.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	
a. Reliability	
Assessed – not assessed	Not assessed in HD.
Internal consistency	–
Test-retest reliability	–
Inter-rater reliability	–
b. Validity in HD	
Assessed vs. not assessed	Assessed in HD.
Criterion validity (any comparison with gold-standard)	No gold standard available.
Construct validity	
Factor analysis	
Convergent validity	Manifest HD (n=18). Association between Mini-BESTest and: ABC-UK (r2=0.45, p=0.0024), UHDRS-TFC (r2=0.75, p <0.0001) and UHDRS-TMS (r2=0.68, p =0.00003).[2]
Divergent validity	-
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Limited information.
Generalizability	
Shown to be valid at any stage of HD?	Unknown; not assessed comprehensively across stages of HD (n=9 stage 1, n=6 stage 2, n=3 stage 3).[2]
Shown to be valid in any population with dementia or significant cognitive impairment?	Unknown.

Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Not in HD.
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	Not in HD.
Floor and ceiling effects	Likely to have floor effects in HD, not applicable for non-ambulatory individuals.
Score distributions	Mean (95% CI) in non-HD =98 (96-99); in HD- 76 (64-87).
IV. Overall impression	
Advantages	Comprehensively assesses multiple domains of balance impairment.
Disadvantages	Not fully validated in HD.
V. Recommendation	**Suggested for assessing severity of balance impairment in HD.**

Supplemental references 8

1. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. *J Rehabil Med* 2010;42:323-31.
2. Jacobs JV, Boyd JT, Hogarth P, Horak FB. Domains and correlates of clinical balance impairment associated with Huntington's disease. *Gait Posture* 2015;41:867-70.

Supplemental table 9

Physical Performance Test (PPT)	
I. Scale description	
Are there several versions of the scale?	Yes. Two Versions: 9-item scale and 7-item scale. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1104
If you replied YES, which was been assessed?	The 9-item scale.[1]
Scale construct/ overall structure	The scale is composed of multiple domains of physical function using observed performance of tasks that simulate activities of daily living of various degrees of difficulty.[1] A 5-point scale of (0-4) on each item. Minimum score of 0 for both scales. Maximum of 36 for 9-item scale. A higher total score is indicative of better physical performance. Subject is given a command "go" to perform a task. Time to comple, in seconds, is recorded. A corresponding score is given from 0-4 determined by seconds taken to complete the task. Scores from each task are totaled.[1] http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1104
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	No.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	5 point scale (0-4).[1]
c. Is the scale easy to score?	
Approx. time to score patient	5-10 minutes. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1104
d. Raters	
Patient, caregiver, or clinician	Clinician.
If clinician-rated, is training for application required?	No.

e. Access to scale	
Copyright or public domain?	Public domain.
How can the scale be obtained (address or website)?	http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1104
Has the scale been published in other languages?	No.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Unknown.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Yes - writing, eating, dressing, walking, and climbing stairs.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	Balance not specifically covered; items apply to general functional tasks (personal judgment).
Does it score current state or is it based on the patient/caregiver recall?	Current state. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1104
What is the time frame (e.g. "during the past week")?	Current state. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1104
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To measure severity of mobility.
Is there a cut-off score? (for HD, for non-HD)	No (personal judgment).
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	No (personal judgment).
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Possibly pre-manifest. Applicable in late stages.[2]

e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Not applicable.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.
a. Reliability	
Assessed – not assessed	Assessed.
Internal consistency	Not assessed.
Test-retest reliability	n = 11 (Pre-manifest HD) and n=64 (Manifest HD) [2] 1. Pre-manifest HD: ICC=0.76. 2. Manifest HD: ICC=0.95: • early stage (TFC=11–13) - ICC= 0.92. • middle stage (TFC=7–9) - ICC= 0.93. • late stage (TFC<=6) - ICC= 0.94. **NOTE:** MDC: 3 in pre-manifest HD ; 5 in manifest HD.[2]
Inter-rater reliability	Not assessed.
b. Validity in HD	
Assessed vs. not assessed	Assessed.
Overall impression: good – not good	Limited information.
Criterion validity (any comparison with gold-standard)	Not assessed in HD.
Construct validity	
Factor analysis	Not assessed in HD.
Convergent validity	**Busse (2014)**[3]: (although not really set up as study of validity): (n=63) manifest HD, PPT was correlated with: UHDRS TMS (r=-0.4; p<0.01); UHDRS Functional Assessment Scale (r=0.59, p<0.01; and Total Functional Capacity (r=0.48, p<0.05). **Parkinson's disease:** Concurrent validity in PD: Good correlation of the 9-item scale with basic Katz Activities of daily living (r = 0.65).

Divergent validity	Not assessed in HD.
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Good (personal judgment).
Generalizability	
Shown to be valid at any stage of HD?	No. Not in pre-manifest HD.[2]
Shown to be valid in any population with dementia or significant cognitive impairment?	Yes.[4]
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Task-specific training vs. usual care in 30 HD patients. Effect size=0.01.[2] Non-randomized study with no control group of an inpatient rehabilitation program. manifest HD, n=40. After a three-week period of treatment there was a significant average improvement in the PPT (5.21, $p<0.001$)[2, 5] RCT of structured home-based exercise *vs.* usual care: early to moderate HD with walking or balance difficulties, n=25: Mean difference: 4.8 (95% CI: 2.0, 7.7, $p=0.002$).[6]
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	There is a ceiling effect in pre-manifest HD.[2]
Score distributions	n = 11 (Pre-manifest HD) and n=64 (Manifest HD) Mean (SD):[2] 1. Pre-manifest HD: 31 (2). 2. Manifest HD: 23 (7): • early stage (TFC=11–13) - 27 (5). • middle stage (TFC=7–9) - 22 (7). • late stage (TFC≤6) - 20 (7).

	Comment: whether values discriminate among pre- and manifest severity levels is unclear as confidence intervals (CI) on means overlap; may separate pre- and mild from mid/severe levels based on CIs.
IV. Overall impression	
Advantages	Measures range of physical functioning; timed measures are quantitative and not subjective.
Disadvantages	Cut-offs for categories (converting time to numbers) may not be appropriate for HD. Need some equipment - e.g., coffee can; beans, lab coat. Don't know if converted scores are appropriate for HD (have been validated in other populations).
V. Recommendation	**Suggested for severity of impairment of physical function (activities of daily living).**

Supplemental references 9

1. Reuben DB, Siu AL. An objective measure of physical function of elderly outpatients. The Physical Performance Test. *J Am Geriatr Soc* 1990;38:1105-12.
2. Quinn L, Khalil H, Dawes H, Fritz NE, Kegelmeyer D, Kloos AD*, et al.* Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. *Phys Ther* 2013;93:942-56.
3. Busse M, Quinn L, Khalil H, McEwan K. Optimising mobility outcome measures in Huntington's disease. *J Huntingtons Dis* 2014;3:175-88.
4. Farrell MK, Rutt RA, Lusardi MM, Williams AK. Reliability of the Physical Performance Test in People with Dementia. *Physical & Occupational Therapy In Geriatrics* 2010;28:144-53.
5. Zinzi P, Salmaso D, De Grandis R, Graziani G, Maceroni S, Bentivoglio A*, et al.* Effects of an intensive rehabilitation programme on patients with Huntington's disease: a pilot study. *Clin Rehabil* 2007;21:603-13.
6. Khalil H, Quinn L, van Deursen R, Dawes H, Playle R, Rosser A*, et al.* What effect does a structured home-based exercise programme have on people with Huntington's disease? A randomized, controlled pilot study. *Clinical rehabilitation* 2013;27:646-58.

Physical Performance Test

Testing Protocol: Administer the test as outlined below. Subjects are given up to two chances to complete each item. Assistive devices are permitted for tasks 6 – 9.

1. Ask the subject, when given the command to "go" to write the sentence "whales live in the blue ocean." Time from the word "go" until the pen is lifted from the page at the end of the sentence. All words must be included and legible. Period need not be included for task to be considered completed.
2. Five kidney beans are placed in a bowl, 5 inches from the edge of the desk in front of the patient. An empty coffee can is placed on the table at the patient's non-dominant side. A teaspoon is place in the patient's dominant hand. Ask the subject on the command "go" to pick up the beans, one at a time and place each in the coffee can. Time from the command "go" until the last bean is heard hitting the bottom of the can.
3. Place a Physician's Desk Reference or other heavy book on a table in front of the patient. Ask the patient, when given the command "go" to place the book on a shelf above shoulder level. Time from the command "go" to the time the book is resting on the shelf.
4. If the subject has a jacket cardigan sweater, ask them to remove it. If not, give the subject a lab coat. Ask the subject, on the command "go" to put the coat on completely such that it is straight on their shoulders and then remove the garment completely. Time from the command "go" until the garment has been complexly removed.
5. Place a penny approximately1 foot from the patient's foot on the dominant side. Ask the patient, on the command "go" to pick up the penny from the floor and stand up. Time from the command "go" until the subject is standing erect with a penny in hand.
6. With subject in a corridor or in and open room, ask the subject to turn 360 degrees. Evaluate using the scale on PPT scoring sheet.
7. Bring subject to start on a 50 –foot walk test course (25 feet out and 25 feet back) and ask the subject, on the command "go" to walk to the 25-foot mark and back. Time from the command "go" until the starting line is crossed on the way back.
8. Bring subject to foot of stairs (nine to 12 steps) and ask subject, on the command "go" to begin climbing stairs until they feel tired and wishes to stop. Before beginning this task, alert the subject to the possibility of developing chest pain or shortness of breath and inform the subject to tell you if any of these symptoms occur. Escort the subject up the stairs. Time from the command "go" until the subjects' first foot reaches the top of the first flight of stairs. Record the number of flights (maximum is four) completed (up and down is one flight).

Physical Performance Test

Scoring Sheet

			Time	Scoring	Score
1.	Write a sentence. (Whales live in the blue ocean.)	Seconds		≤ 10 sec = 4 10.5-15 sec = 3 15.5 – 20 sec = 2 >20 sec = 1 unable = 0	
2.	Simulated eating	Seconds		≤ 10 sec = 4 10.5-15 sec = 3 15.5 – 20 sec = 2 >20 sec = 1 unable = 0	
3.	Lift a book and put it on a shelf Book PDR 1988: 5.5 lbs Bed height 59 cm Shelf height 118 cm All sitting with feet on floor	Seconds		≤ 2 sec = 4 2.5- 4 sec = 3 4.5 – 6 sec = 2 > 6 sec = 1 unable = 0	
4.	Put on and remove a jacket 1. Standing 2. Use of bathrobe; button down shirt; hospital gown.	Seconds		≤ 10 sec = 4 10.5-15 sec = 3 15.5 – 20 sec = 2 >20 sec = 1 unable = 0	
5.	Pick up a penny from floor.	Seconds		≤ 2 sec = 4 2.5- 4 sec = 3 4.5 – 6 sec = 2 > 6 sec = 1 unable = 0	
6.	Turn 360 degrees		Discontinuous steps = 0 Continuous steps = 2		
			Unsteady (grabs, staggers) = 0 Steady = 2		
7.	50-foot walk test.	Seconds		≤ 15 sec = 4	

	Starting sitting for instructions.			15.5- 20 sec = 3 20.5 – 25 sec = 2 >25 sec = 1 unable = 0	
8.	Climb one flight of stairs.+	Seconds		≤ 5 sec = 4 5.5- 10 sec = 3 10.5 – 15 sec = 2 >15 sec = 1 unable = 0	
9.	Climb stairs.+		Number of flights of stairs up and down (maximum 4)		
	TOTAL SCORE (maximum 36 for nine-item, 28 for seven-item)				
	(*Round time measurements to nearest 0.5 seconds.) (+ omit for 7 item test)			9-item score	

Supplemental table 10

Six condition Romberg test (in s)	
I. Scale description	
Are there several versions of the scale?	The 6 condition Romberg test can be scored as individual conditions. Stance with feet together, tandem stance and one limb stance have been assessed in HD.
If you replied YES, which was been assessed?	The 6 condition Romberg test has been assessed.
Scale construct/ overall structure	Assessed standing balance under various conditions. Romberg and Sharpened Romberg tests are performed with and without cognitive loading and with eyes open and eyes closed. Starting position is feet close together, with both eyes open and then eyes closed, for a maximum of 30 seconds under each condition. Participants then perform the Sharpened Romberg test by standing with one foot placed directly in front of the other, with the front heel touching the toes of the back foot (tandem standing), again with eyes open and eyes closed. Both conditions of the Sharpened Romberg test are performed as a dual task, with the addition of a secondary cognitive task (counting backward by 3 from100). The amount of time the patient maintains the position without loss of balance for all 6 conditions is recorded (maximum score 180 seconds, with higher scores indicating better balance). http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1173; http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1160
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	No.
b. Response scale	

Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Time is scored.
c. Is the scale easy to score?	
Approx. time to score patient	5 minutes (personal judgment).
d. Raters	
Patient, caregiver, or clinician	Clinician (personal judgment).
If clinician-rated, is training for application required?	No.
e. Access to scale	
Copyright or public domain?	Public domain.
How can the scale be obtained (address or website)?	Not applicable
Has the scale been published in other languages?	–
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	No.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Not applicable. Performance based test.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	Not applicable. Performance based test.
Does it score current state or is it based on the patient/caregiver recall?	Current state.
What is the time frame (e.g. "during the past week")?	Not applicable. Performance based test.
c. Use	

Purpose: to measure severity, screen or diagnosis of the domain?	To measure severity of balance impairment.
Is there a cut-off score? (for HD, for non-HD)	No.
d. Acceptability	
Is the length of the scale appropriate?	Not applicable.
Are there ambiguities in instructions to patient/rater (as applicable)?	Not applicable.
Are there ambiguities in rating anchors?	Not applicable.
Are the questions appropriate for use in an HD population?	Not applicable.
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	No, not applicable in non-ambulatory HD.
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Not applicable.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	
a. Reliability	
Assessed – not assessed	Assessed in HD.[1]
Internal consistency	–
Test-retest reliability	n = 11 (Pre-manifest HD) and n=64 (Manifest HD).[1] 1) Pre-manifest HD: ICC=0.73. 2) Manifest HD: ICC=0.89: • early stage (TFC=11–13) - ICC= 0.91. • middle stage (TFC=7–9) - ICC= 0.86.

	• late stage (TFC<=6) - ICC= 0.84. **Personal comment:** (relatively consistent across disease stages).
Inter-rater reliability	—
b. Validity in HD	
Assessed vs. not assessed	Not assessed in HD.
Criterion validity (any comparison with gold-standard)	No defined gold standard.
Construct validity	
Factor analysis	N/A
Convergent validity	Not assessed.
Divergent validity	Not assessed.
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	No information for an impression to be given.
Generalizability	
Shown to be valid at any stage of HD?	Yes, if the patient is ambulatory.[1]
Shown to be valid in any population with dementia or significant cognitive impairment?	No (personal judgment).
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	No.[2]
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	MDC: pre-manifest HD=29.70 seconds, manifest HD=37.43 seconds (fairly consistent across stages).[1]
Floor and ceiling effects	Likely to have floor effects for non-ambulatory individuals.
Score distributions	Pre-manifest HD: mean (SD) 158.77 (22.22), manifest HD: mean (SD) 69.98 (41.06).[1]

IV. Overall impression	
Advantages	Useful to differentiate between pre-manifest and manifest HD. May have potential as a tool for identifying early stage clinical improvement
Disadvantages	Difficult to administer in the presence of cognitive deficits.
V. Recommendation	**Suggested for assessing severity of balance impairment in HD.**

Supplemental references 10

1. Quinn L, Khalil H, Dawes H, Fritz NE, Kegelmeyer D, Kloos AD*, et al.* Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. *Phys Ther* 2013;93:942-56.
2. Kegelmeyer DA, Kloos AD, Fritz NE, Fiumedora MM, White SE, Kostyk SK. Impact of tetrabenazine on gait and functional mobility in individuals with Huntington's disease. *J Neurol Sci* 2014;347:219-23.

Supplemental table 11

Functional Reach Test	
I. Scale description	
Are there several versions of the scale?	Yes, one conducted in a standing position and a modified version conducted in a sitting position. http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=950
If you replied YES, which was been assessed?	Standing version.[3]
Scale construct/ overall structure	The test assesses a patient's stability. It consists of a single task, the patient is asked to reach outside the base of support, and the furthest distance reached is measured in inches/centimeters. Instructions:[4, 5] • The patient is instructed to stand close to, but not touching, a wall, and position the arm that is closest to the wall at 90 degrees of shoulder flexion with a closed fist. • The assessor records the starting position at the 3rd metacarpal head on the yardstick. • Instruct the patient to "Reach as far as you can forward without taking a step". • The location of the 3rd metacarpal is recorded. • The difference between the start and end position is the distance reached, usually measured in inches. • The test allows for five total trials: two practice trials, followed by three "test" trials. The distances of the last three trials are averaged to obtain the patient's score. http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=950
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	No.

b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	No, measured as continuous values (inches).
c. Is the scale easy to score?	
Approx. time to score patient	< 2 minutes.[3, 6] < 5 minutes (personal judgment).
d. Raters	
Patient, caregiver, or clinician	Clinician.
If clinician-rated, is training for application required?	No, but standardized instructions must be followed.
e. Access to scale	
Copyright or public domain?	Public domain.
How can the scale be obtained (address or website)?	http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=950
Has the scale been published in other languages?	Not applicable.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Not applicable.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Not applicable.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	Not applicable.
Does it score current state or is it based on the patient/caregiver recall?	Current time.
What is the time frame (e.g. "during the past week")?	Current time.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To screen for risk of falls.
Is there a cut-off score? (for HD, for non-HD)	No.

	NOTE: non-HD, cut-off values have been looked at in various conditions in relation to fall risk. http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=950
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	Not applicable.
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	No, did not identify changes in pre-manifest HD.[6]
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Not applicable.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.
a. Reliability	Not assessed in HD.
Assessed – not assessed	Not assessed in HD.
Internal consistency	Not assessed in HD.
Test-retest reliability	Not assessed in HD. <u>Other conditions:</u> Community dwelling elderly. ICC = 0.89-0.92.[4, 5] Parkinson's disease: ICC = 0.84[7] in one study, while another reported ICC= 0.42 in PD with no falls history and a ICC=0.93 if falls were present in history.[8]
Inter-rater reliability	Not assessed in HD. <u>Other conditions:</u> Multiple studies outside PD: all generally ≥ 0.90, [4, 5, 7] In PD ICC was reported to be

	0.64.[9]
b. Validity in HD	
Assessed vs. not assessed	Assessed in HD.
Criterion validity (any comparison with gold-standard)	No.
Construct validity	
Factor analysis	No.
Convergent validity	Pre-manifest HD (n=15):[6] 1) FRT *vs.* gait parameters of velocity, stride length, cadence (correlations <0.23 and not significantly different than 0). 2) FRT *vs.* dynamic balance, measures: double support % and support base (correlations n.s).[6] Manifest HD (n=64, excluded: severe dementia or if a walking aid was used, scored average of three tests): 1) FRT *vs.* gait parameters of velocity: stride length, cadence (correlations = 0.70, 0.81, 0.60, all significant). 2) FRT *vs.* dynamic balance measures: double support % and support base (correlations =0.581 and -0.440, all significant).[3] 3) FRT *vs* UHDRS-TFC (correlation = 0.66, p<0.001).[3] 4) FRT *vs* HD-ADL (correlation = - 0.451, p<0.05).[3]
Divergent validity	
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Good; no concerns based on existing data.
Generalizability	
Shown to be valid at any stage of HD?	Not for pre-manifest HD.[6] NOTE: Predicted symptom onset was not correlated with FRT (Correlations = 0.23, P = 0.44).[6]

	Manifest HD: increasing HD severity correlated with decreased scores in the FRT: stage I - 14.39 cm, stage II - 12.39 cm, stage III - 8.44 cm, overall and pair-wise comparisons were significant (p values <0.01).[3]
Shown to be valid in any population with dementia or significant cognitive impairment?	Not tested in severe dementia.
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Not assessed in HD.
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	Not assessed in HD.
Floor and ceiling effects	Not assessed in HD.
Score distributions	Manifest HD: increasing HD severity correlated with decreased scores in the FRT: stage I - 14.39 cm, stage II - 12.39 cm, stage III - 8.44 cm, overall and pair-wise comparisons were significant (p values <0.01).[3]
IV. Overall impression	
Advantages	It is a short, easy to use scale that is discriminative.
Disadvantages	One used by one group, and because there are some data available – albeit minimal – the criteria for "suggested" are met.
V. Recommendation	**Suggested *with caveats*** *Caveats relate with disadvantages listed above.*

Supplemental references 11

1. Quinn L, Khalil H, Dawes H, Fritz NE, Kegelmeyer D, Kloos AD, *et al.* Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. *Phys Ther* 2013;93:942-56.
2. Kegelmeyer DA, Kloos AD, Fritz NE, Fiumedora MM, White SE, Kostyk SK. Impact of tetrabenazine on gait and functional mobility in individuals with Huntington's disease. *J Neurol Sci* 2014;347:219-23.

3. Rao AK, Muratori L, Louis ED, Moskowitz CB, Marder KS. Clinical measurement of mobility and balance impairments in Huntington's disease: validity and responsiveness. *Gait Posture* 2009;29:433-6.
4. Weiner DK, Duncan PW, Chandler J, Studenski SA. Functional reach: a marker of physical frailty. *J Am Geriatr Soc* 1992;40:203-7.
5. Duncan PW, Weiner DK, Chandler J, Studenski S. Functional reach: a new clinical measure of balance. *J Gerontol* 1990;45:M192-7.
6. Rao AK, Louis ED, Marder KS. Clinical assessment of mobility and balance impairments in pre-symptomatic Huntington's disease. *Gait Posture* 2009;30:391-3.
7. Schenkman M, Cutson TM, Kuchibhatla M, Chandler J, Pieper C. Reliability of impairment and physical performance measures for persons with Parkinson's disease. *Phys Ther* 1997;77:19-27.
8. Smithson F, Morris ME, Iansek R. Performance on clinical tests of balance in Parkinson's disease. *Phys Ther* 1998;78:577-92.
9. Lim LI, van Wegen EE, de Goede CJ, Jones D, Rochester L, Hetherington V*, et al.* Measuring gait and gait-related activities in Parkinson's patients own home environment: a reliability, responsiveness and feasibility study. *Parkinsonism Relat Disord* 2005;11:19-24.

Supplemental table 12

5 Times Sit to Stand Test (FTSST)	
I. Scale description	
Are there several versions of the scale?	No.
If you replied YES, which was been assessed?	Not applicable.
Scale construct/ overall structure	A measure of functional lower limb muscle strength. May be useful in quantifying functional change of transitional movements. Measures time to complete 5 repetitions of chair stands from a standard chair with arms, with arms crossed at shoulders. Inability to complete five repetitions without assistance or use of upper extremity support indicates failure of test. (any modifications should be documented). http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=1015
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	No.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Continuous value as it is a measure based on timing.
c. Is the scale easy to score?	
Approx. time to score patient	1 minute. http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=1015
d. Raters	
Patient, caregiver, or clinician	Clinician. http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=1015
If clinician-rated, is training for application required?	No.
e. Access to scale	

Copyright or public domain?	Public domain.
How can the scale be obtained (address or website)?	http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=1015
Has the scale been published in other languages?	No.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Not applicable.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Not applicable.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	Not applicable.
Does it score current state or is it based on the patient/caregiver recall?	Current time.
What is the time frame (e.g. "during the past week")?	Current time – performance based.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	Measures severity, i.e., the ability to move from sitting to standing.
Is there a cut-off score? (for HD, for non-HD)	No.
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	No (personal judgment).
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Unknown; possibly applicable across later stages (personal judgment).
e. Has this scale been specifically developed for use	No.

in HD (yes/no)?	
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Not applicable.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	No
a. Reliability	
Assessed – not assessed	Not assessed in HD.
Internal consistency	Not assessed in HD.
Test-retest reliability	Not assessed in HD.
Inter-rater reliability	Not assessed in HD.
b. Validity in HD	
Assessed vs. not assessed	Not assessed in HD.
Criterion validity (any comparison with gold-standard)	No.
Construct validity	
Factor analysis	Not assessed in HD.
Convergent validity	Not assessed in HD.
Divergent validity	Not assessed in HD.
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Fair.
Generalizability	
Shown to be valid at any stage of HD?	Not assessed.
Shown to be valid in any population with dementia or significant cognitive impairment?	No.
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Yes. Manifest HD (n=15): Open label assessment OFF (at least 18 h before testing) and ON

	tetrabenazine (TBZ): OFF-TBZ - 15.52 (3.91) seconds; ON-TBZ: 12.61(3.00) seconds.[1]
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	Yes.
Score distributions	Unknown. Only one report. Values described above in ***"Demonstrated to be sensitive to change (change over time or to treatment)?"***
IV. Overall impression	
Advantages	Easy to administer. Quick; continuous variable.
Disadvantages	Does not consider how someone performs the task. Speed may not be the primary problem. *There is limited indirect data on responsiveness based on one study.*
V. Recommendation	**Suggested *with caveats*** *Caveats relate with disadvantage in italic listed above.*

Supplemental references 12

1. Kegelmeyer DA, Kloos AD, Fritz NE, Fiumedora MM, White SE, Kostyk SK. Impact of tetrabenazine on gait and functional mobility in individuals with Huntington's disease. *J Neurol Sci* 2014;347:219-23.

