**Summary Statement for Cognitive CDEs**

This summary is the companion document to the NINDS Parkinson’s disease CDE Working Group – Cognitive Table (see Table 1: NINDS CDE Workgroup – Cognitive). The goal of the NINDS CDE Cognitive subgroup was to recommend scales that can be readily used in data collection for clinical trials that have broad applicability to all stages of PD. In the following sections, the Cognitive CDE group has endeavored to provide an objective assessment of 11 published scales including 3 PD-specific and 8 non-specific scales that may be used in clinical trials. All scales that were available in the English literature were reviewed.

The decision to provide ratings of global scales rather than individual neuropsychological tests or specific cognitive domains was based on several considerations. There are two basic approaches in cognitive testing: battery versus individual tests. Both approaches are widely applicable to PD. A battery is a collection of individual tests that assess different domains whereas individual tests can be used to look at either a single domain or subcomponents of a domain. We have chosen the battery approach for the following reasons:

First, we have determined that it was beyond the scope of this panel to rate specific neuropsychological tests. As part of the process for the creation of the NIH Toolbox, an assessment tool that has been developed for the assessment of cognitive, motor, sensory and emotional domains in individuals aged 3 – 85 years, a library of extant instruments was created for each of these major domains. The purpose of this information was to determine what the gold-standard instruments were for the specific domains; which instruments had intellectual property constraints for use; what pros and cons might be present that could influence including the instrument in the Toolbox; as well as what instruments would need to be developed de novo. This library is currently available for general access on the web (www.nihtoolbox.org). For the cognition domain, a battery of instruments has been assembled that measures the subdomains of attention, episodic memory, executive function including cognitive flexibility and inhibitory control, vocabulary comprehension, processing speed, and working memory. Concurrently, the entire set of Toolbox measures is being validated and normed.

Second, the choice of cognitive assessment battery depends on the stage/severity of PD, and the study design. Different batteries may be necessary when cognition is a primary endpoint, when cognition is an exploratory/secondary issue and where cognitive impairment may be an exclusion factor to study participation.

Third, there is no single accepted definition for PD dementia or PD-related Mild Cognitive Impairment. DSM-V was published in 2013, replacing the DSM-IV-TR. In the DSM-V, the diagnostic term, Dementia was replaced by "major neurocognitive disorder" (equivalent to dementia) and "mild neurocognitive disorder" (equivalent to mild cognitive impairment [MCI]). The diagnostic term, mild neurocognitive disorder, was a new addition to the DSM-V. For Parkinson's disease, these terms are listed as "major or minor neurocognitive disorder due to Parkinson's disease". In addition, dementia with Lewy bodies (DLB) is a distinct diagnosis in DSM-V ("major or mild neurocognitive disorder with Lewy bodies"). The Movement Disorder Society (MDS) has also published criteria for dementia (Emre et al., 2014) that require both cognitive and functional decline but do not require that the memory domain is impaired.  The MDS has published criteria for PD-MCI (Litvan et al., 2011) and there have been several publications to validate this construct.  In addition, several publications have focused on the validation of the scales that the Task Force suggested.  We recommend that a separate panel review the four usage categories defined in ‘1’ below, specifically for PD-MCI. This guideline is meant for dementia only. Rather than be prescriptive, we will not recommend scales as core or non-core.  We have provided three levels of information including:

1. NINDS CDE Workgroup-Cognitive Recommended CDEs (Table 1: NINDS CDE Workgroup – Cognitive): Each scale was rated based on its utility in each of four usage categories: (i) as a screening instrument for initial identification of a possible disorder or population of interest, (ii) as a rating scale for measurement of severity of the disorder or performance within the domain of interest, (iii) as a scale that is sensitive to longitudinal change, and (iv) as a diagnostic instrument that can categorize individuals into those with and without a disorder. Each category was rated based on a consensus among panel members using a score of 1 to 3 (1 highest or best, 3 lowest or worst) or not available. The final column indicates the administration time for the instrument that varies based on the level of cognitive impairment.
2. Detailed summaries were provided for each scale including:
	1. Short description of the instrument including constructs measured (domains), generic vs. disease specific, means of administration, intended respondent, number of items, subscales, and items per subscale, and scoring.
	2. Background
	3. Key references
	4. Rationale including: Strengths/weaknesses, psychometric properties, purpose of tool, sensitivity to change, usage, availability. Further information on definitions of psychometric properties can be found in **Appendix A**.

3. Future Directions

* Collaborative effort to determine the domains appropriate for a consensus cognitive battery for the spectrum of cognitive impairment in PD that is evidence–based.
* Multicenter, comparative studies of existing instruments:
	+ Different scales (global vs. PD specific)
	+ Different disease groups
	+ Different race/ethnic groups
* Carrying out psychometric analyses of the ability to assess responsiveness to change so that scales can be used in intervention studies to assess outcomes and in longitudinal descriptive studies.
* Carrying out psychometric analyses designed to develop “harmonizing rules” that allow assessments obtained with one measure to be translated into units of another.
* Review of instruments specifically for PD-MCI.
* Formation of a Task Force to explore development of new scales to meet the recommendations of the evidence-based review.
* There are now several computerized cognitive batteries. We recommend a separate panel to review these batteries to assess cognition in PD. As an example of reviews for Computerized Cognitive Batteries, we have included the CANTAB and Cognitive Drug Research (CDR) both of which have been used in PD.

**References:**

Emre M, Ford PJ, Bilgiç B, Uç EY. Cognitive impairment and dementia in Parkinson's disease: practical issues and management. Mov Disord. 2014;29(5):663-672.

Litvan I, Aarsland D, Adler CH, Goldman JG, Kulisevsky J, Mollenhauer B, Rodriguez-Oroz MC, Tröster AI, Weintraub D. MDS Task Force on mild cognitive impairment in Parkinson's disease: critical review of PD-MCI. Mov Disord. 2011;26(10):1814-1824.

Table 1: NINDS CDE Workgroup – Cognitive

Recommended CDE’s

|  |  |  |
| --- | --- | --- |
| **Instrument Name**  | **Rating Scale Usage** |  |
| Screening Instrumenta | Severityb | Longitudinalc | Diagnostic Instrumentd | Administration Time |
| ADAS COG◊ | 2 | 2 | N/A | 2 | 30 min |
| Addenbrooke’s Cognitive Examination-III ACE-III | 1.5 | 2 | 3 | 2 | 15-20 min |
| CAMCOG-R† | 2 | 2 | 2 | 2 | 25-30 min |
| Clinician Global Impression of Change (CGIC)◊ | N/A | N/A | 1.5 | N/A | 40 min |
| Mattis DRS-2□ | 1 | 1 | 1 | 1 | 30-50 min |
| MMSE□ | 3 | 2.5 | 2 | 3 | 10-12 min |
| Montreal Cognitive Assessment (MoCA)▲ | 1.5 | 1.5 | 1 | 2 | 10-15 min |
| PANDA† | 2 | 2 | N/A | 2.5 | 