Five times Sit to Stand Test:

Method:
Use a straight back chair with a solid seat that is 16” high. Ask participant to sit on the chair with arms folded across their chest.

Instructions:
“Stand up and sit down as quickly as possible 5 times, keeping your arms folded across your chest.”

Measurement:
Stop timing when the participant stands the 5th time.

Supplemental table 13

30 Second Chair Stand (30CST)	
I. Scale description	
Are there several versions of the scale?	No. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1122
If you replied YES, which was been assessed?	Not applicable.
Scale construct/ overall structure	The test measures the strength of the lower extremities. Consists of measuring the number of chair stands a patient can perform in 30 seconds. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1122
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	Yes.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	No, a count.
c. Is the scale easy to score?	
Approx. time to score patient	30 seconds (personal judgment).
d. Raters	
Patient, caregiver, or clinician	Clinician.
If clinician-rated, is training for application required?	No, but standardized instructions should be followed.
e. Access to scale	
Copyright or public domain?	Public domain.
How can the scale be obtained (address or website)?	http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1122
Has the scale been published in other	Not applicable.

languages?	
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	No.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Not applicable.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	Not applicable.
Does it score current state or is it based on the patient/caregiver recall?	Current state.
What is the time frame (e.g. "during the past week")?	Current time.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To measure severity.
Is there a cut-off score? (for HD, for non-HD)	No, but there are age-based normative data for community dwelling elderly. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1122
d. Acceptability	
Is the length of the scale appropriate?	Yes.
Are there ambiguities in instructions to patient/rater (as applicable)?	No.
Are there ambiguities in rating anchors?	No.
Are the questions appropriate for use in an HD population?	Not applicable.
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not	It has only been used in early to mid-stage HD.[1, 2]

applicable?	
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Not applicable.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	No.
a. Reliability	
Assessed – not assessed	Not assessed in HD.
Internal consistency	Not assessed in HD.
Test-retest reliability	Not assessed in HD.
Inter-rater reliability	Not assessed in HD.
b. Validity in HD	
Assessed vs. not assessed	Not assessed in HD.
Criterion validity (any comparison with gold-standard)	–
Construct validity	
Factor analysis	Not applicable.
Convergent validity	Not assessed in HD.
Divergent validity	Not assessed in HD.
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	–
Generalizability	
Shown to be valid at any stage of HD?	Not assessed in HD.
Shown to be valid in any population with dementia or significant cognitive impairment?	Not applicable.

Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Task-specific training *vs.* usual care in a randomized feasibility 8-week trial; Manifest HD, n=30: Treatment effect: 1.3 (95% CI: -0.7, 3.3, n.s.).[2] Trial of structured home-based exercise *vs.* usual care: early to moderate HD with walking or balance difficulties, n=25: Mean difference: 3.4 (95% CI: 1.0–5.7, p=0.008).[1]
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	Not tested.
Score distributions	Unknown.
IV. Overall impression	
Advantages	Short test requiring minimal equipment, and easy to carry out.
Disadvantages	Virtually no clinimetric data
V. Recommendation	**Suggested *with caveats***

Supplemental references 13

1. Khalil H, Quinn L, van Deursen R, Dawes H, Playle R, Rosser A, *et al.* What effect does a structured home-based exercise programme have on people with Huntington's disease? A randomized, controlled pilot study. *Clinical rehabilitation* 2013;27:646-58.
2. Quinn L, Debono K, Dawes H, Rosser AE, Nemeth AH, Rickards H, *et al.* Task-specific training in Huntington disease: a randomized controlled feasibility trial. *Phys Ther* 2014;94:1555-68.

Supplemental table 14

Dynamic Gait Index (DGI)	
I. Scale description	
Are there several versions of the scale?	Yes. There are 8 and 4-item tests available. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=898
If you replied YES, which was been assessed?	The 8-item version.
Scale construct/ overall structure	Assesses an individual's ability to modify balance while walking in the presence of external demands. Performed with a marked distance of 20 feet. Can be performed with or without an assistive device. Tasks include 1) steady state walking, 2) walking with changing speeds, 3) walking with head turns both horizontally and vertically, 4) walking while stepping over and around obstacles, 5) pivoting while walking, and 6) stair climbing. Highest possible score is 24 points. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=898
a. Question items	
Items of presence of symptom/sign?	Not applicable.
Items of severity of symptom/sign?	Not applicable.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Scores are based on a 4-point scale: 3 = No gait dysfunction 2 = Minimal impairment 1 = Moderate impairment 0 = Severe impairment http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=898
c. Is the scale easy to score?	
Approx. time to score patient	<10 minutes http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=898

d. Raters	
Patient, caregiver, or clinician	Clinician.
If clinician-rated, is training for application required?	None.
e. Access to scale	
Copyright or public domain?	Public.
How can the scale be obtained (address or website)?	http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=898
Has the scale been published in other languages?	Yes. Spanish and Arabic. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=898
II. Scale properties	
a. Content validity	See: http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=898 for detailed review of clinimetric properties assessed in non-HD.
Any process for item generation and/or reduction	Not applicable.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	–
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	–
Does it score current state or is it based on the patient/caregiver recall?	Current.
What is the time frame (e.g. "during the past week")?	Current.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To measure severity and screen for risk of falls.
Is there a cut-off score? (for HD, for non-HD)	Not in HD. Only in non-HD (e.g., PD, community dwelling elderly): http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=898
d. Acceptability	

Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	No (personal judgment).
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Not applicable in non-ambulatory.
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Not applicable.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	No.
a. Reliability	
Assessed – not assessed	Not assessed in HD.
Internal consistency	Not assessed in HD.
Test-retest reliability	Not assessed in HD.
Inter-rater reliability	Not assessed in HD.
b. Validity in HD	
Assessed vs. not assessed	Not assessed in HD.
Overall impression: good – not good	–
Criterion validity (any comparison with gold-standard)	–
Construct validity	
Factor analysis	Not assessed in HD.
Convergent validity	Not assessed in HD.
Divergent validity	Not assessed in HD.

Overall impression: good – not good (based on references preferably, personal judgment can be stated)	No data.
Generalizability	
Shown to be valid at any stage of HD?	Unknown. Not applicable to non-ambulatory HD (personal judgment).
Shown to be valid in any population with dementia or significant cognitive impairment?	Unknown.
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Withdrawal of tetrabenazine resulted in significant reduction of DGI scores in an HD cohort (n=10): OFF-TBZ: 14.4 (7.01), ON-TBZ: 17.5 (6.94). NOTE: no change in cognitive or behavioral measures.[1]
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	Unknown.
Score distributions	Unknown.
IV. Overall impression	
Advantages	-
Disadvantages	Very limited use in HD.
V. Recommendation	**Suggested *with caveats*.**

Supplemental references 14

1. Fekete R, Davidson A, Jankovic J. Clinical assessment of the effect of tetrabenazine on functional scales in huntington disease: a pilot open label study. *Tremor Other Hyperkinet Mov (N Y)* 2012;2.

Supplemental table 15

Walking while talking test (WWTT)	
I. Scale description	
Are there several versions of the scale?	Yes, simple and complex versions.[1]
If you replied YES, which was been assessed?	Both.[1, 2]
Scale construct/ overall structure	The WWTT is a dual task measure of divided attention to examine cognitive-motor interactions, especially in the context of identifying fallers. Subjects are asked to recite the letters of the alphabet while sitting,[2] then to walk 40 feet, then asked to walk 40 feet while reciting the letters of the alphabet aloud (WWT-simple).[1] Subjects recite alternate letters of the alphabet (a, c, e etc.) while walking (WWT-complex task).[1, 2] The time to complete the task is recorded and serves as the test score. NOTE: indices have been calculated based on differential performance in different tasks. Dual-task cost (DTC) - the change in performance under dual-task conditions relative to the single task condition, as well as a Gait DTC for Simple and Complex versions and a Cognitive DTC.[2, 3] **Additional references**: http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=1059
a. Question items	
Items of presence of symptom/sign?	No.[1]
Items of severity of symptom/sign?	No.[1]
b. Response scale	
Are the items of the scale scored in discrete steps	Performance-based test (time in seconds and number of errors recorded).[1]

(specify number) or in a visual analogue scale?	
c. Is the scale easy to score?	
Approx. time to score patient	Less than 1 minute, but 3-10 minutes including instructions to participant and warm-up test. http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=1059
d. Raters	
Patient, caregiver, or clinician	Clinician.
If clinician-rated, is training for application required?	No.
e. Access to scale	
Copyright or public domain?	Public.
How can the scale be obtained (address or website)?	Unknown.
Has the scale been published in other languages?	–
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Not applicable.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Not applicable.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	Not applicable.
Does it score current state or is it based on the patient/caregiver recall?	Current state. http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=1059
What is the time frame (e.g. "during the past week")?	Not applicable. http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=1059
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	Screen for risk of falls.
Is there a cut-off score? (for HD, for non-HD)	Not for HD. In non-HD: ≥33 seconds for WWT complex versions and risk of falls in a non-

	demented community-living elderly: sensitivity of 38.5, specificity of 95.6, positive predictive value of 71.4.[1] http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=1059
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	
Are the questions appropriate for use in an HD population?	Yes, in ambulatory HD patients.[2]
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	No. Applicable in ambulatory HD only.[2]
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Not applicable.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.[2]
a. Reliability	
Assessed – not assessed	Not assessed.
Internal consistency	Not assessed.
Test-retest reliability	Not assessed.
Inter-rater reliability	Not assessed.
b. Validity in HD	
Assessed vs. not assessed	Not assessed.
Criterion validity (any comparison with gold-standard)	_
Construct validity	
Factor analysis	

Convergent validity	<u>Manifest HD, n=32</u>[2] a. Time to complete: • WWTT-simple: correlated with UHDRS-TMS (Spearman's Rho of 0.37 [95%CI -0.01 to 0.66] although CI goes over 0, p<0.05), but not age, gender, or UHDRS-TFC (Spearman's -0.29, n.s.). Slower time to complete the WWTT-simple is reported to be correlated with poorer performance on the Trail Making Tests A and B (Spearman's Rho 0.52, and 0.42), Stroop word and interference, and Symbol Digit Modalities Test (SDMT) (Spearman's -0.42, -0.51, and -0.43) all p<0.05). • WWTT-complex: not correlated with UHDRS-TMS (Spearman's Rho of 0.310, n.s.). Correlated with UHDRS-TFC (Spearman's Rho of -0.618 [95%CI -0.832 to 0.321] p<0.01). Slower time to complete the WWTT-complex is reported to be correlated with poorer performance on both Trails A and Trails B (Spearman's Rho 0.53, and 0.51), as well as poorer performance on the Stroop color (-0.37), word (-0.35), and interference (-0.38) and the SDMT (-0.50), all p<0.05). b. The number of prospective falls was reported to be related to WWTT-simple (r = 0.86; p < 0.001; 95% CI (0.62– 0.96)), and moderately WWTT-complex (r = 0.44; p = 0.058; 95% CI (0.01–0.73).
Divergent validity	<u>Manifest HD, n=35</u> [2] No correlation between WWTT and disease-specific measures in individuals with UHDRS-TMS ≥ 35.[2]
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	–
Generalizability	
Shown to be valid at any stage of HD?	No.[2]
Shown to be valid in any population with dementia or significant cognitive impairment?	Unknown.

Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Not assessed.
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	Unknown.
Score distributions	Manifest HD, n=32[2] **UHDRS-TMS < 35.** Simple (s) - 11.9 (3.3). Complex (s) - 17.1 (8.5). **UHDRS-TMS >=35.** Simple (s) - 14.9 (7.0). Complex (s) - 21.9 (17.9).
IV. Overall impression	
Advantages	This test may be useful in predicting future falls in individuals with HD.
Disadvantages	Sparse data in HD limited to a single study.
V. Recommendation	**Suggested *with caveats*.**

Supplemental references 15

1. Verghese J, Buschke H, Viola L, et al. Validity of divided attention tasks in predicting falls in older individuals: a preliminary study. J Am Geriatr Soc 2002;50:1572-1576.
2. Fritz NE, Hamana K, Kelson M, Rosser A, Busse M, Quinn L. Motor-cognitive dual-task deficits in individuals with early-mid stage Huntington disease. Gait Posture 2016;49:283-289.

Supplemental table 16

Timed 25 Foot Walk Test (T25FW)	
I. Scale description	
Are there several versions of the scale?	No.
If you replied YES, which was been assessed?	Not applicable.
Scale construct/ overall structure	The T25FW is a clinical tool to evaluate patients for quantitative mobility and leg function performance in a timed, 25-foot walk. The patient is directed to walk 25 feet as quickly and as safely as possible from one marked end to the other (a straight distance without turns). The time is calculated from the moment the patient is instructed to begin, until the patient has reached the 25-foot mark. The second test is immediately administered again by having the patient walk the same distance. Patients may use assistive devices while doing this task. This is a single measure of time based on average of two tests http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1204
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	No.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Time in seconds; note timing score is problematic if person cannot walk 25 feet.
c. Is the scale easy to score?	
Approx. time to score patient	< 5 minutes, requires stop watch and markings for 25 feet distance on floor. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1204
d. Raters	

Patient, caregiver, or clinician	Clinician. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1204
If clinician-rated, is training for application required?	No.
e. Access to scale	
Copyright or public domain?	Public domain.
How can the scale be obtained (address or website)?	Not applicable.
Has the scale been published in other languages?	Not applicable.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	N/A
b. Face validity	
Do the items of the scale cover different components of the specific domain?	No.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	No, only a single domain is covered.
Does it score current state or is it based on the patient/caregiver recall?	Current state.
What is the time frame (e.g. "during the past week")?	Not applicable.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To measure severity. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1204
Is there a cut-off score? (for HD, for non-HD)	Not in HD. There are age- and sex-based normative data in healthy adults (age range 20s to 70s): Men comfortable speed ranges 1.39 m/s to 1.33, and women 1.41 to 1.27 m/s and range fast-paces men 2.53 to 2.08 m/s and women 2.47 to 1.74 m/s (.n=230 age 20-70 measure over 6.62 m with acceleration and deceleration period so completely compatible with 10 meter walk - really 25 foot walk).[4]

d. Acceptability	
Is the length of the scale appropriate?	Not applicable.
Are there ambiguities in instructions to patient/rater (as applicable)?	No.
Are there ambiguities in rating anchors?	N/A
Are the questions appropriate for use in an HD population?	The task is appropriate.
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Yes, unless the patient is non-ambulatory.
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	The scale has been deployed in HD.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Not in HD. In non-HD: The American Physiotherapy Task Force Neurology Section, studied the T25WT largely for multiple sclerosis but also made recommendations from Parkinson's, Spinal Cord Injury, Stroke, Traumatic Brain Injury and Vestibular task forces. Website provides the summary of measurement properties related to those reported here http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1204
a. Reliability	
Assessed – not assessed	Not assessed in HD.
Internal consistency	Not assessed in HD.
Test-retest reliability	Not assessed in HD.

	Non-HD: healthy controls; ICC=0.88.[5]
Inter-rater reliability	Not assessed in HD. Non-HD: ICC=0.94 in MS patients; ICC=0.88 in healthy controls.[5]
b. Validity in HD	
Assessed vs. not assessed	Not assessed.
Criterion validity (any comparison with gold-standard)	–
Construct validity	
Factor analysis	Not assessed in HD.
Convergent validity	Not assessed in HD. Non-HD: Excellent correlation with Expanded Disability Status Scale for classifying multiple sclerosis (EDSS) (r = 0.6686; p<0.0001, the T100MW (r = 0.9227; p<0.0001). For patients with limited ambulation, there was an excellent correlation with walking distance (r = - 0.7121; n = 53 MS patients). For patients with restricted ambulation, there was an excellent correlation with walking distance (r = - 0.6861; n = 44 MS patients).[5]
Divergent validity	Not assessed in HD.
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	No data in HD.
Generalizability	
Shown to be valid at any stage of HD?	Unknown.
Shown to be valid in any population with dementia or significant cognitive impairment?	No, but there are no concerns about existing measurement properties (personal judgment).
Responsiveness (detect change over time in the construct)	

Demonstrated to be sensitive to change (change over time or to treatment)?	Withdrawal of tetrabenazine resulted in significant reduction of T25WT scores in a manifest HD cohort (n=10): Mean(SD) off: 8.5 (3.6) and on drug 8.1 (2.1), n.s.. Same trial s.s. change for BBS and DGI. Manifest HD with chorea, n=11:[6] No significant change in response to tetrabenazine (TBZ): ON-TBZ -5.4 ± 1.9 OFF-TBZ - 5.3 ± 1.7. Same trial n.s. change for BBS.
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	Unknown
Score distributions	See data listed in ***"Demonstrated to be sensitive to change (change over time or to treatment)?"*** [6]
IV. Overall impression	
Advantages	Quick and requires little equipment.
Disadvantages	Only for ambulatory patients. Not enough data in HD.
V. Recommendation	**Listed**

Supplemental references 16

1. Verghese J, Buschke H, Viola L, Katz M, Hall C, Kuslansky G, *et al.* Validity of divided attention tasks in predicting falls in older individuals: a preliminary study. *J Am Geriatr Soc* 2002;50:1572-6.
2. Fritz NE, Hamana K, Kelson M, Rosser A, Busse M, Quinn L. Motor-cognitive dual-task deficits in individuals with early-mid stage Huntington disease. *Gait Posture* 2016;49:283-9.
3. Hall CD, Echt KV, Wolf SL, Rogers WA. Cognitive and motor mechanisms underlying older adults' ability to divide attention while walking. *Phys Ther* 2011;91:1039-50.
4. Bohannon RW. Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants. *Age Ageing* 1997;26:15-9.
5. Phan-Ba R, Pace A, Calay P, Grodent P, Douchamps F, Hyde R, *et al.* Comparison of the timed 25-foot and the 100-meter walk as performance measures in multiple sclerosis. *Neurorehabil Neural Repair* 2011;25:672-9.

6. Ferrara JM, Mostile G, Hunter C, Adam OR, Jankovic J. Effect of tetrabenazine on motor function in patients with huntington disease. *Neurol Ther* 2012;1:5.

Supplemental table 17

12-meter walking, hand tapping in 30s, and time to drink 120 mL	
I. Scale description	
Are there several versions of the scale?	No.
If you replied YES, which was been assessed?	Not applicable.
Scale construct/ overall structure	The test assesses bradykinesia through timed measures of walking, hand tapping and drinking water. The test consists of assessing walking 12 meters, hand tapping in 30 seconds and the time to drink 120 ml of water. No details of test administration given.[1, 2]
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	No.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Not applicable. Timed performance measure.
c. Is the scale easy to score?	
Approx. time to score patient	5-10 min (personal judgment).
d. Raters	
Patient, caregiver, or clinician	Clinician.
If clinician-rated, is training for application required?	No.
e. Access to scale	
Copyright or public domain?	Not applicable.
How can the scale be obtained (address or website)?	Not applicable.
Has the scale been published in other languages?	Not applicable.
II. Scale properties	

a. Content validity	
Any process for item generation and/or reduction	Not applicable.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Yes.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	The test covers three separate functional times, but all appear to measure bradykinesia.
Does it score current state or is it based on the patient/caregiver recall?	Current state/performance.
What is the time frame (e.g. "during the past week")?	Not applicable.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To measure severity of bradykinesia/time to complete task.
Is there a cut-off score? (for HD, for non-HD)	No.
d. Acceptability	
Is the length of the scale appropriate?	Yes.
Are there ambiguities in instructions to patient/rater (as applicable)?	Not applicable.
Are there ambiguities in rating anchors?	Not applicable.
Are the questions appropriate for use in an HD population?	Not applicable.
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Likely not applicable in more advanced stages, namely in non-ambulatory HD (personal judgment).
e. Has this scale been specifically developed for use in HD (yes/no)?	Yes.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Unknown.
III. Clinimetric/psychometric properties	

Are there clini- or psychometric properties in HD ?	No.
a. Reliability	
Assessed – not assessed	Not assessed in HD.
Internal consistency	Not assessed in HD.
Test-retest reliability	Not assessed in HD.
Inter-rater reliability	Not assessed in HD.
b. Validity in HD	
Assessed vs. not assessed	Not assessed in HD.
Criterion validity (any comparison with gold-standard)	No.
Construct validity	
Factor analysis	Not assessed in HD.
Convergent validity	Not assessed in HD.
Divergent validity	Not assessed in HD.
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Unable to assess due to lack of data.
Generalizability	
Shown to be valid at any stage of HD?	Yes, except in pre-manifest HD and non-ambulatory HD (personal judgment).
Shown to be valid in any population with dementia or significant cognitive impairment?	No.
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	It has been shown to be sensitive to change in time in a longitudinal study.[1]
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	No.

Score distributions	No.
IV. Overall impression	
Advantages	The different timing measures appear to be sensitive to change in time. It measures bradykinesia across three unique tasks, including drinking, which is unique.
Disadvantages	Unknown if it is sensitive to change secondary to treatment. Lack summary measure for the whole test.
V. Recommendation	**Listed.**

Supplemental references 17

1. Barker RA, Mason SL, Harrower TP, Swain RA, Ho AK, Sahakian BJ*, et al.* The long-term safety and efficacy of bilateral transplantation of human fetal striatal tissue in patients with mild to moderate Huntington's disease. *Journal of Neurology, Neurosurgery & Psychiatry* 2013;84:657-65.
2. Michell AW, Goodman AO, Silva AH, Lazic SE, Morton AJ, Barker RA. Hand tapping: a simple, reproducible, objective marker of motor dysfunction in Huntington's disease. *J Neurol* 2008;255:1145-52.

Supplemental table 18

Jebsen-Taylor Hand Function Test	
I. Scale description	
Are there several versions of the scale?	No.
If you replied YES, which was been assessed?	Not applicable.
Scale construct/ overall structure	Seven-item test designed to provide an objective measure of various aspects of hand function. Measures unilateral hand function. Assesses speed, not quality of performance.[1] Participants are timed performing common functional activities: writing, card turning (simulated page turning), picking up small common objects, simulated feeding, stacking checkers, lifting light cans, and lifting weighted cans (scored as total time to complete tasks, high score=impaired).[2] Time spent to perform each task has also been reported.[3] Maximum time allotted per subtest is 120 seconds. Each item performed with each hand separately – non-dominant hand first.[1]
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	No.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Not applicable. Timed performance measure.
c. Is the scale easy to score?	
Approx. time to score patient	15 minutes.[1]
d. Raters	
Patient, caregiver, or clinician	Clinician.
If clinician-rated, is training for application	No.

required?	
e. Access to scale	
Copyright or public domain?	Copyright.
How can the scale be obtained (address or website)?	A test kit is sold commercially through multiple vendors.[1]
Has the scale been published in other languages?	Not applicable.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Not applicable.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Yes.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	The test covers various functional tasks that cover different aspects of upper extremity function (personal judgment).
Does it score current state or is it based on the patient/caregiver recall?	Current state/performance.
What is the time frame (e.g. "during the past week")?	Not applicable.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	Measure impairment of speed of hand function tasks (severity).[1]
Is there a cut-off score? (for HD, for non-HD)	Not applicable.
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	Not applicable.
Are there ambiguities in rating anchors?	Not applicable.
Are the questions appropriate for use in an HD population?	Not applicable.
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not	Not applicable.

applicable?	
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Not applicable.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	No.
a. Reliability	
Assessed – not assessed	Not assessed in HD.
Internal consistency	Not assessed in HD.
Test-retest reliability	Not assessed in HD.
Inter-rater reliability	Not assessed in HD.
b. Validity in HD	
Assessed vs. not assessed	Not assessed in HD.
Criterion validity (any comparison with gold-standard)	No.
Construct validity	
Factor analysis	Not assessed in HD.
Convergent validity	Not assessed in HD.
Divergent validity	ON- and OFF-tetrabenazine open label study (n=11)[3]: negative correlation between multiple items of the JTHFT and the MoCA score, stronger for dominant hand.
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Unable to assess due to lack of data.
Generalizability	
Shown to be valid at any stage of HD?	No. Used only in symptomatic HD with chorea. [2, 3]
Shown to be valid in any population with dementia or significant cognitive impairment?	No (personal judgment).
Responsiveness (detect change over time in the construct)	

Demonstrated to be sensitive to change (change over time or to treatment)?	ON- and OFF-tetrabenazine open label study (n=11)[3]: no s.s. difference. ON- and OFF-tetrabenazine open label study (n=10)[2] : **dominant hand** – OFF, 131.2 (73.3); ON, 125.0 (57.1), p=0.647; non-**dominant hand** – OFF, 193.1 (111.6); ON, 217.3 (111.3), p=0.285
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	Unknown in HD.
Score distributions	Scores on the JTHFT were globally slower.[3]
IV. Overall impression	
Advantages	Attempts to measure upper limb function in HD.
Disadvantages	Little use in HD.
V. Recommendation	**Listed.**

Supplemental references 18

1. Rehab Measures: Jebsen Hand Function Test. In; 2012.
2. Fekete R, Davidson A, Jankovic J. Clinical assessment of the effect of tetrabenazine on functional scales in huntington disease: a pilot open label study. *Tremor Other Hyperkinet Mov (N Y)* 2012;2.
3. Ferrara JM, Mostile G, Hunter C, Adam OR, Jankovic J. Effect of tetrabenazine on motor function in patients with huntington disease. *Neurol Ther* 2012;1:5.

RATING SCALES

Supplemental table 19

The Unified Huntington's Disease Rating Scale (UHDRS) Total Functional Capacity (TFC)	
I. Scale description	
Are there several versions of the scale?	No.[1]
If you replied YES, which was been assessed?	Not applicable.[1]
Scale construct/ overall structure	The UHDRS-Total Functional Capacity (TFC) is part of a multi-component scale designed originally to prospectively evaluate all patients with HD and individuals at risk for HD. The UHDRS-TFC focuses on assessment of capacity rather than actual performance. UHDRS-TFC is a brief interview involving the patient and a close family member or friend familiar with the patient's functioning. There are 5 items covering basic activities of living: 1) occupation, 2) handling finances, 3) domestic responsibilities, 4) ADLs (eating, dressing, bathing), and 5) level of care (home or facility). The UHDRS-TFC places emphasis on the clinician's judgment and does not require rigorous documentation of performance. Higher scores on the function scales indicate better functioning than lower scores. The Shoulson and Fahn HD Staging system categorizes the total UHDRS-TFC scores in the stages I (11-13), II (7-10), III (3-6), IV (1-2), and V (0).[2]
a. Question items	
Items of presence of symptom/sign?	No.[2]
Items of severity of symptom/sign?	Yes.[2]
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Discrete steps. Variable range (3 to 4 steps).[2]
c. Is the scale easy to score?	

Approx. time to score patient	2-5 min (personal judgment). The full UHDRS will take approx. 30 min.[1]
d. Raters	
Patient, caregiver, or clinician	Clinician (with information from patient and caregiver) (personal judgment).
If clinician-rated, is training for application required?	No (personal judgment).
e. Access to scale	
Copyright or public domain?	Copyright.[2]
How can the scale be obtained (address or website)?	HSG, prior written permission is required. E-mail: info@hsglimited.org
Has the scale been published in other languages?	Yes (Portuguese, French, German, Dutch, Danish, Italian, Polish, Russian, Czech, Norwegian, Swedish).[3]
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Yes. [1] 1) Creation of a single scale based on pre-existing scales: Quantitated neurological exam (QNE), HD functional capacity scale (HDFCS), the HD motor rating scale (HDMRS), the Physical Disability and Independence scales, Marsden and Quinn's chorea severity scale, the HD Activities of Daily Living scale, and other relevant measures. 2) Followed by "several months of pilot experience". 3) Neurologists, psychiatrists, neuropsychologists, and other professionals participated in the drafting of the scale.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Yes (occupation, financial, domestic chores, activities of the daily living, care level).[2]
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	–
Does it score current state or is it based on the patient/caregiver recall?	Current state based on clinical best judgment and patient/caregiver report.
What is the time frame (e.g. "during the past	Current state.

week")?	
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	Measure severity of functional capacity (personal judgment).
Is there a cut-off score? (for HD, for non-HD)	No (personal judgment).
d. Acceptability	
Is the length of the scale appropriate?	Yes (5 items) (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	No (personal judgment).
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Yes. No. (personal judgment).
e. Has this scale been specifically developed for use in HD (yes/no)?	Yes.[2]
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Yes.[2, 4-6]
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.
a. Reliability	
Assessed – not assessed	Assessed in HD.
Internal consistency	Not assessed.
Test-retest reliability	Not assessed.
Inter-rater reliability	Agreement among 7 raters - fully concordant in 27% ratings, within one unit 65% concordance. Spearman's correlation for identical item score. Range: 0.44 - 0.82 (mean 0.62). [7] n=29, UHDRS-TFC modified for assessment of ability and not capacity: HD patient and caregiver - ICC 0.96 (0.92, 0.98).[8]

b. Validity in HD	
Assessed vs. not assessed	Assessed in HD.
Criterion validity (any comparison with gold-standard)	There is no gold-standard (personal judgment).
Construct validity	
Factor analysis	No.
Convergent validity	Manifest HD, n=25: UHDRS-TFC X HD-ADL r=-0.89 p<0.001, vs disease duration r=-0.38 p<0.05, vs MMSE r=0.71 p<0.001.[9] Manifest HD, n=489: UHDRS-TFC X UHDRS-FAS r=0.94 p<0.005, UHDRS-TFC X UHDRS-IS r=0.94 p<0.005. [1] Manifest HD, n=22: UHDRS-TFC X Unspecified 5-point response QoL scale, r=0.54 p<0.05.[10] Manifest HD, n= 70: UHDRS-TFC x SF-36 mental summary score, r=0.42, p=0.000, vs SF-36 physical summary score, r=0.68, p=0.000.[11] Manifest HD, n=30, UHDRS-TFC x Functional reach test r=0.66, p<0.001, UHDRS-TFC x Timed up and go r=-0.68, p<0.001, UHDRS-TFC x Berg Balance Test r=0.60, p<0.01, UHDRS-TFC x UHDRS-TMS r=-0.546 p<0.01.[12] Manifest HD, n=132, *UHDRS-TFC (modified version to capture self-reported functional capacity)*) x HD-PRO-TRIAD r=0.72 p<0.05.[8] Manifest HD, n=18, Association between UHDRS-TFC and MiniBESTest Score: r2=0.45, p=0.0024.[14] Manifest HD, n=69/46 (2 cohorts), UHDRS-TFC x UHDRS-TMS r= -0.87/ -0.83 both p<0.001.[15] Manifest HD, n=82, UHDRS-TFC x UHDRS-TMS, r=-0.08 p<0.005, x several

	cognitive assessments including object recall, word fluency, Stroop, all p<0.0005. UHDRS-TFC x PBA-HD Subscales: apathy r = -0.85, p<0.0001, irritability and depression subscales, p=ns.[16] Manifest HD with UHDRS-TFC≤5, n=53, UHDRS-TFC x UHDRS-FAS r= -0.90 p<0.001, UHDRS-TFC x UHDRS-TMS r=-0.69 p<0.001, UHDRS-TFC x UHDRS - behavioral p=n.s., x UHDRS-TFC x UHDRS cognitive assessment r=0.76, p<0.001.[17]
Divergent validity	-
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Lacks reliability data, despite widespread use (personal judgment).
Generalizability	
Shown to be valid at any stage of HD?	Yes (personal judgment).
Shown to be valid in any population with dementia or significant cognitive impairment?	Potentially with caregiver information (personal judgment).
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Manifest HD, n=22, mean f/u of 27 months. Decline from a mean±SEM of 7.9 ± 0.72 to 4.0 ± 0.67 (p < 0.001), at a rate of 1.8 units per yr. The rate of decline was similar for minimally disabled patients (stages I and II) and for those with more advanced disability (stages III through V).[2] Manifest HD, n=47, mean 2.2 years. The average UHDRS-TFC score changed from 8.2 ± 0.50 units (mean ± SEM) at the initial examination to 5.9 ± 0.51 units at the last examination. Average TFC decline of 1.05 UHDRS-TFC units/yr. Average rate of decline per initial HD stage: 1.6 units/yr (stage I), 0.9 units/yr (stage II), 0.5 units/yr (stage III), 0.3 units/yr (stage IV).[5] Manifest HD, n=26, RCT of baclofen vs. placebo, follow-up 30 months. Change in UHDRS-TFC Units/year (mean±SD): -0.53 ± 0.45 / year (placebo), -0.85 ± 0.64 / year (baclofen).[4]

	Manifest HD, n=50, mean ± SD f/u=4.4 ± 2.9 years, mean ± SD UHDRS-TFC rate of decline per year = 0.5 ± 0.6.[18] Manifest HD, n=129**:** mean ± SD f/u = 3.6 ± 2.3 years, mean ± SD UHDRS-TFC rate decline per year= -0.63 ± 0.75.[19] Manifest HD, n=202, mean ± SD f/u = 8.0 ± 2.4 months, mean ± SD UHDRS-TFC decline in 6 months= -0.3 ± 1.6.[1] Manifest HD, RCT Fluoxetine vs. placebo, n=12/11, f/u=4 months, Mean change ± SD scores of UHDRS-TFC: 0.25 ± 2.7 (Fluoxetine) *vs.* 0.09 ± 2.0 (placebo) p=n.s.[20] Manifest HD, RCT OPC-14117 vs. placebo, n=40/16, f/u=20 weeks, no statistically significant differences.[21] Manifest HD, n=72, UHDRS-TFC decline after one year: 0.56 95% CI: 0.02-1.09, p=0.042. [22] Manifest HD, RCT Lamotrigine vs. placebo, n= 28/27, f/u=30 months. Mean change (± SD) scores of UHDRS-TFC: 1.89 ± 2.46/ 2.11 ± 1.00, n.s..[23] Manifest HD, n=960, mean follow-up= 18.3 months.[24] Mean (SE) UHDRS-TFC decline -0.72(0.04)/yr. Rate of UHDRS-TFC decline in function of symptom's duration: [24] 1. 1.11 (0.16) units/yr for those with 0 to 2 yrs duration. 2. 0.85 (0.09) units/yr for those with 2-5 yrs duration. 3. 0.60 (0.07) units/yr for those with 5-10 yrs duration. 4. 0.66 (0.08) units/yr for those with 10-20 yrs duration. Rate of functional decline for stage I HD: 1.15 (0.09) units/yr; 0.84 (0.08) units/yr for stage II; 0.38 (0.08) units/yr for stage III; 0.06 (0.10) units/yr for stages IV and V.[24] Manifest HD, RCT CoQ10 vs. remacemide vs. combination vs. placebo,

	n=87/86/87/87 f/u=31months. Mean change ± SD scores of UHDRS-TFC: placebo =-2.7 ± 2.3; CoQ10 treated = -2.4 ± 2.2; Non-CoQ10 treated = -2.7 ± 2.3; combination = -2.4 ± 2.1, comparison between arms all n.s..[25] Manifest HD, n=815, mean f/u =2.7 years, UHDRS-TFC decline = -0.73/ yr; 95% CI: 0.67–0.78, p<0.0001.[26] Manifest HD, RCT placebo vs. riluzole 100 mg vs. riluzole 200 mg, n=22/18/23 f/u=8 weeks. Mean change ± SD scores of UHDRS-TFC: - 0.3 ±1.1/ 0.1 ± 0.9/ -0.1 ± 1.4 p=n.s.[27] Manifest HD, RCT placebo vs. minocycline 100 mg vs. minocycline 200 mg, n=23/18/19, f/u= 8 weeks, Mean change ± SD scores of UHDRS-TFC at =+0.04 ±1.26/ -0.22 ± 0.73/ +0.11 ± 0.94, p=n.s.[28] Manifest HD, RCT ethyl-EPA/placebo, n=39/44, f/u=12 months, UHDRS-TFC decline: n.s. between arms.[29] Manifest HD, RCT ethyl-EPA vs. placebo, n=316, f/u = 6 + 6 (open label) months, UHDRS-TFC at 6 months: -0.2 vs. -0.3, p=n.s; UHDRS-TFC at 12 months: -0.6/ -0.4 p=n.s. arm comparison.[30] Manifest HD, RCT placebo/donepezil, n=12/12 1:1 f/u=12w, Median change UHDRS-TFC 0 / 0.5, p=0.07 for difference between arms.[31] Manifest HD, RCT placebo/riluzole,: n=180/357, f/u=3yrs, Mean change ± SD scores of UHDRS-TFC : -4.4 ± 4.1/-4.6 ± 4.2, n.s. (ITT population), n.s. arm comparison.[32] Manifest HD, n=335, f/u=30 months, Mean change ± SD scores of UHDRS-TFC: -2.7 ± 2.3.[33]

	Manifest HD, RCT latrepirdine/placebo, n=46/44, f/u= 90 days, Mean change ± SD scores of UHDRS-TFC: -0.04 ± 0.15/ 0.01 ± 0.15, treatment comparison n.s.[34] **TRACK-HD:** Pre-manifest (pre-HD A and B) and Early manifest (HD1/earlier and HD2/later: a) n=330 f/u=12 months, Mean change scores of UHDRS-TFC compared with controls: HD: -0.73 (HD1: -0.91, HD2: -0.44), preHD: -0.06 (preHDa: -0.07, preHDb: -0.05).[35] b) n=334 f/u=36 months, Mean change scores of UHDRS-TFC compared with controls: HD1: -1.67, HD2: - 1.48, preHDa: - 0.21, preHDb: - 0.07.[36] Manifest HD, RCT citalopram vs. placebo, n=16/15, f/u=17 weeks, Mean ± SEM change scores of UHDRS-TFC at 17 weeks: -0.54 ± 0.46/ -0.06 ± 0.5, n.s. arm comparison.[37] Manifest HD, RCT placebo/ Selisistat 10mg or 100mg, n=19/17/19, f/u 14 days, Mean change scores of UHDRS-TFC n.s. between arms.[38]
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	Ceiling effect for early stage, floor effect for late stage (personal judgment).[24]
Score distributions	–
IV. Overall impression	
Advantages	Widely used scale. Easy and quick to administer.
Disadvantages	More extensive clinimetric data is required, considering purpose proposed for UHDRS.
V. Recommendation	**Suggested for assessing severity of limitation in functional capacity in HD.**

Supplemental references 19

1. Huntington Study Group. Unified Huntington's Disease Rating Scale: reliability and consistency. *Mov Disord* 1996;11:136-42.

2. Shoulson I. Huntington disease: functional capacities in patients treated with neuroleptic and antidepressant drugs. *Neurology* 1981;31:1333-5.
3. Enroll-HD Portal. In; 2016.
4. Shoulson I, Odoroff C, Oakes D, Behr J, Goldblatt D, Caine E*, et al.* A controlled clinical trial of baclofen as protective therapy in early Huntington's disease. *Annals of neurology* 1989;25:252-59.
5. Young AB, Shoulson I, Penney JB, Starosta-Rubinstein S, Gomez F, Travers H*, et al.* Huntington's disease in Venezuela: neurologic features and functional decline. *Neurology* 1986;36:244-9.
6. Bamford KA, Caine ED, Kido DK, Plassche WM, Shoulson I. Clinical-pathologic correlation in Huntington's disease: a neuropsychological and computed tomography study. *Neurology* 1989;39:796-801.
7. Munsat TL. Quantification of neurological deficit : Workshop on quantification in clinical neurology : Papers: Butterworths, 1989.
8. Carlozzi NE, Victorson D, Sung V, Beaumont JL, Cheng W, Gorin B*, et al.* HD-PRO-TRIAD Validation: A Patient-reported Instrument for the Symptom Triad of Huntington's Disease. *Tremor Other Hyperkinet Mov (N Y)* 2014;4:223.
9. Bylsma. Assessment of Adaptive Functioning in Huntington's Disease. *Mov Disord* 1993;8:183-90.
10. Ready RE, Mathews M, Leserman A, Paulsen JS. Patient and caregiver quality of life in Huntington's disease. *Mov Disord* 2008;23:721-6.
11. Ho AK, Gilbert AS, Mason SL, Goodman AO, Barker RA. Health-related quality of life in Huntington's disease: Which factors matter most? *Mov Disord* 2009;24:574-8.
12. Rao AK, Muratori L, Louis ED, Moskowitz CB, Marder KS. Clinical measurement of mobility and balance impairments in Huntington's disease: validity and responsiveness. *Gait Posture* 2009;29:433-6.
13. Jacobs JV, Boyd JT, Hogarth P, Horak FB. Domains and correlates of clinical balance impairment associated with Huntington's disease. *Gait Posture* 2015;41:867-70.
14. Siesling S, Zwinderman AH, van Vugt JP, Kieburtz K, Roos RA. A shortened version of the motor section of the Unified Huntington's Disease Rating Scale. *Mov Disord* 1997;12:229-34.
15. Thompson JC, Snowden JS, Craufurd D, Neary D. Behavior in Huntington's disease: dissociating cognition-based and mood-based changes. *J Neuropsychiatry Clin Neurosci* 2002;14:37-43.
16. Youssov K, Dolbeau G, Maison P, Boisse MF, Cleret de Langavant L, Roos RA*, et al.* Unified Huntington's disease rating scale for advanced patients: validation and follow-up study. *Mov Disord* 2013;28:1717-23.
17. Kieburtz K, MacDonald M, Shih C, Feigin A, Steinberg K, Bordwell K*, et al.* Trinucleotide repeat length and progression of illness in Huntington's disease. *J Med Genet* 1994;31:872-4.
18. Feigin A, Kieburtz K, Bordwell K, Como P, Steinberg K, Sotack J*, et al.* Functional decline in Huntington's disease. *Mov Disord* 1995;10:211-4.

19. Como PG, Rubin AJ, O'Brien CF, Lawler K, Hickey C, Rubin AE, *et al.* A controlled trial of fluoxetine in nondepressed patients with Huntington's disease. *Movement disorders* 1997;12:397-401.
20. Safety and tolerability of the free-radical scavenger OPC-14117 in Huntington's disease. The Huntington Study Group. *Neurology* 1998;50:1366-73.
21. Siesling S, van Vugt JP, Zwinderman KA, Kieburtz K, Roos RA. Unified Huntington's disease rating scale: a follow up. *Mov Disord* 1998;13:915-9.
22. Kremer B, Clark CM, Almqvist EW, Raymond LA, Graf P, Jacova C, *et al.* Influence of lamotrigine on progression of early Huntington disease: a randomized clinical trial. In; 1999. p. 1000-11.
23. Marder K, Zhao H, Myers RH, Cudkowicz M, Kayson E, Kieburtz K, *et al.* Rate of functional decline in Huntington's disease. Huntington Study Group. *Neurology* 2000;54:452-8.
24. Huntington Study Group. A randomized, placebo-controlled trial of coenzyme Q10 and remacemide in Huntington's disease. *Neurology* 2001;57:397-404.
25. Mahant N, McCusker EA, Byth K, Graham S. Huntington's disease: clinical correlates of disability and progression. *Neurology* 2003;61:1085-92.
26. Huntington Study Group. Dosage effects of riluzole in Huntington's disease: A multicenter placebo-controlled study. *Neurology* 2003;61:1551-56.
27. Huntington Study G. Minocycline safety and tolerability in Huntington disease. *Neurology* 2004;63:547-9.
28. Puri BK, Leavitt BR, Hayden MR, Ross CA, Rosenblatt A, Greenamyre JT, *et al.* Ethyl-EPA in Huntington disease: a double-blind, randomized, placebo-controlled trial. In; 2005. p. 286-92.
29. Dorsey ER. Randomized controlled trial of ethyl-eicosapentaenoic acid in huntington disease the trend-hd study. *Archives of neurology* 2008;65:1582-89.
30. Cubo E, Shannon KM, Tracy D, Jaglin JA, Bernard BA, Wuu J, *et al.* Effect of donepezil on motor and cognitive function in Huntington disease. *Neurology* 2006;67:1268-71.
31. Landwehrmeyer GB, Dubois B, De Yebenes JG, Kremer B, Gaus W, Kraus PH, *et al.* Riluzole in Huntington's disease: A 3-year, randomized controlled study. *Annals of Neurology* 2007;62:262-72.
32. Ravina B, Romer M, Constantinescu R, Biglan K, Brocht A, Kieburtz K, *et al.* The relationship between CAG repeat length and clinical progression in Huntington's disease. *Mov Disord* 2008;23:1223-7.
33. Kieburtz K, McDermott MP, Voss TS, Corey-Bloom J, Deuel LM, Dorsey ER, *et al.* A randomized, placebo-controlled trial of latrepirdine in Huntington disease. In; 2010. p. 154-60.
34. Tabrizi SJ, Scahill RI, Durr A, Roos RAC, Leavitt BR, Jones R, *et al.* Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: the 12-month longitudinal analysis. *The Lancet Neurology* 2011;10:31-42.

35. Tabrizi SJ, Scahill RI, Owen G, Durr A, Leavitt BR, Roos RA, *et al.* Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. *The Lancet Neurology* 2013;12:637-49.
36. Beglinger LJ, Adams WH, Langbehn D, Fiedorowicz JG, Jorge R, Biglan K, *et al.* Results of the citalopram to enhance cognition in Huntington disease trial. *Mov Disord* 2014;29:401-5.
37. Sussmuth SD, Haider S, Landwehrmeyer GB, Farmer R, Frost C, Tripepi G, *et al.* An exploratory double-blind, randomized clinical trial with selisistat, a SirT1 inhibitor, in patients with Huntington's disease. *British journal of clinical pharmacology* 2015;79:465-76.

Supplemental table 20

The Unified Huntington's Disease Rating Scale (UHDRS) Functional Assessment Scale	
I. Scale description	
Are there several versions of the scale?	No[1]
If you replied YES, which was been assessed?	N/A
Scale construct/ overall structure	The UHDRS-FAS is part of a multi-component scale originally designed to prospectively evaluate all patients with HD as well as those at risk for HD. It consists of 25 questions which screen capacity to complete the tasks mentioned in the assessment. It is considered an extension of the Total Functional Capacity and is more detailed in certain tasks.[1] The checklist is summed by giving a score of 1 to all "yes" replies. A higher score indicates better functioning than a lower score.[1]
a. Question items	
Items of presence of symptom/sign?	Yes.[1]
Items of severity of symptom/sign?	No.[1]
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Yes (2, Y/N).[1]
c. Is the scale easy to score?	
Approx. time to score patient	5 -10 min (personal judgment).
d. Raters	
Patient, caregiver, or clinician	Clinician (in the presence of a family or friend to get the clinician's best judgment based on both responses).[1]
If clinician-rated, is training for application required?	Written instructions (personal judgment).
e. Access to scale	
Copyright or public domain?	Copyright.
How can the scale be obtained (address or website)?	Huntington Study Group (HSG), prior written permission is required. E-mail: info@hsglimited.org

Has the scale been published in other languages?	Yes (Portuguese, French, German, Dutch, Danish, Italian, Polish, Russian, Czech, Norwegian, Swedish).[2]
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Yes. 1) Creation of a single scale based on pre-existing scales: Quantitated neurological exam (QNE), HD functional capacity scale (HDFCS), the HD motor rating scale (HDMRS), the Physical Disability and Independence scales, Marsden and Quinn's chorea severity scale, the HD Activities of Daily Living scale, and other relevant measures. 2) Followed by "several months of pilot experience". 3) Neurologists, psychiatrists, neuropsychologists, and other professionals participated in the drafting of the scale.[1]
b. Face validity	
Do the items of the scale cover different components of the specific domain?	No (personal judgment).
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	N/A.
Does it score current state or is it based on the patient/caregiver recall?	Based on clinician's impression with input from patient/caregiver (personal judgment).
What is the time frame (e.g. "during the past week")?	N/A.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	Severity, longitudinal measurement (personal judgment).
Is there a cut-off score? (for HD, for non-HD)	No (personal judgment).
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).

Are there ambiguities in rating anchors?	No (personal judgment).
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Yes (personal judgment).
e. Has this scale been specifically developed for use in HD (yes/no)?	Yes.[1]
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Yes.[1, 3-17]
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.[1]
a. Reliability	
Assessed – not assessed	Assessed in HD.[1]
Internal consistency	Manifest HD, n=489: Cronbach's alpha 0.95.[1]
Test-retest reliability	—
Inter-rater reliability	—
b. Validity	
Assessed vs. not assessed	Assessed in HD.
Criterion validity (any comparison with gold-standard)	There is no gold-standard (personal judgment).
Construct validity	
Factor analysis	—
Convergent validity	Manifest HD, n=489: [1] UHDRS-FA X UHDRS-TFC r=0.94 p<0.005, X UHDRS-TMS r=-0.75 p<0.005, X Verbal fluency r=0.59 p<0.005, X Symbol Digit r=0.65 p<0.005, X Stroop word r=0.60 p<0.005, X Stroop color word r=0.61 p<0.005, X UHDRS-behavior total r=-0.07 p=n.s., X subscale mood r=0.06 r=n.s., X subscale behavior r=-0.13 p=n.s., X subscale psychosis r=-0.14 p<0.005. Manifest HD, n=69/46 (2 cohorts, Dutch/US)[18]: UHDRS-FA X UHDRS-TMS r=0.88

	p<0.001/r=0.83 p<0.001. Manifest HD, n=21, UHDRS-FA X UHDRS-TMS r=-0.686 p<0.001.[19] Manifest HD, n=80, UHDRS-FA X SF-36 r=0.46 p<0.05, X CBI r=-0.56 p<0.05, X UHDRS-TMS r=-0.82 p<0.05, X UHDRS cognitive r=0.76 p<0.05, X HAM-D r=-0.43 p<0.05, X UHDRS behavioral r=-0.35 p<0.05, X UHDRS apathy r=-0.47 p<0.05, X UHDRS psychotic symptoms r=-0.25 p=n.s., X UHDRS anxiety r=-0.20 p=n.s., X UHDRS irritability r=0.2 p=n.s., X UHDRS aggression -0.19 p=n.s.[20] Manifest HD, n=48[21], UHDRS-FA X UHDRS-TFC r=-0.9 p<0.001, X UHDRS-IS r=-0.91 p<0.001, X UHDRS-TMS r=0.77 p<0.001, X UHDRS behavior r=-0.10 p=0.47, X UHDRS cognitive r=-0.85 p<0.001, X UHDRS-FAP motor r=0.90 p<0.001, X UHDRS-FAP behavioral r=0.00 p=0.97, X UHDRS-FAP somatic r=0.71 p<0.001, X UHDRS-FAP cognitive r=-0.71 p<0.001.
Divergent validity	
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	More extensive clinimetric analyses are required (personal judgment).
Generalizability	
Shown to be valid at any stage of HD?	Yes (personal judgment).
Shown to be valid in any population with dementia or significant cognitive impairment?	Yes (personal judgment).
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Manifest HD, n=171, f/u=6 m, Mean change ± SD scores UHDRS-FA at 6m= - 0.9 ± 3.0.[1] Manifest HD, n=71, f/u=1.07 year (SD 0.38), UHDRS-FA at last follow-up= -1.5, 95% CI: 0.76, 2.33, p<0.0001.[3] Manifest HD, RCT CoQ10 vs. remacemide vs. combination vs. placebo,

	n=87/86/87/87, f/u=31months, Mean change ± SD scores of UHDRS-FA: placebo -4.0 ± 4.5; Q10 -3.1 ± 3.6; Remacemide -4.3 ± 4.5; Combination -3.4 ± 4.0. [4] Manifest HD, RCT placebo vs. riluzole 100 mg vs. riluzole 200 mg, n=22/18/23, f/u=8 weeks, mean change ± SD scores of UHDRS-FA =-0.8 ±1.3 / -0.2 ± 1.2/ -0.2 ±1.8 p=0.50.[6] Manifest HD, n=815, mean f/u =2.7 yrs, estimated rate of progression (points/year)for UHDRS-FA: -1.4, 95% CI: 1.3 - 1.6. [17] Manifest HD, RCT placebo vs. minocycline 100 mg vs. minocycline 200 mg, n=23/18/19, f/u= 8 weeks, Mean change ± SD scores of UHDRS-FA: -0.30 ± 1.69/ -0.39 ± 1.14/ 0.58 ± 1.95. [7] Manifest HD, RCT ethyl-EPA/placebo, n=39/44, f/u=12 months, UHDRS- FA decline= n.s. between arms.[8] Manifest HD, RCT placebo/donepezil, n=12/12 1:1 f/u=12w, Median change UHDRS-FA 0 / +0.5, p=0.07 for difference between arms.[9] Manifest HD, RCT placebo/TBZ, n=30/54 f/u=9w, Mean change ± SD scores of UHDRS-FA: +0.4 ± 0.4 / -0.8 ± 0.3, p=0.02.[10] Manifest HD, RCT placebo/riluzole, n=128/251 (PPT population), f/u=3yrs, UHDRS-FA at 3yrs=-3.6 ± 4.2/-3.3 ± 3.7.[22] Manifest HD, n=335, f/u=30 months, Mean change ± SD scores of UHDRS-FA: -4.0 ± 4.5.[12] Manifest HD, RCT minocycline/placebo, n=87/27, f/u=18m, Mean change ± SD scores of UHDRS-FA at 18m = -2.4 ± 4.04.[13]

	Manifest HD, RCT latrepirdine/placebo, n=46/44, f/u= 90 days, Mean change ± SD scores of UHDRS-FA at 90 d: 0.01 ± 0.25/ 0.11 ± 0.26, p=0.79.[14] Manifest HD, RCT placebo/ Selisistat 10mg or 100mg, n=19/17/19, f/u 14 days, [16] UHDRS-FA at 14 d vs baseline: -0.05/-0.12/-0.21. Treatment comparison n.s.
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	—
Score distributions	—
IV. Overall impression	
Advantages	Used extensively.
Disadvantages	More extensive clinimetric analyses are necessary.
V. Recommendation	**Suggested for assessing severity of limitation in functional capacity in HD.**

Supplemental references 20

1. Huntington Study Group. Unified Huntington's Disease Rating Scale: reliability and consistency. *Mov Disord* 1996;11:136-42.
2. Enroll-HD Portal. In; 2016.
3. Siesling S, van Vugt JP, Zwinderman KA, Kieburtz K, Roos RA. Unified Huntington's disease rating scale: a follow up. *Mov Disord* 1998;13:915-9.
4. Huntington Study Group. A randomized, placebo-controlled trial of coenzyme Q10 and remacemide in Huntington's disease. *Neurology* 2001;57:397-404.
5. Vaddadi KS, Soosai E, Chiu E, Dingjan P. A randomised, placebo-controlled, double blind study of treatment of Huntington's disease with unsaturated fatty acids. In; 2002. p. 29-33.
6. Huntington Study Group. Dosage effects of riluzole in Huntington's disease: A multicenter placebo-controlled study. *Neurology* 2003;61:1551-56.
7. Huntington Study G. Minocycline safety and tolerability in Huntington disease. *Neurology* 2004;63:547-9.

8. Puri BK, Leavitt BR, Hayden MR, Ross CA, Rosenblatt A, Greenamyre JT, *et al.* Ethyl-EPA in Huntington disease: a double-blind, randomized, placebo-controlled trial. In; 2005. p. 286-92.
9. Cubo E, Shannon KM, Tracy D, Jaglin JA, Bernard BA, Wuu J, *et al.* Effect of donepezil on motor and cognitive function in Huntington disease. *Neurology* 2006;67:1268-71.
10. Huntington Study G. Tetrabenazine as antichorea therapy in Huntington disease: a randomized controlled trial. *Neurology* 2006;66:366-72.
11. Landwehrmeyer GB, Dubois B, De Yebenes JG, Kremer B, Gaus W, Kraus PH, *et al.* Riluzole in Huntington's disease: A 3-year, randomized controlled study. *Annals of Neurology* 2007;62:262-72.
12. Ravina B, Romer M, Constantinescu R, Biglan K, Brocht A, Kieburtz K, *et al.* The relationship between CAG repeat length and clinical progression in Huntington's disease. *Mov Disord* 2008;23:1223-7.
13. Huntington Study Group DI. A futility study of minocycline in Huntington's disease. *Movement Disorders* 2010;25:2219-24.
14. Kieburtz K, McDermott MP, Voss TS, Corey-Bloom J, Deuel LM, Dorsey ER, *et al.* A randomized, placebo-controlled trial of latrepirdine in Huntington disease. In; 2010. p. 154-60.
15. Verbessem P, Lemiere J, Eijnde BO, Swinnen S, Vanhees L, Leemputte M, *et al.* Creatine supplementation in Huntington's disease: a placebo-controlled pilot trial. In; 2003. p. 925-30.
16. Sussmuth SD, Haider S, Landwehrmeyer GB, Farmer R, Frost C, Tripepi G, *et al.* An exploratory double-blind, randomized clinical trial with selisistat, a SirT1 inhibitor, in patients with Huntington's disease. *British journal of clinical pharmacology* 2015;79:465-76.
17. Mahant N, McCusker EA, Byth K, Graham S. Huntington's disease: clinical correlates of disability and progression. *Neurology* 2003;61:1085-92.
18. Siesling S, Zwinderman AH, van Vugt JP, Kieburtz K, Roos RA. A shortened version of the motor section of the Unified Huntington's Disease Rating Scale. *Mov Disord* 1997;12:229-34.
19. Tumas V, Camargos ST, Jalali PS, Galesso Ade P, Marques Jr W. Internal consistency of a Brazilian version of the unified Huntington's disease rating scale. *Arq Neuropsiquiatr* 2004;62:977-82.
20. Banaszkiewicz K, Sitek EJ, Rudzinska M, Soltan W, Slawek J, Szczudlik A. Huntington's disease from the patient, caregiver and physician's perspectives: three sides of the same coin? *J Neural Transm* 2012;119:1361-5.
21. Youssov K, Dolbeau G, Maison P, Boisse MF, Cleret de Langavant L, Roos RA, *et al.* Unified Huntington's disease rating scale for advanced patients: validation and follow-up study. *Mov Disord* 2013;28:1717-23.
22. Landwehrmeyer GB, Dubois B, de Yebenes JG, Kremer B, Gaus W, Kraus PH, *et al.* Riluzole in Huntington's disease: a 3-year, randomized controlled study. *Ann Neurol* 2007;62:262-72.

Supplemental table 21

The Unified Huntington's Disease Rating Scale (UHDRS)- Independence Scale (IS)	
I. Scale description	
Are there several versions of the scale?	No.[1]
If you replied YES, which was been assessed?	N/A.
Scale construct/ overall structure	The UHDRS- IS is part of a multi-component scale originally designed to prospectively evaluate all patients with HD and at risk for HD. The UHDRS-IS assesses functional disability.[1] The UHDRS-IS is a useful clinical tool to follow progression of functional disability. It covers a wide range of functioning. The scale is rated from 100 (no special care needed) to 0 (tube-fed, total bed care); descriptors are provided to gauge function levels at every 10 points (personal judgment).
a. Question items	
Items of presence of symptom/sign?	Yes (personal judgment).
Items of severity of symptom/sign?	Yes (personal judgment).
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Yes, from 100 to 10.[1]
c. Is the scale easy to score?	
Approx. time to score patient	5 min (personal judgment).
d. Raters	
Patient, caregiver, or clinician	Clinician.[1]
If clinician-rated, is training for application required?	Written instructions (personal judgment).
e. Access to scale	
Copyright or public domain?	Copyright.
How can the scale be obtained (address or website)?	Myers 1985.[1]
Has the scale been published in other languages?	Yes (Portuguese, French, German, Dutch, Danish, Italian, Polish, Russian, Czech,

	Norwegian, Swedish).[2]
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Yes. 1) Creation of a single scale based on pre-existing scales: Quantitated neurological exam (QNE), HD functional capacity scale (HDFCS), the HD motor rating scale (HDMRS), the Physical Disability and Independence scales, Marsden and Quinn's chorea severity scale, the HD Activities of Daily Living scale, and other relevant measures. 2) Followed by "several months of pilot experience". 3) Neurologists, psychiatrists, neuropsychologists, and other professionals participated in the drafting of the scale.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Yes (personal judgment).
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	No (personal judgment).
Does it score current state or is it based on the patient/caregiver recall?	Based on clinician's impression with input from patient/caregiver (personal judgment).
What is the time frame (e.g. "during the past week")?	N/A.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	Severity (personal judgment).
Is there a cut-off score? (for HD, for non-HD)	No (personal judgment).
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	No (personal judgment).

Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Yes (personal judgment).
e. Has this scale been specifically developed for use in HD (yes/no)?	Yes.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Yes.[3-7]
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.[1, 3, 4, 8-13]
a. Reliability	
Assessed – not assessed	Assessed in HD.[8]
Internal consistency	—
Test-retest reliability	—
Inter-rater reliability	Patient (n=132) vs. carers (n=40): ICC - 0.71 (0.48, 0.85).[8] **NOTE: modified version**
b. Validity in HD	
Assessed vs. not assessed	Assessed in HD.[1, 3, 4, 8-13]
Criterion validity (any comparison with gold-standard)	There is no gold standard (personal judgment).
Construct validity	–
Factor analysis	–
Convergent validity	Manifest HD, n=489, UHDRS-IS X UHDRS-TFC r=0.92 p<0.005, X UHDRS-FA r=0.90 p<0.005, X UHDRS-TMS r=-0.75 p<0.005, X visual fluency r=0.58 p<0.005, X Sym Digit r=0.63 p=<0.005, X Stroop word r=0.62 p<0.005, X Stroop color r=0.63 p<0.005, X Stroop color word r=0.54 p<0.005, X UHDRS behavior r=-0.05 p=ns, X UHDRS sub mood r=0.09 p=ns, X UHDRS sub behavior r-0.14 p<0.005, X UHDRS sub psychosis r=-0.14 p<0.005.[3] Manifest HD, n=69/46, UHDRS-IS X UHDRS-TMS r=-0.88 p<0.001/-0.91 p<0.001.[14]

	Manifest HD, n=21, UHDRS-IS X UHDRS-TMS r=-0.745 p<0.001.[10] 65 HD patients and 56 carers, UHDRS-IS correlated significantly with the majority SF-36 and SIP sub-items.[11] Manifest HD, n=53, UHDRS-IS X UHDRS-TFC r=0.86 p<0.001, X UHDRS-FA r=-0.91 p<0.001, X UHDRS-TMS r=0.77 p<0.001, X UHDRS behavior r=-0.10 p=0.47, X UHDRS cognitive r=-0.85 p<0.001, X UHDRS-FAP behavior r=0.02 p=0.86, UHDRS-FAP r=-0.88 p<0.001, X UHDRS-FAP somatic r=-0.70 p<0.001, X UHDRS-FAP cognitive r=0.75 p<0.001.[13]
Divergent validity	—
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Good, but needs further clinimetric evaluation.
Generalizability	
Shown to be valid at any stage of HD?	Yes. n=960, mean f/u=18.3 months, rate of IS decline per disease duration: 0 to 2 years, 5.70 units/year (SE 0.76). 2 to 5 years, 4.87 units/year (SE 0.44). 5 to 10 years, 4.08 units/year (SE 0.37). 10 to 20 years, 4.50 units/year (SE 0.48).[5]
Shown to be valid in any population with dementia or significant cognitive impairment?	Yes (personal judgment).
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Manifest HD, n=78, mean±SD f/u =1.07 y ± 0.38, Mean change score UHDRS-IS at 1yr = -3.60.[7] Manifest HD, n=960 f/u mean=18.3m ± 9.7, UHDRS-IS at 1yr = -4.52 SE 0.23.[5] Manifest HD, RCT CoQ10 vs. remacemide vs. combination vs. placebo,

	n=73/76/66/63, f/u=30m, Mean change ± SD UHDRS-IS: placebo -12.2 ± 11.0; Q10 -10.0 ± 10.6; Remacemide -11.4 ± 10.7; combination -9.4 ± 10.2.[15] Manifest HD, RCT placebo vs. riluzole 100 mg vs. riluzole 200 mg, n=22/18/23, placebo/riluzole 100mg/200mg f/u=8w, mean change ± SD scores of UHDRS-IS at 8w = -3.0 ± 5.9/-4.2 ± 5.5/-1.5 ± 4.6.[6] Manifest HD, n=815, mean f/u =2.7 yrs, estimated rate of progression (points/year)for UHDRS-IS: -4.3, 95% CI: 4.0 - 4.6.[16] Manifest HD, RCT placebo vs. minocycline 100 mg vs. minocycline 200 mg, n=23/18/19, f/u= 8 weeks, Mean change ± SD scores of UHDRS-IS at 8w=-0.30 ± 1.69/-0.38 ± 1.14/+0.58 ± 1.95.[17] Manifest HD, RCT ethyl-EPA/placebo, n=39/44, f/u=12 months, UHDRS-IS at 12m=-1.78 /-2.58, n.s. difference between arms.[18] Manifest HD, RCT placebo/donepezil, n=12/12 1:1 f/u=12w, Median change UHDRS-IS at 12w=0/+0.5.[19] Manifest HD, RCT placebo/riluzole, n=128/251 (PPT population), f/u=3yrs, UHDRS-IS at 3yrs=-11.7 ± 11.7 / -9.9 ± 10.2.[20] Manifest HD, n=335, f/u=30 months, Mean change ± SD scores of UHDRS-IS = -11.3 ± 10.8.[21] Manifest HD, n=158/158 ethyl-EPA/placebo f/u=12m, UHDRS-IS at 6m=-1.2/-1.8 p=0.50; UHDRS-IS at 12m=-3.5/-2.8 p=0.50.[22] Manifest HD, RCT minocycline/placebo, n=87/27, f/u=18m, Mean change ± SD scores of UHDRS-IS at 18m=-8.81 ± 10.77. Data not available for the placebo arm.[23]

	Manifest HD, RCT latrepirdine/placebo, n=46/44, f/u= 90 days, Mean change ± SD scores of UHDRS-IS at 90=-0.48 ± 0.77/-0.58 ± 0.78, p=0.93.[24] Manifest HD, RCT placebo/ Selisistat 10mg or 100mg, n=19/17/19, f/u 14 days, UHDRS-IS at 14d=0.27/-0.59/0.79. [25]
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No (personal judgment).
Floor and ceiling effects	There is a ceiling effect for presymptomatic HD (personal judgment).
Score distributions	—
IV. Overall impression	
Advantages	Quick and easy to apply (personal judgment).
Disadvantages	Lack of clinimetric validation (personal judgment).
V. Recommendation	**Suggested for assessing severity of limitation in functional ability in HD..**

Supplemental references 21

1. Myers RH, Sax DS, Schoenfeld M, Bird ED, Wolf PA, Vonsattel JP, *et al.* Late onset of Huntington's disease. *J Neurol Neurosurg Psychiatry* 1985;48:530-4.
2. Enroll-HD Portal. In; 2016.
3. Huntington Study Group. Unified Huntington's Disease Rating Scale: reliability and consistency. *Mov Disord* 1996;11:136-42.
4. Siesling S, van Vugt JP, Zwinderman KA, Kieburtz K, Roos RA. Unified Huntington's disease rating scale: a follow up. *Mov Disord* 1998;13:915-9.
5. Marder K, Zhao H, Myers RH, Cudkowicz M, Kayson E, Kieburtz K, *et al.* Rate of functional decline in Huntington's disease. Huntington Study Group. *Neurology* 2000;54:452-8.
6. Huntington Study Group. Dosage effects of riluzole in Huntington's disease: A multicenter placebo-controlled study. *Neurology* 2003;61:1551-56.
7. Safety and tolerability of the free-radical scavenger OPC-14117 in Huntington's disease. The Huntington Study Group. *Neurology* 1998;50:1366-73.

8. Carlozzi NE, Tulsky DS, Chiaravalloti ND, Beaumont JL, Weintraub S, Conway K, *et al.* NIH Toolbox Cognitive Battery (NIHTB-CB): the NIHTB Pattern Comparison Processing Speed Test. *J Int Neuropsychol Soc* 2014;20:630-41.
9. Myers RH, Sax DS, Koroshetz WJ, Mastromauro C, Cupples LA, Kiely DK, *et al.* Factors associated with slow progression in Huntington's disease. *Arch Neurol* 1991;48:800-4.
10. Tumas V, Camargos ST, Jalali PS, Galesso Ade P, Marques Jr W. Internal consistency of a Brazilian version of the unified Huntington's disease rating scale. *Arq Neuropsiquiatr* 2004;62:977-82.
11. Ho AK, Robbins AO, Walters SJ, Kaptoge S, Sahakian BJ, Barker RA. Health-related quality of life in Huntington's disease: a comparison of two generic instruments, SF-36 and SIP. *Mov Disord* 2004;19:1341-8.
12. Ho AK, Robbins AO, Barker RA. Huntington's disease patients have selective problems with insight. *Mov Disord* 2006;21:385-9.
13. Youssov K, Dolbeau G, Maison P, Boisse MF, Cleret de Langavant L, Roos RA, *et al.* Unified Huntington's disease rating scale for advanced patients: validation and follow-up study. *Mov Disord* 2013;28:1717-23.
14. Siesling S, Zwinderman AH, van Vugt JP, Kieburtz K, Roos RA. A shortened version of the motor section of the Unified Huntington's Disease Rating Scale. *Mov Disord* 1997;12:229-34.
15. Huntington Study Group. A randomized, placebo-controlled trial of coenzyme Q10 and remacemide in Huntington's disease. *Neurology* 2001;57:397-404.
16. Mahant N, McCusker EA, Byth K, Graham S. Huntington's disease: clinical correlates of disability and progression. *Neurology* 2003;61:1085-92.
17. Huntington Study G. Minocycline safety and tolerability in Huntington disease. *Neurology* 2004;63:547-9.
18. Puri BK, Leavitt BR, Hayden MR, Ross CA, Rosenblatt A, Greenamyre JT, *et al.* Ethyl-EPA in Huntington disease: a double-blind, randomized, placebo-controlled trial. In; 2005. p. 286-92.
19. Cubo E, Shannon KM, Tracy D, Jaglin JA, Bernard BA, Wuu J, *et al.* Effect of donepezil on motor and cognitive function in Huntington disease. *Neurology* 2006;67:1268-71.
20. Landwehrmeyer GB, Dubois B, De Yebenes JG, Kremer B, Gaus W, Kraus PH, *et al.* Riluzole in Huntington's disease: A 3-year, randomized controlled study. *Annals of Neurology* 2007;62:262-72.
21. Ravina B, Romer M, Constantinescu R, Biglan K, Brocht A, Kieburtz K, *et al.* The relationship between CAG repeat length and clinical progression in Huntington's disease. *Mov Disord* 2008;23:1223-7.
22. Randomized controlled trial of ethyl-eicosapentaenoic acid in Huntington disease: the TREND-HD study. *Arch Neurol* 2008;65:1582-9.
23. Huntington Study Group DI. A futility study of minocycline in Huntington's disease. *Movement Disorders* 2010;25:2219-24.
24. Kieburtz K, McDermott MP, Voss TS, Corey-Bloom J, Deuel LM, Dorsey ER, *et al.* A randomized, placebo-controlled trial of latrepirdine in Huntington disease. In; 2010. p. 154-60.
25. Sussmuth SD, Haider S, Landwehrmeyer GB, Farmer R, Frost C, Tripepi G, *et al.* An exploratory double-blind, randomized clinical trial with selisistat, a SirT1 inhibitor, in patients with Huntington's disease. *British journal of clinical pharmacology* 2015;79:465-76.

Supplemental table 22

Rivermead Mobility Index	
I. Scale description	
Are there several versions of the scale?	No.
If you replied YES, which was been assessed?	—
Scale construct/ overall structure	The Rivermead Mobility Index (RMI) consists of 14-self-reported items about a patient's ability to perform a wide range of activities, from turning over in bed to running, and 1 direct observation item (standing for 10 seconds without any aid).[1] http://www.rehabmeasures.org/lists/rehabmeasures/dispform.aspx?id=926
a. Question items	
Items of presence of symptom/sign?	–
Items of severity of symptom/sign?	–
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Discrete steps (0 or 1). Scores are reported as either unable or able (0–1) and added to produce a total score (0–15). A higher score reflects better mobility.
c. Is the scale easy to score?	
Approx. time to score patient	5 minutes. http://www.rehabmeasures.org/lists/rehabmeasures/dispform.aspx?id=926
d. Raters	
Patient, caregiver, or clinician	Clinician and patient.
If clinician-rated, is training for application required?	No.
e. Access to scale	
Copyright or public domain?	Provided courtesy of Dr. Derick Wade and the Oxford Centre for Enablement. http://www.rehabmeasures.org/lists/rehabmeasures/dispform.aspx?id=926
How can the scale be obtained (address or website)?	http://www.rehabmeasures.org/lists/rehabmeasures/dispform.aspx?id=926

Has the scale been published in other languages?	No.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	—
b. Face validity	
Do the items of the scale cover different components of the specific domain?	—
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	—
Does it score current state or is it based on the patient/caregiver recall?	Both (personal judgment).
What is the time frame (e.g. "during the past week")?	Present time (undefined)
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To measure severity.
Is there a cut-off score? (for HD, for non-HD)	No.
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	No (personal judgment).
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Yes (personal judgment).

e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.
a. Reliability	
Assessed – not assessed	Assessed in HD.
Internal consistency	Not assessed.
Test-retest reliability	ICC (pre-manifest HD, n=11): 0.81; ICC (manifest HD, n=62) : 0.94. NOTE: consistent across stages.[2]
Inter-rater reliability	Not assessed in HD.
b. Validity in HD	
Assessed vs. not assessed	Not assessed in HD.
Criterion validity (any comparison with gold-standard)	
Construct validity	
Factor analysis	Not assessed.
Convergent validity	–
Divergent validity	–
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	–
Generalizability	
Shown to be valid at any stage of HD?	No. Little differentiation across stages.[2]
Shown to be valid in any population with dementia or significant cognitive impairment?	No.

Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	No.
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No. The Minimum Detectable Change (MDC) has been determined: MDC=1 in premanifest (n=11), MDC=2 in manifest HD.[2]
Floor and ceiling effects	Ceiling effect in pre-manifest HD.
Score distributions	Premanifest HD: 15 (0.5), range: [14–15], n=11.[2] Manifest HD:13 (2), range: [13–14], n=64.[2]
IV. Overall impression	
Advantages	Quick and easy to administer.
Disadvantages	Very limited development in HD Ceiling effect in early stages of HD
V. Recommendation	**Suggested for assessing severity of mobility restriction (as a generic measure)**

Supplemental references 22

1. Collen FM, Wade DT, Robb GF, Bradshaw CM. The Rivermead Mobility Index: a further development of the Rivermead Motor Assessment. *Int Disabil Stud* 1991;13:50-4.
2. Quinn L, Khalil H, Dawes H, Fritz NE, Kegelmeyer D, Kloos AD*, et al.* Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. *Phys Ther* 2013;93:942-56.

Supplemental table 23

Activity-specific balance scale (ABC)	
I. Scale description	
Are there several versions of the scale?	Yes.[1] ABC, modified version for UK (ABC-UK).
If you replied YES, which was been assessed?	Primarily, ABC.[1]
Scale construct/ overall structure	The ABC Scale measures confidence and fear of falling and has proven reliability and valid in the elderly and with some neurological populations. Individuals rate their balance confidence from 0 to 100 in each of 16 tasks; higher scores indicate greater confidence and lower fall risk.
a. Question items	
Items of presence of symptom/sign?	–
Items of severity of symptom/sign?	–
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Visual analogue scale from 0 to 100.
c. Is the scale easy to score?	
Approx. time to score patient	6-30 minutes. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=949
d. Raters	
Patient, caregiver, or clinician	Can be self-administered, face-to-face interview is recommended.[1]
If clinician-rated, is training for application required?	No.
e. Access to scale	
Copyright or public domain?	Public domain.
How can the scale be obtained (address or website)?	http://www.exercisepd.com/uploads/3/5/3/1/3531021/activities_specific_balance_scale_nov_5_2012.pdf
Has the scale been published in other languages?	Yes.

II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Items for the newly developed 16-item ABC Scale were generated by 15 clinicians and 12 elderly outpatients. Psychometric testing involved 60 community seniors (aged 65-95) self-classified as either high or low in mobility confidence according to their perceived need for a walking aid and personal assistance to ambulate outdoors.[1]
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Yes.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	No.
Does it score current state or is it based on the patient/caregiver recall?	Patient self-assessment.
What is the time frame (e.g. "during the past week")?	
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To measure severity of falls risk.
Is there a cut-off score? (for HD, for non-HD)	Not in HD. Cut-off scores have been established in Parkinson's disease (69%, with 93% sensitivity and 69% specificity)[2] and stroke (81.1%)[3] patients.
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	No (personal judgment).
Are the questions appropriate for use in an HD population?	Mostly (personal judgment).
Is the scale applicable across HD disease	N/A for non-ambulatory. (personal judgment).

stages? Are there HD stages in which the scale is not applicable?	
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	–
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	
a. Reliability	
Assessed – not assessed	Assessed.
Internal consistency	Not assessed in HD. NOTE: found to have good internal consistency in older people.[1]
Test-retest reliability	Manifest HD, n = 20, ICC = 0.74, 95% CI: 0.58, 1.0.[4]
Inter-rater reliability	Not assessed.
b. Validity in HD	
Assessed vs. not assessed	Assessed.
Overall impression: good – not good	Not good (when compared to TMT and four square step test) (personal judgment).
Criterion validity (any comparison with gold-standard)	
Construct validity	
Factor analysis	
Convergent validity	Manifest HD, n = 20[4] **Gait parameters** — **Activities-specific Balance Confidence Scale** **Forward walking** Velocity 0.42 Stride length 0.41 Swing percent –0.18

	Double support percent	–0.15
	Base of support	–0.58
	CV step time	–0.72
	CV stride length	–0.53
	CV swing time	–0.74
	Backward walking	
	Velocity	0.34
	Stride length	0.39
	Swing percent	0.28
	Double support percent	–0.29
	Base of support	–0.24
	CV step time	–0.33
	CV stride length	–0.44
	CV swing time	–0.01
Divergent validity	–	
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Limited information.	
Generalizability		
Shown to be valid at any stage of HD?	No.	
Shown to be valid in any population with dementia or significant cognitive impairment?	No.	
Responsiveness (detect change over time in the construct)		
Demonstrated to be sensitive to change (change over time or to treatment)?	Yes. Manifest HD, n = 20, the ABC-UK 'walking-up-and-down-stairs' and 'Walking around-the-house component' improved following a 9-month multidisciplinary rehabilitation program.[5]	
Interpretability		
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No. The Minimum Detectable Change (MDC) has been determined: Manifest HD, n = 20**,** MDC: 27.33.[4]	
Floor and ceiling effects	Unlikely, this is a self-assessment of confidence (personal judgment).	

Score distributions	
IV. Overall impression	
Advantages	Easy to rate.
Disadvantages	Known to be subject to discrepancies between the self-assessment of the person with HD and the corresponding carer assessment. Questionable use, since lack of insight is a feature in HD.
V. Recommendation	**Suggested for assessment of self-reported balance confidence in HD.**

Supplemental references 23

1. Powell LE, Myers AM. The Activities-specific Balance Confidence (ABC) Scale. *J Gerontol A Biol Sci Med Sci* 1995;50A:M28-34.
2. Mak MK, Pang MY. Fear of falling is independently associated with recurrent falls in patients with Parkinson's disease: a 1-year prospective study. *J Neurol* 2009;256:1689-95.
3. Beninato M, Portney LG, Sullivan PE. Using the International Classification of Functioning, Disability and Health as a framework to examine the association between falls and clinical assessment tools in people with stroke. *Phys Ther* 2009;89:816-25.
4. Kloos AD, Fritz NE, Kostyk SK, Young GS, Kegelmeyer DA. Clinimetric properties of the Tinetti Mobility Test, Four Square Step Test, Activities-specific Balance Confidence Scale, and spatiotemporal gait measures in individuals with Huntington's disease. *Gait Posture* 2014;40:647-51.
5. Thompson JA, Cruickshank TM, Penailillo LE, Lee JW, Newton RU, Barker RA*, et al.* The effects of multidisciplinary rehabilitation in patients with early-to-middle-stage Huntington's disease: a pilot study. *Eur J Neurol* 2013;20:1325-9.

Supplemental table 24

HD Activities of Daily Living (HD-ADL)	
I. Scale description	
Are there several versions of the scale?	No, but there are 20-item[1] and 17-item[2] versions available.
If you replied YES, which was been assessed?	17-item.
Scale construct/ overall structure	Instrumental activities of daily living. The HD-ADL Scale was modeled after the Scale for Instrumental Activities of Daily Living,[3] and has been reported both as a 20- [1] or 17-item[2] informant-completed instrument on which the informant rates the HD patient's ability to perform specific activities, covering the domains of personal care, household care (domestic activities, household upkeep), work and money, social relationships, and communication. For each item the patient is rated on a 4-point scale, from normal to severely impaired. The score ranges from 0 (normal) to 51 (maximal impairment). When items cannot be rated, a pro-rated value is calculated or the item is coded as not impaired (score = 0).[2]
a. Question items	
Items of presence of symptom/sign?	Yes (depending on the item).[2]
Items of severity of symptom/sign?	Yes (depending on the item).[2]
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Yes, discrete steps from 0 (no impairment) to 3 (maximal impairment).[2]
c. Is the scale easy to score?	
Approx. time to score patient	Unknown.
d. Raters	
Patient, caregiver, or clinician	Caretaker: spouse, caretaker, or whoever knows the patient or person at risk the best. **NOTE:** In Brandt 1984[1], a structured interview was mentioned. In Bylsma 1993,[2] HD-ADL was mailed to informant.
If clinician-rated, is training for application required?	N/A

e. Access to scale	
Copyright or public domain?	Copyright, The Johns Hopkins University Press, 1989.[2]
How can the scale be obtained (address or website)?	Bylsma 1993..[2]
Has the scale been published in other languages?	No.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Scarce information. The HD-ADL Scale was modeled after the Scale for Instrumental Activities of Daily Living.[3]
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Yes (personal judgment).
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	No (personal judgment).
Does it score current state or is it based on the patient/caregiver recall?	Caregiver recall for "current" state and for some items there is comparison with the premorbid functional level.[1, 2]
What is the time frame (e.g. "during the past week")?	Not specified.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	Severity (personal judgment).
Is there a cut-off score? (for HD, for non-HD)	No (personal judgment).
d. Acceptability	
Is the length of the scale appropriate?	Moderate (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	Yes, time frame that applies to item score (personal judgment).
Are there ambiguities in rating anchors?	No, although strategies for rating change from item to item (personal judgment).
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not	No. The HD-ADL scale is not adequate for assessing adaptive functioning in patients in the later stages of disease.[2] A ceiling effect for early HD patients would be expected

applicable?	(personal judgment).
e. Has this scale been specifically developed for use in HD (yes/no)?	Yes.[1]
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	Yes.[2, 4, 5]
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.[2]
a. Reliability	
Assessed – not assessed	Assessed in HD.[2]
Internal consistency	Clinical diagnosis of HD, pre-genetic testing era, n=163: Test sample, n=93: coefficient alpha=0.91. Replication sample, n=70: coefficient alpha = 0.96.[2]
Test-retest reliability	Not assessed.
Inter-rater reliability	Not assessed.
b. Validity	
Assessed vs. not assessed	Assessed.[2]
Criterion validity (any comparison with gold-standard)	There is no gold standard.
Construct validity	
Factor analysis	Clinical diagnosis of HD, pre-genetic testing era, n=163.[2] Test sample, n=93; Replication sample, n=70. Principal Component Analysis with VARIMAX rotation revealed 4 factors: 1) General Functioning (personal care and functioning in the community. 2) Domestic Activities, (meals and housework). 3) Home Upkeep (house maintenance and repairs, as well as job performance), and 4) Family Relationships (intrafamilial interactions). Four factors account for 72-74% of the total variance. Authors documented a stable structure from test to replication samples.
Convergent validity	Clinical diagnosis of HD, n=163, n=25[2]**:** TFC x total HD-ADL : r= -0.89, p < 0.001,

	General Functioning: r = - 0.85, p < 0.001, Domestic Activities: r = - 0.79, p < 0.001, Home Upkeep: r = - 0.57, p < 0.002), Family Relationships: n.s.[1, 6]
Divergent validity	**Rothlind (1993)**: n=80: multiple cognitive measures vs. HD-ADL total (p<0.001).[6]
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Requires further testing (personal judgment).
Generalizability	
Shown to be valid at any stage of HD?	Multiple correlations with measures of progression in HD were found in different studies: 1) Greater motor disability, Clinical diagnosis of HD, n=57. QNE: r=0.68, p<0.01.[1] Clinical diagnosis of HD [2]: Test sample, n=93 / Replication sample, n=70: QNE total (MIS and chorea scores are also available) vs. general function (r=0.70 p<0.01/ r=0.76 p<0.001), vs. domestic activities (r=0.44 p<0.001/r=0.53 p<0.001), vs. home upkeep (p=ns/r=0.42 p<0.001), vs. family relationships (both ns), vs. HD-ADL total score (r=0.64 p<0.001/r=0.75 p<0.001). Correlations were not fully reproduced in smaller sample size testing TFC and HD-ADL. 2) Clinical diagnosis of HD. Test sample, n=93; Replication sample, n=70: Duration of chorea vs. general function (r=0.50 p<0.01/r=0.61 p<0.001), vs. domestic activities (ns/r=0.57 p<0.001), vs. home upkeep (both ns), vs. family relationships (both n.s.), HD-ADL (r=0.49 p<0.001/r=0.59 p<0.001). Correlations were not fully reproduced in smaller sample size testing TFC and HD-ADL.[2] Clinical diagnosis of HD, n=57: duration of chorea r=0.55, p<0.01; duration of behavior change r=0.49 p<0.01; duration of symptoms r=0.58 p<0.01; age of onset p=n.s.[1]
Shown to be valid in any population with dementia or significant cognitive impairment?	Yes. Clinical diagnosis of HD only, n=163. Test sample, n=93. Replication sample, n=70: MMSE vs. general function (r=-0.70 p<0.01/r=-0.82 p<0.001), vs. domestic activities (r=-0.45 p<0.001/r=-0.48 p<0.001), vs. home upkeep (r=-0.09 p=ns/r=-0.37 p<0.001),

	vs. family relationships (r=-0.33 p<0.001/r=-0.29 p=n.s.), vs. HD-ADL total score (r=-0.65 p<0.001/r=-0.77 p<0.001).[2]
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Manifest HD, n=73, double blinded placebo-controlled RCT of d-alpha-tocopherol, f/u=12 months: total HD-ADL: -1.7 (d-alpha-tocopherol), -2.2, (placebo), p=n.s.[5] Manifest HD, n=91, double blinded placebo -controlled RCT of idebenone, f/u =12 months:[7] a) based on historical longitudinal data on 49 HD subjects gathered prior to this study - HD-ADL (mean annual change = 3.1 ± 5.3 (no reference given). b) study results. Total HD-ADL: -2.9 ±3.3 (idebenone); 3.1 ± 4.9 (placebo), p=ns, -3.0 ± 4.1 (all participants). Manifest HD, n=46, f/u=2yrs, [8] total HD-ADL: short repeat length (1year =+2.14, 2 years=+3.81), long repeat length (1 year=+2.53, at 3 year=+5.18).
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	Not available.
Floor and ceiling effects	Yes. Floor effect for early HD (personal judgment).
Score distributions	Clinical diagnosis of HD only, n=163: Test sample, n=93; Replication sample, n=70: Mean(SD) - 24.2 (13.2)/22.9 (15.7). [2] Manifest HD, n=91, double blinded placebo -controlled RCT of idebenone, mean (SD): idebenone 11.3 (8.6), placebo 12.5 (8.6).[7]
IV. Overall impression	
Advantages	Comprehensive (more than TFC, includes family related activities).
Disadvantages	Needs an informant, it is not possible to score some items in certain individuals, there are two ways to impute missing value with completely opposite effects on the overall score, needs further clinimetric assessment.
V. Recommendation	**Suggested for assessing severity of limitation in ADLs in HD**

Supplemental references 24

1. Brandt J, Strauss ME, Larus J, Jensen B, Folstein SE, Folstein MF. Clinical correlates of dementia and disability in Huntington's disease. *J Clin Neuropsychol* 1984;6:401-12.
2. Bylsma. Assessment of Adaptive Functioning in Huntington's Disease. *Mov Disord* 1993;8:183-90.
3. Lawton MP. The functional assessment of elderly people. *J Am Geriatr Soc* 1971;19:465-81.
4. Starkstein SE, Brandt J, Folstein S, Strauss M, Berthier ML, Pearlson GD*, et al.* Neuropsychological and neuroradiological correlates in Huntington's disease. *J Neurol Neurosurg Psychiatry* 1988;51:1259-63.
5. Peyser CE, Folstein M, Chase GA, Starkstein S, Brandt J, Cockrell JR*, et al.* Trial of d-alpha-tocopherol in Huntington's disease. *American journal of psychiatry* 1995;152:1771-75.
6. Rothlind JC, Brandt J. A brief assessment of frontal and subcortical functions in dementia. *J Neuropsychiatry Clin Neurosci* 1993;5:73-7.
7. Ranen NG, Peyser CE, Coyle JT, Bylsma FW, Sherr M, Day L*, et al.* A controlled trial of idebenone in Huntington's disease. *Mov Disord* 1996;11:549-54.
8. Brandt J, Bylsma FW, Gross R, Stine OC, Ranen N, Ross CA. Trinucleotide repeat length and clinical progression in Huntington's disease. *Neurology* 1996;46:527-31.

Supplemental table 25

Modified Self-Assessment PD Disability Scale (SPDDS)	
I. Scale description	
Are there several versions of the scale?	A modified version of the original reported scales was used.[1, 2]
If you replied YES, which was been assessed?	Not applicable
Scale construct/ overall structure	The SPDDS is a unidimensional questionnaire that assesses disability in nine daily activities and was developed for PD patients living at home. The original SPDDS consisted of 25 items, however, an item has been added and two items have been dropped due to high nonresponse. The SPDDS contains 24 items. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1148 In HD it was used in a 21-item version.[2]
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	Yes.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Discrete. A five-point scale ranging from 'able to do alone without difficulty' to 'unable to do at all'. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1148
c. Is the scale easy to score?	
Approx. time to score patient	5 minutes. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1148
d. Raters	
Patient, caregiver, or clinician	Patient.
If clinician-rated, is training for application required?	Not applicable.

e. Access to scale	
Copyright or public domain?	Public domain.
How can the scale be obtained (address or website)?	Unknown.
Has the scale been published in other languages?	Unknown.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	–
b. Face validity	
Do the items of the scale cover different components of the specific domain?	–
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	–
Does it score current state or is it based on the patient/caregiver recall?	Patient.
What is the time frame (e.g. "during the past week")?	Unknown.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	Severity.
Is there a cut-off score? (for HD, for non-HD)	–
d. Acceptability	
Is the length of the scale appropriate?	Unknown.
Are there ambiguities in instructions to patient/rater (as applicable)?	Unknown.
Are there ambiguities in rating anchors?	Unknown.
Are the questions appropriate for use in an HD population?	Unknown.
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Unknown.

e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	
a. Reliability	
Assessed – not assessed	Not assessed in HD.
Internal consistency	–
Test-retest reliability	–
Inter-rater reliability	–
b. Validity in HD	
Assessed vs. not assessed	Not assessed in HD.
Criterion validity (any comparison with gold-standard)	–
Construct validity	
Factor analysis	–
Convergent validity	–
Divergent validity	–
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Unknown. No information available.
Generalizability	
Shown to be valid at any stage of HD?	No.
Shown to be valid in any population with dementia or significant cognitive impairment?	No.
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	In neuroleptic naive patients, Clozapine (n=7) vs. Placebo (n=11), mean differences (SD) after 30-day treatment: 5.7 (9.2) (better) *vs.* -3.8 (7.7) (worse) p=0.02. [2]

Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	Unknown.
Score distributions	–
IV. Overall impression	
Advantages	
Disadvantages	Not validated in HD. Lack of insight of patient may be a limitation.
V. Recommendation	**Suggested *with caveats***

Supplemental references 25

1. Brown RG, MacCarthy B, Jahanshahi M, Marsden CD. Accuracy of self-reported disability in patients with parkinsonism. *Arch Neurol* 1989;46:955-9.
2. Vugt JP, Siesling S, Vergeer M, Velde EA, Roos RA. Clozapine versus placebo in Huntington's disease: a double blind randomised comparative study. In; 1997. p. 35-9.

Supplemental table 26

Barthel Index	
I. Scale description	
Are there several versions of the scale?	Yes: 20 point earlier version, still in use (not HD).[1] 10-point scale more commonly used.[2]
If you replied YES, which was been assessed?	The 10-point version,[2] as it is the one used in HD studies.[3-5]
Scale construct/ overall structure	An ordinal scale that evaluates the level of assistance needed by patients to perform 10 basic activities of daily living: Feeding, moving from wheelchair to bed and return, personal toilet, getting on and off toilet, bathing self, walking on level surface, ascend and descend stairs, dressing, controlling bowels, controlling bladder.
a. Question items	
Items of presence of symptom/sign?	Yes.
Items of severity of symptom/sign?	Yes.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Yes.
c. Is the scale easy to score?	
Approx. time to score patient	5 minutes (personal judgment).
d. Raters	
Patient, caregiver, or clinician	Self-report and clinician.
If clinician-rated, is training for application required?	No.
e. Access to scale	
Copyright or public domain?	Copyright, but free for non-funded academic users.
How can the scale be obtained (address or website)?	https://eprovide.mapi-trust.org/instruments/barthel-index
Has the scale been published in other languages?	Yes.

	Danish, Dutch, English, French, German, Italian for Italy, Norwegian, Portuguese, Russia, Spanish, Thai, Chinese, Japanese, Korean.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Yes, from a 20 to a 10 version.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Yes.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	N/A
Does it score current state or is it based on the patient/caregiver recall?	Patient recall and clinician observation.
What is the time frame (e.g. "during the past week")?	Last two days.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To measure severity.
Is there a cut-off score? (for HD, for non-HD)	For acute stroke, but not HD.
d. Acceptability	
Is the length of the scale appropriate?	Yes.
Are there ambiguities in instructions to patient/rater (as applicable)?	No.
Are there ambiguities in rating anchors?	N/A
Are the questions appropriate for use in an HD population?	Yes, partially.
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	No, only appropriate in later stages.
e. Has this scale been specifically developed for use in HD (yes/no)?	No.

e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	
a. Reliability	
Assessed – not assessed	Assessed in HD **NOTE:** Well assessed in stroke, elderly patients and neurological rehabilitation.
Internal consistency	Not assessed.
Test-retest reliability	Not assessed.
Inter-rater reliability	Manifest HD, n=64, ICC=0.97. [4]
b. Validity in HD	
Assessed vs. not assessed	Not assessed in HD.
Criterion validity (any comparison with gold-standard)	Not assessed in HD
Construct validity	
Factor analysis	N/A
Convergent validity	N/A
Divergent validity	N/A
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	Good as a generic test, interesting in order to compare HD with other neurological disease populations.
Generalizability	
Shown to be valid at any stage of HD?	No.
Shown to be valid in any population with dementia or significant cognitive impairment?	No.
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	-

Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	Poor.
Score distributions	Manifest HD, HD stage I – III, n=40, 86.3 (19.0). [3] Pre-manifest HD (n=11), 100 (0), Manifest HD (n=64), 93 (12).[4]
IV. Overall impression	
Advantages	Generic scale, used in many studies.
Disadvantages	No validation in HD. Rarely used in HD
V. Recommendation	**Suggested *with caveats***

Supplemental references 26

1. Mahoney FI, Barthel DW. Functional Evaluation: The Barthel Index. *Md State Med J* 1965;14:61-5.
2. Granger CV, Dewis LS, Peters NC, Sherwood CC, Barrett JE. Stroke rehabilitation: analysis of repeated Barthel index measures. *Arch Phys Med Rehabil* 1979;60:14-7.
3. Zinzi P, Salmaso D, De Grandis R, Graziani G, Maceroni S, Bentivoglio A*, et al.* Effects of an intensive rehabilitation programme on patients with Huntington's disease: a pilot study. *Clin Rehabil* 2007;21:603-13.
4. Quinn L, Khalil H, Dawes H, Fritz NE, Kegelmeyer D, Kloos AD*, et al.* Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. *Phys Ther* 2013;93:942-56.
5. Piira A, van Walsem MR, Mikalsen G, Nilsen KH, Knutsen S, Frich JC. Effects of a One Year Intensive Multidisciplinary Rehabilitation Program for Patients with Huntington's Disease: a Prospective Intervention Study. *PLoS Curr* 2013;5.

Supplemental table 27

Self-report HD Work function (HDWF)	
I. Scale description	
Are there several versions of the scale?	No.
If you replied YES, which was been assessed?	
Scale construct/ overall structure	Perceptions of work function. HDWF is a brief self-assessment that may be used to monitor work function. It captures perceptions of work function as reported by individuals with pre-manifest HD and their companions. It asks questions related to work role limitations and effort, two components of work function that may be affected by cognitive, behavioral, and motor changes in people with pre-manifest HD. The HDWF contains 20 items. The response categories are on a seven-point Likert scale with verbal anchors only at the lowest end (1), “not at all like me”, and at the highest end (7) “very much like me”. The instrument includes a checklist for the employment level that best matches the worker’s current situation, and what workplace adjustments, if any, have been made. Higher scores on the HDWF indicate better function.
a. Question items	
Items of presence of symptom/sign?	N/A
Items of severity of symptom/sign?	N/A
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	20 items scored on 7-point Likert scale.[1]
c. Is the scale easy to score?	
Approx. time to score patient	—
d. Raters	
Patient, caregiver, or clinician	Patient.
If clinician-rated, is training for application	Not applicable.

required?	
e. Access to scale	
Copyright or public domain?	—
How can the scale be obtained (address or website)?	—
Has the scale been published in other languages?	No.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Literature review, review of existing measures, focus groups and interviews (expert evaluation and cognitive interviews with patients with pre-manifest HD and their companions.[1]
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Motor, behavioral, cognitive and compensatory strategies.[1]
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	Weighted toward work function impairment in pre-manifest HD.[1]
Does it score current state or is it based on the patient/caregiver recall?	Unknown (unable to procure a copy of the scale).
What is the time frame (e.g. "during the past week")?	Unknown (unable to procure a copy of the scale).
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	To detect work function ability.[1]
Is there a cut-off score? (for HD, for non-HD)	No.
d. Acceptability	
Is the length of the scale appropriate?	Unknown (unable to procure a copy of the scale).
Are there ambiguities in instructions to patient/rater (as applicable)?	Unknown (unable to procure a copy of the scale).
Are there ambiguities in rating anchors?	No, cognitive interviews with HD participants were conducted to refine the questions.
Are the questions appropriate for use in an HD population?	Yes.

Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	No, only for those in employment.
e. Has this scale been specifically developed for use in HD (yes/no)?	Yes.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	No.
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.
a. Reliability	
Assessed – not assessed	Not assessed in HD.
Internal consistency	–
Test-retest reliability	–
Inter-rater reliability	–
b. Validity in HD	
Assessed vs. not assessed	Assessed in HD.
Criterion validity (any comparison with gold-standard)	No gold standard available.
Construct validity	
Factor analysis	—
Convergent validity	Pre-manifest (n=238) + Companion (n=70): HDWF X Endicott work productivity scale (r=-0.56); X Social Adjustment Scale self-report (r=-0.29); X Everyday cognition (r=-0.70).[1]
Divergent validity	—
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	—
Generalizability	
Shown to be valid at any stage of HD?	No.
Shown to be valid in any population with dementia	No.

or significant cognitive impairment?	
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	No.
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	—
Score distributions	Total score (mean (sd)) for pre-manifest HD: 98.27 (18.59). [1]
IV. Overall impression	
Advantages	Clearly described development process.
Disadvantages	Not used in studies outside the PREDICT group, difficulty in obtaining a copy for review.
V. Recommendation	**Listed**

Supplemental references 27

1. Brossman B, Williams JK, Downing N, Mills JA, Paulsen JS. Development of the Huntington disease work function scale. *J Occup Environ Med* 2012;54:1300-8.

Listed

Supplemental table 28

Behavior Observation Scale Huntington (BOSH) - ADL subscale	
I. Scale description	
Are there several versions of the scale?	No.[1]
If you replied YES, which was been assessed?	N/A.
Scale construct/ overall structure	Repeated monitoring for longitudinal assessment, of an inventory of the behavior in the later stages of the disease.[1] The BOSH contains 32 items in 3 subscales: 1) Activities of daily living (ADL). 2) Social–cognitive functioning, and 3) Mental rigidity and aggression. Only the ADL component was considered.[1]
a. Question items	
Items of presence of symptom/sign?	Yes.[1]
Items of severity of symptom/sign?	Yes. Severity by a graded statement expressed in terms of degree of autonomy or frequency of a behavior.[1]
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Discrete (4 steps): no standardized response key for items.[1]
c. Is the scale easy to score?	
Approx. time to score patient	10 to 15 minutes with a mean of 14 minutes (for the full scale, including the 3 components); possibly 3-4 minutes (divide total by number of valid items),[1] (personal judgment).
d. Raters	
Patient, caregiver, or clinician	Clinician as observation-based (observations of the staff of the nursing home).[1]
If clinician-rated, is training for application	No training required. Instructions are printed on the questionnaire.[1]

required?	
e. Access to scale	
Copyright or public domain?	Public domain.[1]
How can the scale be obtained (address or website)?	Reinier Timman @ Erasmus Medical Center Rotterdam, Medical Psychology and Psychotherapy, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands (e-mail: r.timman@erasmusmc.nl).[1]
Has the scale been published in other languages?	No, published in English but only tested in Dutch.[1]
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Yes. Two pilot questionnaires - both in Dutch - preceded the final version of the BOSH. 1) Experts, psychologists of the specialized HD wards, reached consensus based on HD patient characteristics for the items of the first pilot. Consensus was reached on 11 characteristics of the HD patient according to the observations of nursing home staff: 1) inflexible behavior, 2) need for social care, 3) need for mental care, 4) need for physical care, 5) communication problems, 6) choking problems, 7) uncontrolled eating and drinking behaviors, 8) self-oriented behavior, 9) repetitive behavior, 10) aggressive behavior, and 11) inability to perform complex actions. Principal component analysis (PCA) revealed 6 components: speech capability, mental rigidity– aggression, social–cognitive capacities, obsessive-compulsive behavior, voraciousness, and deterioration of ADLs. 2) For the construction of a second version, the items were restructured in line with the 6 components that emerged from the first pilot. Twenty-four items with the highest loadings on each component selected on the premise that a large conceptual overlap was not present. Four items with lower component loadings, which were considered clinically essential aspects of HD, were added. These items involved the ability to stop current activities, information processing and memory, behavior when a fellow patient needs immediate help, and behavior when a fellow patient is helped first. Four items from the functional assessment subscale of the UHDRS were included. Experts reached consensus on these items that they considered essential to the manifestation of

	HD. Ambiguous items, double questions and items with an overlap in the response possibilities, as well as items with gaps between answer possibilities were reformulated. The second version was administered to 84 patients in one Belgian and 3 Dutch nursing homes. 3) For the third and final version administered in Dutch to 91 patients in the 4 nursing homes, 32 items were reformulated to avoid overlap and gaps, as well as ambiguity, and presented in a more logical order. Ratings for outpatients and for tube-fed patients were introduced.[1]
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Yes, the ADL component has 9 items - going to the toilet, going to bed, mobility, comprehensibility through nonverbal communication, voice control and articulation, eating and drinking, washing and getting dressed, intelligibility, and choking while eating or drinking.[1]
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	No, it is adapted to late HD.
Does it score current state or is it based on the patient/caregiver recall?	Clinician recall.[1]
What is the time frame (e.g. "during the past week")?	Observation over previous two weeks.[1]
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	Monitor severity longitudinally (personal judgment).
Is there a cut-off score? (for HD, for non-HD)	No (personal judgment).
d. Acceptability	
Is the length of the scale appropriate?	Yes, the ADL subscale has 9 items (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No, but it requires consistency of the clinician over the past two weeks (note completed by nurses in testing) (personal judgment).
Are there ambiguities in rating anchors?	These items have some overlapping responses that are not clearly differentiated: item 2, regarding going to bed: needs "some assistance" vs just " needs assistance"; item 4, regarding going to the toilet: "some assistance vs "almost full assistance"; item 7,

	regarding voice control: "affected" vs "bad" (personal judgment).
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	No, developed and tested in stage 3 and 4 Shoulson and Fahn's staging system (late stage).[1]
e. Has this scale been specifically developed for use in HD (yes/no)?	Yes.[1]
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	No
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	Yes.[1]
a. Reliability	
Assessed – not assessed	Assessed in HD.[1]
Internal consistency	Cronbach's alpha (ADL Component) = 0.94 (Sample 1 and 2).[1]
Test-retest reliability	N/A
Inter-rater reliability	Intraclass correlation coefficient (nurses) = 0.95.[1]
b. Validity in HD	
Assessed vs. not assessed	Assessed in HD.[1]
Criterion validity (any comparison with gold-standard)	There is no gold standard, but no association with longer disease duration has been reported. There is an association with duration of care (beta:0.5, p=0.004) (personal judgment).
Construct validity	
Factor analysis	PCA and subscale supported by factor structure; VARIMAX rotation. The scree test resulted in the selection of a 3-component solution in both samples. (ADL, social–cognitive capabilities, and rigidity–aggression).[1]
Convergent validity	Not assessed.[1]
Divergent validity	Not assessed.[1]
Overall impression: good – not good (based on references preferably, personal judgment can be	Not good at the time of writing, further testing is required.[1]

stated)	
Generalizability	
Shown to be valid at any stage of HD?	No, only at stages 4 and 5 (personal judgment).
Shown to be valid in any population with dementia or significant cognitive impairment?	Yes (personal judgment).
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Not tested over time.
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No data.
Floor and ceiling effects	No data.
Score distributions	No data.
IV. Overall impression	
Advantages	Easy and quick to apply, attempts to fill a gap in late HD stages where specific scales are not available (personal judgment).
Disadvantages	Limited testing of measurement properties. Lacks external validation and translation into other languages (personal judgment).
V. Recommendation	**Listed**

Supplemental references 28

1. Timman R, Claus H, Slingerland H, van der Schalk M, Demeulenaere S, Roos RA*, et al.* Nature and development of Huntington disease in a nursing home population: The Behavior Observation Scale Huntington (BOSH). *Cogn Behav Neurol* 2005;18:215-22.

Supplemental table 29

Alzheimer's Disease Cooperative Study Activities of Daily Living Scale (ADCS-ADL)	
I. Scale description	
Are there several versions of the scale?	Yes[1]
If you replied YES, which was been assessed?	Unknown.
Scale construct/ overall structure	An inventory of informant based items to assess activities of daily living and instrumental activities of daily living, i.e. functional performance, of Alzheimer's disease (AD). The ADCS-ADL was the first ADL scale to be developed for use specifically in clinical trials with people with AD across the range of severity. Scores on the 24-item ADCS-ADL range from 0 to 78, higher scores reflect greater competence.
a. Question items	
Items of presence of symptom/sign?	Yes.
Items of severity of symptom/sign?	Yes.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Discrete.
c. Is the scale easy to score?	
Approx. time to score patient	–
d. Raters	
Patient, caregiver, or clinician	Caregiver or clinician.
If clinician-rated, is training for application required?	
e. Access to scale	
Copyright or public domain?	Copyright.

How can the scale be obtained (address or website)?	Alzheimer's Disease Cooperative Study
Has the scale been published in other languages?	—
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Galasko et al. (1997)[2] selected the items for the ADCS-ADL from a pool of 45 items thought to be relevant to the target population on the basis of existing scales and clinical experience. An item was included in the final measure fit the criteria. It was performed either premorbidly or at baseline by >90% of participants (showing it was applicable to the target group), had a kappa agreement statistic at 1–2 months of >0.4 (indicating good test-retest reliability), had a significant correlation with MMSE score (indicating appropriate scaling and validity), and showed decline over 12 months in at least 20% of participants (indicating validity and sensitivity to change).
b. Face validity	
Do the items of the scale cover different components of the specific domain?	–
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	–
Does it score current state or is it based on the patient/caregiver recall?	It can be completed by a caregiver in questionnaire format, or administered by a clinician/researcher as a structured interview with a caregiver.
What is the time frame (e.g. "during the past week")?	'In the past 4 weeks'.[2]
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	Severity.
Is there a cut-off score? (for HD, for non-HD)	Not for HD.
d. Acceptability	
Is the length of the scale appropriate?	Yes.

Are there ambiguities in instructions to patient/rater (as applicable)?	–
Are there ambiguities in rating anchors?	–
Are the questions appropriate for use in an HD population?	–
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	–
e. Has this scale been specifically developed for use in HD (yes/no)?	No.
e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	No.
a. Reliability	
Assessed – not assessed	Not assessed in HD.
Internal consistency	–
Test-retest reliability	–
Inter-rater reliability	–
b. Validity in HD	
Assessed vs. not assessed	Not assessed in HD.
Overall impression: good – not good	–
Criterion validity (any comparison with gold-standard)	
Construct validity	
Factor analysis	–
Convergent validity	–
Divergent validity	–
Overall impression: good – not good (based on references preferably, personal judgment can be	No information available in HD.

stated)	
Generalizability	
Shown to be valid at any stage of HD?	No.
Shown to be valid in any population with dementia or significant cognitive impairment?	No.
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	No. Treatment Effects on Efficacy Outcomes at Week 26 in RCT of latrepirdine vs Placebo: −1.8 vs. −0.8.[3]
Interpretability	
Has the minimal clinically important change and minimal clinically relevant incremental difference been assessed?	No.
Floor and ceiling effects	Unknown.
Score distributions	Mean Scores (SD) in RCT of latrepirdine Group (n = 200) 59.6 (14.1) vs Placebo Group (n = 203) 59.3 (14.4). [3]
IV. Overall impression	
Advantages	Unknown.
Disadvantages	Not validated in HD. Used in a single clinical trial.
V. Recommendation	**Listed.**

Supplemental references 29

1. Fish J. Alzheimer's Disease Cooperative Study ADL Scale. In: Kreutzer JS, DeLuca J, Caplan B, eds. Encyclopedia of Clinical Neuropsychology. New York, NY: Springer New York, 2011:111-12.
2. Galasko D, Bennett D, Sano M, Ernesto C, Thomas R, Grundman M, *et al.* An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord* 1997;11 Suppl 2:S33-9.
3. Dorsey ER. A randomized, double-blind, placebo-controlled study of latrepirdine in patients with mild to moderate huntington disease: HORIZON investigators of the huntington study group and european huntington's disease network. In; 2013. p. 25-33.

Supplemental table 30

Quick Disabilities of Arm, Shoulder & Hand (Quick DASH)	
I. Scale description	
Are there several versions of the scale?	There is the full version also (DASH).
If you replied YES, which was been assessed?	Quick DASH.
Scale construct/ overall structure	The purpose of the QuickDASH is to use 11 items to measure physical function and symptoms in people with any or multiple musculoskeletal disorders of the upper limb. The QuickDASH is a widely used reference of self-reported disability. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1267
a. Question items	
Items of presence of symptom/sign?	No.
Items of severity of symptom/sign?	Yes.
b. Response scale	
Are the items of the scale scored in discrete steps (specify number) or in a visual analogue scale?	Discrete (5 steps).
c. Is the scale easy to score?	
Approx. time to score patient	10 minutes. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1267
d. Raters	
Patient, caregiver, or clinician	Patient
If clinician-rated, is training for application required?	No training required. http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1267
e. Access to scale	
Copyright or public domain?	Free of charge. Sole property of the Institute for Work & Health. NOTE: some uses require the issue of a license (Commercial or profit publications) http://dash.iwh.on.ca/conditions-use?n=quickdash

How can the scale be obtained (address or website)?	http://dash.iwh.on.ca/conditions-use?n=quickdash
Has the scale been published in other languages?	Unknown.
II. Scale properties	
a. Content validity	
Any process for item generation and/or reduction	Yes.
b. Face validity	
Do the items of the scale cover different components of the specific domain?	Yes.
Is the scale heavily weighted towards one aspect of the domain it measures and which components of the domain are not covered?	–
Does it score current state or is it based on the patient/caregiver recall?	Patient.
What is the time frame (e.g. "during the past week")?	Last week.
c. Use	
Purpose: to measure severity, screen or diagnosis of the domain?	Severity.
Is there a cut-off score? (for HD, for non-HD)	
d. Acceptability	
Is the length of the scale appropriate?	Yes (personal judgment).
Are there ambiguities in instructions to patient/rater (as applicable)?	No (personal judgment).
Are there ambiguities in rating anchors?	Yes: subjective judgment without concrete anchors (personal judgment).
Are the questions appropriate for use in an HD population?	Yes (personal judgment).
Is the scale applicable across HD disease stages? Are there HD stages in which the scale is not applicable?	Unknown.
e. Has this scale been specifically developed for use in HD (yes/no)?	No.

e1. If yes to the above, has the scale been deployed in HD by groups other than the developers?	
III. Clinimetric/psychometric properties	
Are there clini- or psychometric properties in HD ?	No.
a. Reliability	
Assessed – not assessed	Not assessed in HD.
Internal consistency	–
Test-retest reliability	–
Inter-rater reliability	–
b. Validity in HD	
Assessed vs. not assessed	Not assessed in HD.
Criterion validity (any comparison with gold-standard)	–
Construct validity	
Factor analysis	–
Convergent validity	–
Divergent validity	–
Overall impression: good – not good (based on references preferably, personal judgment can be stated)	–
Generalizability	
Shown to be valid at any stage of HD?	No.
Shown to be valid in any population with dementia or significant cognitive impairment?	No.
Responsiveness (detect change over time in the construct)	
Demonstrated to be sensitive to change (change over time or to treatment)?	Manifest HD, n=10, On – off tetrabenazine evaluation: Off 43.2 (27.5), On 37.3 (26.5), p=0.307.[1]
Interpretability	
Has the minimal clinically important change and	No.

minimal clinically relevant incremental difference been assessed?	
Floor and ceiling effects	Unknown.
Score distributions	–
IV. Overall impression	
Advantages	Unknown in HD.
Disadvantages	Unknown in HD.
V. Recommendation	**Listed**.

Supplemental references 30

1. Fekete R, Davidson A, Jankovic J. Clinical assessment of the effect of tetrabenazine on functional scales in huntington disease: a pilot open label study. *Tremor Other Hyperkinet Mov (N Y)* 2012;2.

Rating Scales and Performance-based Measures For Assessment of Functional Ability In Huntington's Disease: Critique And Recommendations

Tiago A. Mestre MD MSc,[1*] Monica Busse BSc. BSc (Med) Hons. MSc (Med) PhD,[2] Aileen M. Davis PhD,[3] Lori Quinn, EdD, PT,[4] Filipe B. Rodrigues, MD, [5] Jean-Marc Burgunder, [6] Noelle E. Carlozzi PhD,[7] Francis Walker MD,[8] Aileen K. Ho PhD,[9] Cristina Sampaio MD PhD,[10] Christopher G. Goetz MD,[11] Esther Cubo MD,[12] Pablo Martinez-Martin PhD,[13] Glenn T. Stebbins PhD,[11] and the Members of the MDS Committee on Rating Scales Development

1 Parkinson's disease and Movement Disorders Center, Division of Neurology, Department of Medicine, The Ottawa Hospital Research Institute, University of Ottawa Brain and Mind Institute, Canada.

2 Centre for Trials Research, Cardiff University, Wales, UK.

3 Krembil Research Institute, University Health Network and Institute of Health Policy, Management and Evaluation and Rehabilitation Institute, University of Toronto, Canada.

4 Department of Biobehavioral Sciences, Teachers College, Columbia University, USA.

5 Huntington's Disease Centre, Institute of Neurology, University College London, UK
Clinical Pharmacology Unit, Instituto de Medicina Molecular, Portugal
Laboratory of Clinical Pharmacology and Therapeutics, Faculty of Medicine, University of Lisbon, Portugal

6 Swiss HD Center, NeuroZentrumSiloah and Department of Neurology, University of Bern, Switzerland

7 Department of Physical Medicine and Rehabilitation, University of Michigan, USA.

8 Department of Neurology, Wake Forest School of Medicine, USA

9 School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK.
10 CHDI Foundation/CHDI management, USA
11 Department of Neurological Sciences, Rush University Medical Center, Chicago, USA
12 Department of Neurology, Hospital Universitário HermanosYagüe, Burgos, Spain
13 National Center of Epidemiology and CIBERNED, Carlos III Institute of Health, Madrid, Spain

*** Corresponding author:**

Tiago A. Mestre

Parkinson's disease and Movement Disorders Center

Division of Neurology, Department of Medicine

University of Ottawa

1053 Carling Avenue, Ottawa ON K1Y 4E9, Canada

Telephone: +1 613 979 1513; E-mail: tmestre@toh.on.ca

Potential conflict of interest: Nothing to report

Word count of abstract: 267

Word count of main text: 4337~~4~~8

Number of tables: 3

Key words: Huntington's disease, physical function, performance measures, rating scales

Abstract

Limitation of functional ability is a major feature of Huntington's disease (HD). The International Parkinson and Movement Disorder Society (MDS) commissioned the ~~assessment of the clinimetric properties of clinical measures of functional ability in HD to make recommendations regarding their use based on standardized criteria~~ **appraisal of the use and clinimetric properties of clinical measures of functional ability that have been applied in HD studies and trials to date, to make recommendations regarding their use based on standardized criteria.** After a systematic literature search, we included a total of 29 clinical measures grouped into two categories: 1) performance-based measures (e.g., balance, walking, reaching/grasping), and 2) rating scales. Three performance-based measures are rated as "recommended": the Tinetti Mobility Test for screening of fall risk and for severity assessment of mobility in patients with manifest HD (up to stage III); the Berg Balance Scale for severity of balance impairment; and the Six-Minute Walk Test for assessment of walking endurance (severity) in HD subjects with preserved ambulation. No rating scale targeting functional ability reached a "recommended" status, either for screening or severity measurement.

The main challenges identified in this review include applying widely accepted conceptual frameworks to the identified measures, the lack of validation of clinical measures to detect change over time, and absence of validated measures for upper limb function. Furthermore, measures of capacity or ability to perform activities of daily living had ceiling effects in people with early and pre-manifest HD. We recommend that the MDS prioritize the development of new scales that capture small but meaningful changes in function over time for outcome assessment in clinical trials, particularly in earlier stages of HD.

Introduction

The ability to perform daily life activities depends on the integration of motor, cognitive and behavioral functioning. These domains are progressively impaired in Huntington's disease (HD). A measure of functional ability based on key life activities is thus an attractive outcome in clinical studies, namely for treatment trials. A single measure pertinent to patient overall function would be useful to capture changes occurring simultaneously in the different symptom domains in HD. Further, functional ability measures are valued as an outcome for drug development by regulatory agencies.[1]

There is a need to identify and critically appraise the measurement properties of clinical measures currently used to capture functional ability in people with HD to inform optimal application in clinical research. The scope of this review is directed towards physical function and includes a wide spectrum of clinical measures from those capturing motor tasks such as walking and balance ability, to those assessing the ability to perform activities of daily living (ADL).

~~The current review aims to provide recommendations and identify gaps in the use of these clinical measures for HD populations.~~ **The current review aims to provide recommendations and identify gaps in the use and validation of these functional measures that have been used in HD studies and trials to date.** Such information will inform the field, identifying where additional testing of measurement properties or development of new measures may be required.

METHODS

We followed the methodology proposed by the MDS Committee on Rating Scales Development described elsewhere,[2] and includes i) Organization and Critique Process, ii) Selection of Scales, iii) Inclusion/Exclusion for Review, iv) Criteria for Rating Scales Recommendation (Table 1). For selection of measures, the keywords selected for this review were "Huntington*" OR "Westphal variant" OR "juvenile Huntington*", and the terms "scale" OR "questionnaire" OR "index" OR "measure" as well as keywords: "function", "activit* daily li*", "capacity", "*ability", "impairment". Manuscripts published before October 17, 2016 were retrieved using the above search strategy and thoroughly screened by the chair of the sub-committee (T.A.M.) to ascertain which clinical measure had been used in each study. To aid our categorization of clinical measures in this review, we applied a widely accepted classification of the health components of functioning and disability: The International Classification of Functioning, Disability and Health (ICF).[3] The ICF defines: 1) impairments or problems in body function or structure such as a significant deviation or loss, 2) activity or the execution of a task, 3) and participation or involvement in a life situation.[3] By consensus, we included clinical measures in this review that captured a) activity or the execution of a task or tasks, and b) participation or involvement in a life situation.

Identified Clinical Measures and Their Utilization in Clinical Research

A total of 47 potentially relevant clinical measures were identified. After screening for exclusion criteria with abstract screening and in-depth review, a total of 29 measures were included and divided in performance-based measures defined as functional assessments based on the live performance of a task (e.g., balance, walking, reaching/grasping) (n=17) and rating scales (n=12) capturing the assessment of various aspects of functional ability based on recall. (See *Supplementary material* for more details)

123

124 **Critique of Measures of Functional Ability**

125 We provide a summary description of the performance-based measures and rating scales
126 classified as "recommended" or "suggested". See *Supplementary material* for a full
127 description of all clinical measures included for full review, including those that were
128 included in the "suggested *with caveats*" or "listed" categories.

129 **1) Performance-based measures**

130 **"RECOMMENDED"**

131 **Tinetti Mobility Test (TMT)**

132 The TMT is a 16-item clinician-administered performance measure, which consists of balance
133 and gait subscales that measure static and dynamic balance. It was originally developed to
134 measure balance and screen for risk of falls in the elderly,[4] but has been used in other patient
135 populations.[4] During the 10-15 minute test, patients perform a series of balance and walking
136 tasks and are rated on a 0-2 scale based on qualitative assessment of performance.[4] The TMT
137 has been used in several studies in HD and demonstrates good test-retest reliability in early-,
138 mid-, and late stage HD (ICC = 0.8-0.9).[5, 6] Higher scores in the TMT correlated positively
139 with spatio-temporal measures of gait (e.g., velocity r=0.68; stride length r=0.74), with higher
140 scores of the UHDRS-FAS (r=0.44) and UHDRS-TFC (r=0.42) and lower scores of the
141 UHDRS-Total Motor Score (TMS) (r=-0.59).[5, 7, 8] The TMT has demonstrated responsiveness
142 in the context of interventional studies, including an intensive rehabilitation intervention
143 program in patients with HD stages I–III (pre= 15.97, post=20.79, p<0.001), [9] and off- (17.09
144 ± 4.04) and on-tetrabenazine (19.91 ± 3.53, p<0.02) study of manifest HD patients. [10]
145 However, there was no significant change in the TMT following a video-based balance

training program.[11] A cut-off score of 21 has 74% sensitivity and 60% specificity in identifying fallers in HD.[5]

Recommendation: The TMT is "recommended" for assessment of mobility in patients with manifest HD (up to stage III) and "recommended" for screening for risk of falls .

The Berg Balance Scale (BBS)

The BBS is a performance measure consisting of 14 subtests of various activities related to balance that takes 10 to 15 minutes to complete. These activities include static postures (e.g., sitting, standing), transitions (e.g., sitting to standing, transferring between chairs), and challenging positions (e.g., standing with eyes closed). Quality of performance for each item is scored using a 4-point scale, with higher scores indicating better balance, and a possible maximum score of 56. Although originally developed to measure balance in older people, the BBS has been widely used in HD, although it has limited applicability in non-ambulatory HD due to the nature of the activities.[6, 12-19] The available clinimetric data show that it has good test-retest reliability in both pre-manifest (ICC=0.86) and manifest HD (ICC=0.96).[6] A minimal detectable change (MDC) of 5 in people with manifest HD has been reported.[6] Convergent validity has been reported between the BBS and the HD-ADL (r= -0.47), UHDRS TFC (r=0.60[19] and r=0.43[7]), UHDRS-FAS (r=0.48)[7], and UHDRS-TMS (r= -0.55).[7] Sensitivity to change following treatment withdrawal (tetrabenazine) was reported in a small open-label cohort.[14] A cut-off score of 40 was used as a cut-off to predict being a "faller" for a plotted probability of 60%.[86]

Recommendation: The BBS is "recommended" for assessing severity of balance impairment in ambulatory HD. The BBS is "suggested" for screening for fall risk, as no sensitivity or specificity data for falls have been reported.

The Six-Minute Walk Test

The Six-Minute Walk test measures how many meters an individual can walk in 6 minutes. [20, 21] Two practice tests are recommended, but not always carried out.[22, 23] It has been applied as a measure of endurance in neurological conditions, in contrast to shorter walk tests that generally measure velocity of walking speed.[6] It has been used in patients with pre-manifest and manifest HD, although it cannot be used for those who are non-ambulatory. Excellent test-retest reliability data have been reported in pre-manifest (ICC = 0.98) and manifest HD (IC=0.94; early and late HD = 0.97, and mid-stage HD=0.86).[6, 24] It is unclear how values discriminate among pre- and manifest HD severity levels as there is an overlap of the 95% confidence interval (CI) around mean values in both groups. On the other hand, values may separate pre- and early manifest HD from mid- to late stage HD.[6] Low correlations have been reported between the Six-Minute Walk Test and the UHDRS-FAS,[7] but higher correlations are not expected due to the limited overlap of the measure constructs. The MDC has been reported to be 39.2 meters for pre-manifest HD and 86.6 meters for manifest HD (range: 56.6 to 126.1 meters).[6]

Recommendation: The Six-Minute Walk test is "recommended" for the assessment of walking endurance (severity) across HD severity.

"SUGGESTED"

Timed 'up and go' Test (TUG)

The TUG is a simple and quick (<3 minutes) to use test that assesses mobility, balance and risk of falls. Although not specifically developed for use in HD, it has been used in pre-manifest and manifest HD to measure severity and screen for risk of falls.[13, 25] The TUG

measures the time it takes for a patient to rise from a chair, walk three meters, turn around, walk back to the chair, and sit down. One practice test is recommended before scoring the test.[25] Mean scores for patients with manifest HD range from 9-17 seconds [6, 19] and a cut-off score of 14 seconds has been reported to predict being a "faller" for a plotted probability of 60%.[13] Test-retest reliability in HD has been shown to be excellent (ICC = 0.93 [pre-manifest HD], 0.96 [manifest HD]) and the MDC has been reported to be 1.34 seconds in pre-manifest HD and 2.98 seconds in manifest HD.[6] The TUG was not statistically significantly correlated with the UHDRS-TMS or the UHDRS-TFC and correlated weakly with the UHDRS-FAS ($r = -0.33$, $p<0.01$).[7] Pre-post scores improved by an average of 1.3 seconds following training in a non-controlled study, that follow within the MDC.[26] The TUG can be used in early to mid-stages of HD, but not in pre-manifest or late stage HD, and it appears to be sensitive to disease progression, but does not discriminate between disease subtypes.[6, 19, 27]

Recommendation: The TUG is "suggested" for assessing severity of balance and mobility, and "suggested" for screening for fall risk. There is no sensitivity or specificity data for the reported cut-off point. Construct validity needs further assessment.

The Ten-Meter Walk Test

The Ten-Meter Walk test is a quick and easy performance-based measure that assesses walking speed. The score is based on the mean of two tests. The test has been used in pre-manifest and manifest HD with varying walking speeds: self-paced[6, 7, 24] and fast-paced.[6, 17, 24] Test-retest reliability has been shown to be good in both pre-manifest and manifest HD for the self-paced version.[6] For the self-paced version there was no correlation with the UHDRS-TMS, a weak correlation was reported with the UHDRS-FAS ($r=0.35$, $p<0.01$) and none with the UHDRS-TFC.[7] The fast-paced version of the test has been shown to be sensitive to

change following a rehabilitation program intervention in mild to moderate manifest HD (improvement of 0.27 m/s).[17] Following a 12-week community-based exercise program there was no significant change for either the self- or fast-paced versions.[24]

Recommendation: The Ten-Meter walk test is "suggested" for assessing walking speed in manifest HD. The vast majority of the clinimetric data sustaining this recommendation was obtained using the self-paced version.

Four Square Step Test (FSST)

The FSST is a 5-10 minute test of dynamic balance. The FSST clinically assesses a patient's ability to step over canes positioned in a cross shape in three directions in a set sequence: forward, sideways, and backwards. The test was not specifically developed for use in HD, but has been used in three studies in HD, and some clinimetric data are available in pre- and manifest HD.[6, 8, 11] Test-retest reliability has been reported to be excellent in pre-manifest HD (ICC=0.91), and good in manifest HD (ICC=0.78).[6] The MDC is higher in manifest HD (15.2) than in pre-manifest HD (1.9).[6] Moderate to high correlation has been shown between the FSST and the ABC (Pearson correlations: –0.57; $p<0.05$); the Tinetti Mobility Test (Pearson correlations: –0.67, $p<0.01$), and gait velocity (Pearson correlations: –0.69, $p<0.01$).[8] The FSST has not been shown to be sensitive to change in one exercise study.[11]

Recommendation: The FSST is "suggested" for assessing dynamic balance in HD

Mini Balance Evaluation Systems Test (Mini-BESTest)

The Mini-BESTest is a 14-item measure of dynamic balance. Derived from the Balance Evaluation Systems Test (BESTest), factor analysis was used for item reduction to include

dynamic balance only, and to improve clinical utilization.[28] Administered in 10-15 minutes, the Mini-BESTest evaluates domains of postural control. Each question is rated from normal to severe and scored between 0 and 2, for a maximum total score of 28 points. The test was not specifically developed for HD, and has not been assessed comprehensively across stages of HD. The test is not applicable to non-ambulatory patients.[29] Convergent validity has been shown between the Mini-BESTest and the ABC (r^2=0.45), UHDRS-TFC (r^2=0.75) and UHDRS-TMS (r^2=0.68).[29]

Recommendation: The Mini-BESTest is "suggested" for assessing severity of balance impairment in HD, as it has been used in only one study with a very small sample size across HD severity with a partial clinimetric assessment.

Physical Performance Test (PPT)

The PPT is a ten-minute test, which assesses multiple domains of physical function using observed performance of tasks that simulate activities of daily living (ADL) of various degrees of difficulty (writing, eating, dressing, walking, and climbing stairs).[30] Each activity is timed and rated from 0-4, a higher score indicating better physical performance. The test was not specifically developed for use in HD, but some of its clinimetric properties have been assessed in both pre- and manifest HD. Good test-retest reliability has been recorded in pre-manifest HD (ICC = 0.76) and excellent reliability in manifest HD (ICC=0.95). The MDC was 3 points for pre-manifest HD and 5 points for manifest HD respectively.[6] Convergent validity has been reported in manifest HD between the PPT and the UHDRS-TMS (r = -0.41 n=63, $p<0.01$), the UHDRS-FAS (r = 0.59, $p<0.01$); and the UHDRS-TFC (r= 0.48, $p<0.05$).[7]A ceiling effect has been reported in pre-manifest HD.[6] It has also been shown to be valid in patients with cognitive impairment.[31]

Recommendation: The PPT is "suggested" for assessing severity of impairment of physical function in performance of tasks that simulate activities of daily living.

Six-condition Romberg test

The six-condition Romberg test is a 5-minute easy to administer performance-based measure of balance developed in the context of myelopathies and neuropathies with an associated sensory dysfunction. The amount of time the patient maintains the position without loss of balance for 6 standard conditions is recorded, for a maximum score of 180 seconds. Higher scores indicate better balance. The test has been used in some HD studies [6, 10] and the clinimetric data available document good test-retest reliability in both pre-manifest (ICC=0.73) and manifest HD (ICC=0.89).[6] The six-condition Romberg test is a valid tool that can be used across all stages of HD provided that the patient is ambulatory as it is likely to have floor effects in non-ambulatory patients.[6] It has not been shown to be sensitive to change in treatment.[10] People with pre-manifest HD (158.8±22.2) have higher scores (better performance) than those with manifest HD (70.0±41.1).[6]

Recommendation: The Six-Condition Romberg Test is "suggested" for assessing severity of balance impairment in HD

2) Rating Scales

"SUGGESTED":

The Unified Huntington's Disease Rating Scale (UHDRS) - Total Functional Capacity (TFC)

The UHDRS-TFC is part of a multi-component rating scale originally designed to prospectively evaluate all patients with HD and individuals at risk for HD.[34] It assesses capacity as opposed to actual performance, and consists of a 5-item interview between a clinician, and the patient and a person familiar with the patient's functioning. It takes < 5 minutes to complete and covers basic activities of living: occupation, handling finances, domestic responsibilities, ADLs such as eating, dressing, bathing, and level of care (home or facility). A higher score indicates better functional capacity. The UHDRS-TFC has been used in pre-manifest and manifest HD populations in multiple observational studies and randomized controlled trials.[34-51] The TFC total score can be categorized into Shoulson and Fahn HD stages.[35] There is evidence of excellent inter-rater reliability, but only for a modified version of the UHDRS-TFC that is filled by patient or the caregiver (ICC = 0.96, 95% CI: 0.92, 0.98).[52] Data from multiple studies suggest good convergent validity with other components of the UHDRS assessing the functional domain and quality of life, and good divergent validity with motor disability, cognitive deficits and behavioral problems.[19, 29, 34, 53-60] Extensive data from multiple observational studies and clinical trials suggest sensitivity to change over time.[34-51, 61-70] There appears to be a ceiling effect for early stage HD and a floor effect for late stage HD.[41]

Recommendation: The UHDRS-TFC is "suggested" for assessing severity of limitation in functional capacity in HD, because it lacks core clinimetric data, namely, test-retest reliability and internal consistency to reach a "recommended" status.

The UHDRS - Functional Assessment Scale (FAS)

The UHDRS-FAS is an extensively-used checklist that is also part of the UHDRS. It is a clinician-administered questionnaire with 25 items which screen an individual's capacity to complete specific tasks, enables the clinician to assess severity, and make longitudinal assessments. The questionnaire takes 5-10 minutes to complete. It is considered an extension of the TFC and is more detailed in certain tasks.[34] A total score is obtained by giving 1 point to all "yes" replies, and a higher score indicates better functioning.[34] It has been used in multiple observational studies and randomized controlled trials in manifest HD populations.[34, 39, 43, 48, 49, 61, 62, 64, 68, 70-72] The UHDRS-FAS has been shown to have high internal consistency (Cronbach's $\alpha = 0.95$).[34] There are no available data on test-retest reliability or inter-rater reliability. Good convergent validity with other components of the UHDRS has been shown, as well as with motor disability, cognitive and behavioral deficits.[34, 54, 58, 73, 74] The UHDRS-FAS has been shown to be sensitive to change over time in several studies.[34, 39, 42, 43, 48, 49, 61, 62, 64, 68, 70, 71, 75]

Recommendation: The UHDRS-FAS is "suggested" for assessing severity of limitation in functional capacity in HD, because it lacks core clinimetric data, namely, test-retest or inter-rater reliability data.

The UHDRS-Independence Scale (IS)

The UHDRS-IS is a clinician-rated tool which assesses the actual reduction of functional ability.[76] It is rated from 100 (no special care needed) to 0 (tube-fed, total bed care) and takes approximately 5 minutes to complete. It has been used in many observational and randomized controlled trials in manifest HD populations.[34, 41-44, 46, 48-50, 61, 62, 64, 68, 70] The clinimetric data available show that the UHDRS-IS has moderate inter-rater reliability but in a modified

version that compares caregiver report with patient self-report (ICC = 0.71, 95% CI: 0.48, 0.85). [59] Good correlation with other components of the UHDRS, as well as motor disability, cognitive and behavioral deficits has been shown in various studies.[34, 54, 58, 59, 73, 76-79] Data from clinical trials suggest sensitivity of the UHDRS-IS to change over time and across disease stages.[35, 41]

Recommendation: The UHDRS-IS is "suggested" for assessing severity of limitation in functional ability in HD, because reliability data are missing, including test-retest, inter-rater (for clinicians) and internal consistency.

HD Activities of Daily Living (HD-ADL) 17-item

The HD-ADL Scale, which was developed to be used specifically in HD, was modeled after the Scale for Instrumental Activities of Daily Living.[80] It is a 17-item informant-completed instrument on which the informant rates the HD patient's ability to perform specific activities, covering the domains of personal care, household care, work and money, social relationships, and communication. For each item, the patient is rated on a 4-point scale, from normal to severely limited. The total score of the HD-ADL scale ranges from 0 (normal) to 51 (maximal limitation). [53] With exception of one study,[19] the scale has not been used outside the John Hopkins group who developed it. Clinimetric testing show that the HD-ADL has good internal consistency (α = 0.91-0.96).[53] Principal Component Analysis showed that four factors account for 72-74% of the total variance.[53] Convergent validity has been shown between the total score of the HD-ADL and the UHDRS-TFC ($r = -0.89$, $p < 0.001$), as well as all factors except for the domain "family relationships".[53] Multiple correlations have been reported with measures of cognitive impairment or disease duration.[53, 81, 82] The HD-ADL failed to show differences in treatment compared to placebo.[83, 84]

Recommendation: The HD-ADL is "suggested" for assessing severity of limitation in ADL, because studies of the scale's clinimetric properties are lacking, namely for any type of reliability.

Activity-Specific Balance Scale (ABC)

The ABC is a patient-completed scale that measures balance confidence and fear of falling. The ABC can take anywhere between 6 and 30 minutes to complete depending on the patient. Although it is a self-administered scale, a face-to-face interview is recommended.[85] Patients rate their balance confidence on a visual analogue scale ranging from 0 to 100 for each of 16 tasks, with higher scores indicating greater confidence and lower fall risk. The ABC has been widely used in HD,[8, 17, 29] including a modified ABC-UK version adapted for British culture,[86] but normative cut-off scores have not been established. The clinimetric data available show that the ABC has good test-retest reliability (ICC = 0.74 95% CI: 0.58, 1.0),[8] the MDC has been reported to be 27.33.[8] There is good convergent validity with the Mini-BESTest,[29] and the modified ABC-UK can distinguish between non-fallers and fallers in HD (mean score: 77.5 vs. 47.9).[86] While the ABC has been shown to be sensitive to change in one study (after a 9-month multidisciplinary rehabilitation program),[87] no change was reported in two other studies.[8, 17]

Recommendation: The ABC is "suggested" for assessing level of self-reported balance confidence in HD. The use of the ABC is challenged since the lack of insight is a feature of HD.

Rivermead Mobility Index (RMI)

The RMI is an extension of the Rivermead Motor Assessment Gross Function Scale that assesses functional mobility and was initially developed for stroke. The RMI consists of 14 questions about a patient's ability to perform a wide range of activities, from turning over in bed to running, and one observation (standing for 10 seconds without any aid). Questions are answered as "able" (1 point) or "unable" (0 points) and summed to produce a total score, with a higher score reflecting better mobility.[88] Test-retest reliability has been reported in HD (ICC in pre-manifest HD= 0.81; ICC in manifest HD = 0.94).[6] A MDC of 2 points has been reported in manifest HD; ceiling effects are present in pre-manifest HD.[6] There are no cut-off scores established in HD, which limits its use as a screening tool in HD.

Recommendation: The RMI is "suggested" for the assessment of severity of restriction of mobility

Discussion

The current critique focuses on performance-based measures and rating scales assessing functional ability in HD. In the process of developing the protocol for the review, we found a variety of scale constructs and other instruments that could be associated with various aspects of function ability. We used the ICF[3] as a conceptual framework related with function to guide us in the inclusion or exclusion of rating scales based on the adequacy of their constructs. Nevertheless, we realize that the measures included in this review represent a wide variety of concepts that apply across the components of the ICF. Many of these measures included multiple ICF components, raising challenges for conceptual clarity and subsequent evaluation of validity. For example, balance can be seen as a sheer impairment but it can overlap with activity/function, depending how it is captured in a given clinical measure.

Considering these aspects, we decided to be inclusive and included balance measures in this review. Ultimately, there is a need for clear definitions for future measures to better enable validation and application in HD populations.

We identified and included a range of performance-based measures. We provide a "recommended" level of recommendation for both screening purposes related to balance, gait and/or risk of falling, and measurement of severity of impairment of specific motor tasks. There were however no "recommended" performance-based measures covering upper limb function. It is also important to emphasize that the majority of these performance measures were only used in ambulatory HD populations.

We did not identify a rating scale that met the criteria for "recommended". If further testing of the measurement properties is conducted, we agreed that UHDRS sub-scales related with function (TFC, FAS and IS) are in a good position to reach the higher level of recommendation in the future due to their widespread use, specific development in HD and known initial clinimetric development. For each one of these scales, important shortcomings in terms of clinimetric development were identified, namely incomplete reliability testing, which precluded a "recommended" level of recommendation. In addition, these scales have limiting ceiling effects that make them unattractive for use in earlier stages of HD. For example, the use of these UHDRS subscales in a clinical trial conducted with the purpose of capturing a disease-modifying effect in an ideal HD subgroup of individuals with a high level of functional ability would be performed at the cost of a prohibitively long trial duration to capture a meaningful change. Rating scales such as Functional Rating Scale Task force for pre-Huntington Disease 2 (FuRST-pHD)[89, 90] are currently being developed and are expected to fill this gap in the future.

The assessment of functional ability as a clinical outcome is deemed essential for therapeutic approval by regulatory agencies such as the FDA.[1] In this regulatory context, it is important to

emphasize that there was no recommendation for the purpose of measuring change over time in individuals or groups of subjects in either a pure observational study or in an interventional context. In fact, formal testing for responsiveness was missing in all the included rating scales, and important measures of reliability such as test-retest had not been evaluated in many cases. Along the same lines, there is a need to assess the validity of each rating scale in different subgroups of patients with HD, as these data are presently lacking for most of the measures. The knowledge about responsiveness and its variation in important patient subgroups can determine sample size requirements and help with the interpretation of clinical trial results, respectively.[1]

Looking towards the future, the committee concludes that there are well-validated performance-based measures that capture motor tasks such as walking or balance, but further clinimetric development is required for performance-based measures that capture other aspects of physical function such as upper limb function. For rating scales, including those evaluating activities of daily living, we are cannot endorse an existing scale at a "recommended" level and encourage the MDS to prioritize the development of such instruments for clinical care and research purposes. Further validation of HD-specific scales such as the UHDRS-TFC are warranted, as is the development of new scales designed to have greater sensitivity in capturing function in HD subgroups who have a relatively well preserved functional ability as measured by currently available rating scales.

Author Roles:

Tiago A. Mestre

1. Research project: A. Conception, B. Organization, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: A. Writing of the first draft; B. Review and Critique;

Monica Busse

1. Research project: A. Conception, B. Organization, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

Aileen M. Davis

1. Research project: A. Conception, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

Lori Quinn

1. Research project: A. Conception, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

Filipe B.Rodrigues

1. Research project: A. Conception, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

Jean-Marc Burgunder

1. Research project: A. Conception, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

Noelle Carlozzi

1. Research project: A. Conception, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

Francis Walker

1. Research project: A. Conception, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

Aileen Ho

1. Research project: A. Conception, C. Execution;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

Cristina Sampaio

1. Research project: A. Conception, C. Execution;
2. Statistical Analysis: not applicable;

3. Manuscript Preparation: B. Review and Critique;

Christopher G. Goetz

1. Research project: A. Conception;
2. Statistical Analysis: not applicable;
3. Manuscript Preparation: B. Review and Critique;

Esther Cubo

1. Research Project: B. Organization
2. Statistical Analysis: not applicable;
3. Manuscript Prep: B. Review and critique.

Pablo Martinez-Martin

1. Research Project: A. Conception
2. Statistical Analysis: not applicable;
3. Manuscript Prep: B. Review and critique.

Glenn T. Stebbins

4. Research Project: A. Conception
5. Statistical Analysis: not applicable;
6. Manuscript Prep: B. Review and critique.

Financial Disclosures:

Tiago A. Mestre

Financial disclosure related to research covered in this article: Consulting for CHDI Foundation/Management

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: Abbvie, CHDI Foundation/Management

Grants and Research: University of Ottawa Medical Associates, PSG/PDF, Parkinson Canada, Parkinson Research Consortium

Honoraria: International Parkinson and Movement Disorders Society, American Academy of Neurology, U of Ottawa, Abbvie

Intellectual Property Rights: none

Ownership interests: none

Royalties: none

Expert Testimony: none

Salary: University of Ottawa Medical Associates

Monica Busse

Financial disclosure related to research covered in this article: none

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: None

Grants and Research: European Framework funding, Health and Care Research Wales (HCRW), Wellcome Trust, Medical Research Council UK, Gossweiler Foundation, National Institute of Health Research (NIHR)

Honoraria: none

Intellectual Property Rights: none

Ownership interests: none

Royalties: none

Expert Testimony: none

Salary: Cardiff University

Aileen M. Davis

Financial disclosure related to research covered in this article: none

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: Flexion Therapeutics Inc

Grants and Research: none

Honoraria: Associate Editor of Osteoarthritis and Cartilage

Intellectual Property Rights: none

Ownership interests: none

Royalties: none

Expert Testimony: none

Salary: U of Toronto.

Lori Quinn

Financial disclosure related to research covered in this article: reimbursement for travel to meetings by Movement Disorders Society

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: None

Grants and Research: Huntington Study Group; Jacques and Gloria Gossweiller Foundation

Honoraria: None

Intellectual Property Rights: None

Ownership interests: None

Royalties: Elsevier Publishers for textbook Documentation for Rehabilitation: A guide to clinical decision making in physical therapy

Expert Testimony: None

Salary: None

Filipe B. Rodrigues

Financial disclosure related to research covered in this article: None

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: None

Grants and Research: CHDI Foundation Clinical Research Fellowship Award (Aug 2015 to Aug 2016)

Honoraria: None

Intellectual Property Rights: None

Ownership interests: None

Royalties: None

Expert Testimony: None

Salary: University College London

Jean-Marc Burgunder

Financial disclosure related to research covered in this article: None

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: Chair of the EHDN Executive Committee

Grants and Research: no conflicts

Honoraria: no conflicts

Intellectual Property Rights: none

Ownership interests: none

Royalties: none

Expert Testimony: none

Salary: no conflict

Noelle Carlozzi

Financial disclosure related to research covered in this article: None.

Full financial disclosure for the last 12 months:

Consulting and Advisory Board Membership with honoraria: Teva Pharmaceuticals; Boston Medical Center

Grants and Research: National Institute for Neurological Disorders and Stroke; National Institute of Nursing Research; National Institute on Aging; CHDI Foundation; Frankel Cardiovascular Center Micro Grant Award;

Honoraria: None

Intellectual Property Rights: None.

Ownership interests: None

Royalties: None.

Expert Testimony: None

Salary: University of Michigan

Francis Walker

Financial disclosure related to research covered in this article: none

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: none

Grants and Research: Grant support by Pfizer, Vaccinex, and Teva; interest free instrument loans from Monarch Medical, Terason, Natus.

Honoraria: Grifols

Intellectual Property Rights: none

Ownership interests: None

Royalties: Elsevier, Up To Date

Expert Testimony: none

Salary: Wake Forest School of Medicine

Aileen Ho

Financial disclosure related to research covered in this article: none

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: Pfizer, National Institute of Health and Care Excellence UK.

Grants and Research: National Institute of Health Research (NIHR), European Huntington's Disease Network.

Honoraria: none

Intellectual Property Rights: none

Ownership interests: none

Royalties: none

Expert Testimony: none

Salary: University of Reading

Cristina Sampaio

646 **Financial disclosure related to research covered in this article:** Salary: CHDI
647 Management
648 **Full financial disclosure For the last 12 months:**
649 Consulting and Advisory Board Membership with honoraria: I received honoraria from
650 Nestle, vTv Therapeutics, Neurotrope Stealth.
651 Grants and Research: none
652 Honoraria: International Parkinson and Movement Disorders Society
653 Intellectual Property Rights: none
654 Ownership interests: none
655 Royalties: none
656 Expert Testimony: none
657 Salary: CHDI Management
658
659 **Christopher G. Goetz**
660 **Financial disclosure related to research covered in this article: None**
661 **Full financial disclosure for the last 12 months:**
662 **Consulting or Advisory Board Membership with honoraria**: Addex, Avanir, Boston
663 Scientific, CHDI Foundation/CHDI management, Clevexel, Kanter Health, Oxford
664 Biomedica, Pfizer, WebMD.
665 **Grants/Research**: Funding to Rush University Medical Center from NIH, Michael J. Fox
666 Foundation for research conducted by Dr. Goetz. Dr. Goetz directs the Rush Parkinson's
667 Disease Research Center that receives support from the Parkinson's Disease Foundation and
668 some of these funds support Dr. Goetz's salary as well as his research efforts. He directs the
669 translation program for the MDS-UPDRS and UDysRS and receives funds directed to Rush

University Medical Center from the International Parkinson and Movement Disorder Society (IPMDS) for this effort.

Honoraria: American Academy of Neurology, Captain James A Lovell Federal Health Care Center, University of Pennsylvania, University of Rochester

Intellectual Property Rights: none

Ownership interests: none

Royalties: Elsevier Publishers, Oxford University Press, Wolters Kluwer,

Salary: Rush University Medical Center

Esther Cubo

Financial disclosure related to research covered in this article: none.

Full financial disclosure For the last 12 months:

Consulting and Advisory Board Membership with honoraria: Abbvie, Allergan

Grants and Research: Junta de Castilla y León, International Parkinson disease and Movement Disorder Society

Honoraria: none

Intellectual Property Rights: none

Ownership interests: none

Royalties: none

Expert Testimony: none

Salary: Hospital Universitário Burgos, Spain**.**

Pablo Martinez-Martin

Financial disclosure related to research covered in this article: None

695

696 **Full financial disclosure For the last 12 months:**

697 Consulting and Advisory Board Membership with honoraria: Abbvie.

698 Grants and Research: International Parkinson and Movement Disorder Society for the Pilot
699 Study of the MDS-Non-Motor Symptoms Scale.

700 Honoraria: Editorial Viguera and Movement Disorder Society; AbbVie

701 Intellectual Property Rights: None

702 Ownership interests: None

703 Royalties: None

704 Expert Testimony: None

705 Salary: Carlos III Institute of Health

706

707 **Glenn T. Stebbins**

708 **Financial disclosure related to research covered in this article:** None

709 **Full financial disclosure For the last 12 months:**

710 Consulting and Advisory Board Membership with honoraria: Acadia, Pharmaceuticals,
711 Adamas Pharmaceuticals, Inc., Ceregene, Inc., CHDI Management, Inc., Ingenix
712 Pharmaceutical Services (i3 Research), Neurocrine Biosciences, Inc., Pfizer, Inc., Ultragenyx
713 Pharmaceutical.

714 Grants and Research: National Institutes of Health, Michael J. Fox Foundation for
715 Parkinson's Research, Dystonia Coalition, CHDI, International Parkinson and Movement
716 Disorder Society, CBD Solutions.

717 Honoraria: International Parkinson and Movement Disorder Society, American Academy of
718 Neurology, Michael J. Fox Foundation for Parkinson's Research, Food and Drug
719 Administration.

720 Intellectual Property Rights: none

721

722 Ownership interests: none

723 Royalties: none

724 Expert Testimony: none

725 Salary: Rush University Medical Center

726

727

Acknowledgments

We would like to thank Anne-Marie Williams for the editorial support, and Theresa Bolton for the assistance in literature search of the current review.

Ethical Compliance Statement: The authors confirm that the approval of an institutional review board was not required for this work. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this work is consistent with those guidelines.

References

1. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;4:79.
2. Schrag A, Barone P, Brown RG, et al. Depression rating scales in Parkinson's disease: critique and recommendations. Mov Disord 2007;22:1077-1092.
3. Towards a Common Language for Functioning, Disability and Health: ICF, The International Classification of Functioning, Disability and Health. Geneva; 2002.
4. Tinetti ME, Williams TF, Mayewski R. Fall risk index for elderly patients based on number of chronic disabilities. Am J Med 1986;80:429-434.
5. Kloos AD, Kegelmeyer DA, Young GS, Kostyk SK. Fall risk assessment using the Tinetti mobility test in individuals with Huntington's disease. Mov Disord 2010;25:2838-2844.
6. Quinn L, Khalil H, Dawes H, et al. Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. Phys Ther 2013;93:942-956.
7. Busse M, Quinn L, Khalil H, McEwan K. Optimising mobility outcome measures in Huntington's disease. J Huntingtons Dis 2014;3:175-188.
8. Kloos AD, Fritz NE, Kostyk SK, Young GS, Kegelmeyer DA. Clinimetric properties of the Tinetti Mobility Test, Four Square Step Test, Activities-specific Balance Confidence Scale, and spatiotemporal gait measures in individuals with Huntington's disease. Gait Posture 2014;40:647-651.
9. Zinzi P, Salmaso D, De Grandis R, et al. Effects of an intensive rehabilitation programme on patients with Huntington's disease: a pilot study. Clin Rehabil 2007;21:603-613.
10. Kegelmeyer DA, Kloos AD, Fritz NE, Fiumedora MM, White SE, Kostyk SK. Impact of tetrabenazine on gait and functional mobility in individuals with Huntington's disease. J Neurol Sci 2014;347:219-223.
11. Kloos AD, Fritz NE, Kostyk SK, Young GS, Kegelmeyer DA. Video game play (Dance Dance Revolution) as a potential exercise therapy in Huntington's disease: a controlled clinical trial. Clinical rehabilitation 2013;27:972-982.
12. Bohlen S, Ekwall C, Hellstrom K, et al. Physical therapy in Huntington's disease--toward objective assessments? Eur J Neurol 2013;20:389-393.
13. Busse ME, Khalil H, Quinn L, Rosser AE. Physical therapy intervention for people with Huntington disease. Phys Ther 2008;88:820-831.
14. Fekete R, Davidson A, Jankovic J. Clinical assessment of the effect of tetrabenazine on functional scales in huntington disease: a pilot open label study. Tremor Other Hyperkinet Mov (N Y) 2012;2.
15. Ferrara JM, Mostile G, Hunter C, Adam OR, Jankovic J. Effect of tetrabenazine on motor function in patients with huntington disease. Neurol Ther 2012;1:5.
16. Khalil H, Quinn L, van Deursen R, et al. What effect does a structured home-based exercise programme have on people with Huntington's disease? A randomized, controlled pilot study. Clinical rehabilitation 2013;27:646-658.
17. Piira A, van Walsem MR, Mikalsen G, Nilsen KH, Knutsen S, Frich JC. Effects of a One Year Intensive Multidisciplinary Rehabilitation Program for Patients with Huntington's Disease: a Prospective Intervention Study. PLoS Curr 2013;5.
18. Quinn L, Debono K, Dawes H, et al. Task-specific training in Huntington disease: a randomized controlled feasibility trial. Phys Ther 2014;94:1555-1568.

786 19. Rao AK, Muratori L, Louis ED, Moskowitz CB, Marder KS. Clinical measurement of
787 mobility and balance impairments in Huntington's disease: validity and responsiveness. Gait
788 Posture 2009;29:433-436.
789 20. Balke B. A Simple Field Test for the Assessment of Physical Fitness. Rep 63-6. Rep
790 Civ Aeromed Res Inst US 1963:1-8.
791 21. Butland RJ, Pang J, Gross ER, Woodcock AA, Geddes DM. Two-, six-, and 12-
792 minute walking tests in respiratory disease. Br Med J (Clin Res Ed) 1982;284:1607-1608.
793 22. Guyatt GH, Thompson PJ, Berman LB, et al. How should we measure function in
794 patients with chronic heart and lung disease? J Chronic Dis 1985;38:517-524.
795 23. Guyatt GH, Pugsley SO, Sullivan MJ, et al. Effect of encouragement on walking test
796 performance. Thorax 1984;39:818-822.
797 24. Busse M, Quinn L, Debono K, et al. A randomized feasibility study of a 12-week
798 community-based exercise program for people with Huntington's disease. J Neurol Phys Ther
799 2013;37:149-158.
800 25. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility
801 for frail elderly persons. J Am Geriatr Soc 1991;39:142-148.
802 26. Quinn L, Debono K, Dawes H, et al. Task-specific training in Huntington's disease: A
803 randomised, controlled feasibility trial. In; 2014. p. A66-a67.
804 27. Rao AK, Louis ED, Marder KS. Clinical assessment of mobility and balance
805 impairments in pre-symptomatic Huntington's disease. Gait Posture 2009;30:391-393.
806 28. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric
807 techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. J Rehabil
808 Med 2010;42:323-331.
809 29. Jacobs JV, Boyd JT, Hogarth P, Horak FB. Domains and correlates of clinical balance
810 impairment associated with Huntington's disease. Gait Posture 2015;41:867-870.
811 30. Reuben DB, Siu AL. An objective measure of physical function of elderly outpatients.
812 The Physical Performance Test. J Am Geriatr Soc 1990;38:1105-1112.
813 31. Farrell MK, Rutt RA, Lusardi MM, Williams AK. Reliability of the Physical
814 Performance Test in People with Dementia. Physical & Occupational Therapy In Geriatrics
815 2010;28:144-153.
816 32. Sharpened Romberg. 2013;
817 http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1160 Last accessed:
818 6 February 2017.
819 33. Romberg Test. 2013;
820 http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1173 Last accessed:
821 6 February 2017.
822 34. Huntington Study Group. Unified Huntington's Disease Rating Scale: reliability and
823 consistency. Mov Disord 1996;11:136-142.
824 35. Shoulson I. Huntington disease: functional capacities in patients treated with
825 neuroleptic and antidepressant drugs. Neurology 1981;31:1333-1335.
826 36. Shoulson I, Odoroff C, Oakes D, et al. A controlled clinical trial of baclofen as
827 protective therapy in early Huntington's disease. Annals of neurology 1989;25:252-259.
828 37. Feigin A, Kieburtz K, Bordwell K, et al. Functional decline in Huntington's disease.
829 Mov Disord 1995;10:211-214.
830 38. Como PG, Rubin AJ, O'Brien CF, et al. A controlled trial of fluoxetine in
831 nondepressed patients with Huntington's disease. Movement disorders 1997;12:397-401.
832 39. Siesling S, van Vugt JP, Zwinderman KA, Kieburtz K, Roos RA. Unified Huntington's
833 disease rating scale: a follow up. Mov Disord 1998;13:915-919.
834 40. Kremer B, Clark CM, Almqvist EW, et al. Influence of lamotrigine on progression of
835 early Huntington disease: a randomized clinical trial. In; 1999. p. 1000-1011.

41. Marder K, Zhao H, Myers RH, et al. Rate of functional decline in Huntington's disease. Huntington Study Group. Neurology 2000;54:452-458.
42. Huntington Study Group. A randomized, placebo-controlled trial of coenzyme Q10 and remacemide in Huntington's disease. Neurology 2001;57:397-404.
43. Huntington Study Group. Dosage effects of riluzole in Huntington's disease: A multicenter placebo-controlled study. Neurology 2003;61:1551-1556.
44. Bonelli RM, Hodl AK, Hofmann P, Kapfhammer HP. Neuroprotection in Huntington's disease: A 2-year study on minocycline. International Clinical Psychopharmacology 2004;19:337-342.
45. Tommaso M, Specchio N, Sciruicchio V, Difruscolo O, Specchio LM. Effects of rivastigmine on motor and cognitive impairment in Huntington's disease. In; 2004. p. 1516-1518.
46. Huntington Study G. Minocycline safety and tolerability in Huntington disease. Neurology 2004;63:547-549.
47. de Tommaso M, Di Fruscolo O, Sciruicchio V, et al. Efficacy of levetiracetam in Huntington disease. Clin Neuropharmacol 2005;28:280-284.
48. Puri BK, Leavitt BR, Hayden MR, et al. Ethyl-EPA in Huntington disease: a double-blind, randomized, placebo-controlled trial. In; 2005. p. 286-292.
49. Cubo E, Shannon KM, Tracy D, et al. Effect of donepezil on motor and cognitive function in Huntington disease. Neurology 2006;67:1268-1271.
50. Landwehrmeyer GB, Dubois B, De Yebenes JG, et al. Riluzole in Huntington's disease: A 3-year, randomized controlled study. Annals of Neurology 2007;62:262-272.
51. Tommaso M, Difruscolo O, Sciruicchio V, Specchio N, Livrea P. Two years' follow-up of rivastigmine treatment in Huntington disease. In; 2007. p. 43-46.
52. Carlozzi NE, Victorson D, Sung V, et al. HD-PRO-TRIAD Validation: A Patient-reported Instrument for the Symptom Triad of Huntington's Disease. Tremor Other Hyperkinet Mov (N Y) 2014;4:223.
53. Bylsma. Assessment of Adaptive Functioning in Huntington's Disease. Mov Disord 1993;8:183-190.
54. Siesling S, Zwinderman AH, van Vugt JP, Kieburtz K, Roos RA. A shortened version of the motor section of the Unified Huntington's Disease Rating Scale. Mov Disord 1997;12:229-234.
55. Thompson JC, Snowden JS, Craufurd D, Neary D. Behavior in Huntington's disease: dissociating cognition-based and mood-based changes. J Neuropsychiatry Clin Neurosci 2002;14:37-43.
56. Ready RE, Mathews M, Leserman A, Paulsen JS. Patient and caregiver quality of life in Huntington's disease. Mov Disord 2008;23:721-726.
57. Ho AK, Gilbert AS, Mason SL, Goodman AO, Barker RA. Health-related quality of life in Huntington's disease: Which factors matter most? Mov Disord 2009;24:574-578.
58. Youssov K, Dolbeau G, Maison P, et al. Unified Huntington's disease rating scale for advanced patients: validation and follow-up study. Mov Disord 2013;28:1717-1723.
59. Carlozzi NE, Tulsky DS, Chiaravalloti ND, et al. NIH Toolbox Cognitive Battery (NIHTB-CB): the NIHTB Pattern Comparison Processing Speed Test. J Int Neuropsychol Soc 2014;20:630-641.
60. Klempir J, Klempirova O, Spackova N, Zidovska J, Roth J. Unified Huntington's disease rating scale: clinical practice and a critical approach. Funct Neurol 2006;21:217-221.
61. Ravina B, Romer M, Constantinescu R, et al. The relationship between CAG repeat length and clinical progression in Huntington's disease. Mov Disord 2008;23:1223-1227.
62. Huntington Study Group DI. A futility study of minocycline in Huntington's disease. Movement Disorders 2010;25:2219-2224.

886 63. Huntington Study Group. A randomized, placebo-controlled trial of coenzyme Q10
887 and remacemide in
888 Huntington's disease. Neurology 2001;57:397-404.
889 64. Kieburtz K, McDermott MP, Voss TS, et al. A randomized, placebo-controlled trial of
890 latrepirdine in Huntington disease. In; 2010. p. 154-160.
891 65. Tabrizi SJ, Scahill RI, Durr A, et al. Biological and clinical changes in premanifest
892 and early stage Huntington's disease in the TRACK-HD study: the 12-month longitudinal
893 analysis. The Lancet Neurology 2011;10:31-42.
894 66. Tabrizi SJ, Reilmann R, Roos RAC, et al. Potential endpoints for clinical trials in
895 premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month
896 observational data. The Lancet Neurology 2012;11:42-53.
897 67. Tabrizi SJ, Scahill RI, Owen G, et al. Predictors of phenotypic progression and disease
898 onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis
899 of 36-month observational data. The Lancet Neurology 2013;12:637-649.
900 68. Verbessem P, Lemiere J, Eijnde BO, et al. Creatine supplementation in Huntington's
901 disease: a placebo-controlled pilot trial. In; 2003. p. 925-930.
902 69. Beglinger LJ, Adams WH, Langbehn D, et al. Results of the citalopram to enhance
903 cognition in Huntington disease trial. Mov Disord 2014;29:401-405.
904 70. Sussmuth SD, Haider S, Landwehrmeyer GB, et al. An exploratory double-blind,
905 randomized clinical trial with selisistat, a SirT1 inhibitor, in patients with Huntington's
906 disease. British journal of clinical pharmacology 2015;79:465-476.
907 71. Vaddadi KS, Soosai E, Chiu E, Dingjan P. A randomised, placebo-controlled, double
908 blind study of treatment of Huntington's disease with unsaturated fatty acids. In; 2002. p. 29-
909 33.
910 72. Mahant N, McCusker EA, Byth K, Graham S. Huntington's disease: clinical correlates
911 of disability and progression. Neurology 2003;61:1085-1092.
912 73. Tumas V, Camargos ST, Jalali PS, Galesso Ade P, Marques Jr W. Internal consistency
913 of a Brazilian version of the unified Huntington's disease rating scale. Arq Neuropsiquiatr
914 2004;62:977-982.
915 74. Banaszkiewicz K, Sitek EJ, Rudzinska M, Soltan W, Slawek J, Szczudlik A.
916 Huntington's disease from the patient, caregiver and physician's perspectives: three sides of
917 the same coin? J Neural Transm 2012;119:1361-1365.
918 75. Huntington Study G. Tetrabenazine as antichorea therapy in Huntington disease: a
919 randomized controlled trial. Neurology 2006;66:366-372.
920 76. Myers RH, Sax DS, Schoenfeld M, et al. Late onset of Huntington's disease. J Neurol
921 Neurosurg Psychiatry 1985;48:530-534.
922 77. Myers RH, Sax DS, Koroshetz WJ, et al. Factors associated with slow progression in
923 Huntington's disease. Arch Neurol 1991;48:800-804.
924 78. Ho AK, Robbins AO, Walters SJ, Kaptoge S, Sahakian BJ, Barker RA. Health-related
925 quality of life in Huntington's disease: a comparison of two generic instruments, SF-36 and
926 SIP. Mov Disord 2004;19:1341-1348.
927 79. Ho AK, Robbins AO, Barker RA. Huntington's disease patients have selective
928 problems with insight. Mov Disord 2006;21:385-389.
929 80. Lawton MP. The functional assessment of elderly people. J Am Geriatr Soc
930 1971;19:465-481.
931 81. Brandt J, Strauss ME, Larus J, Jensen B, Folstein SE, Folstein MF. Clinical correlates
932 of dementia and disability in Huntington's disease. J Clin Neuropsychol 1984;6:401-412.
933 82. Rothlind JC, Brandt J. A brief assessment of frontal and subcortical functions in
934 dementia. J Neuropsychiatry Clin Neurosci 1993;5:73-77.

83. Peyser CE, Folstein M, Chase GA, et al. Trial of d-alpha-tocopherol in Huntington's disease. American journal of psychiatry 1995;152:1771-1775.
84. Ranen NG, Peyser CE, Coyle JT, et al. A controlled trial of idebenone in Huntington's disease. Mov Disord 1996;11:549-554.
85. Powell LE, Myers AM. The Activities-specific Balance Confidence (ABC) Scale. J Gerontol A Biol Sci Med Sci 1995;50A:M28-34.
86. Busse ME, Wiles CM, Rosser AE. Mobility and falls in people with Huntington's disease. J Neurol Neurosurg Psychiatry 2009;80:88-90.
87. Thompson JA, Cruickshank TM, Penailillo LE, et al. The effects of multidisciplinary rehabilitation in patients with early-to-middle-stage Huntington's disease: a pilot study. Eur J Neurol 2013;20:1325-1329.
88. Collen FM, Wade DT, Robb GF, Bradshaw CM. The Rivermead Mobility Index: a further development of the Rivermead Motor Assessment. Int Disabil Stud 1991;13:50-54.
89. FuRST 2.0: Cognitive Pre-Testing Study for a New Functional Rating Scale for Use in Huntington's Disease. 2017; https://clinicaltrials.gov/ct2/show/NCT02881931?term=Furst-&rank=1 Last accessed: 01/24/2017.
90. Vaccarino AL, Sills T, Anderson KE, et al. Assessment of Day-to-Day Functioning in Prodromal and Early Huntington Disease. PLoS Curr 2011;3:RRN1262.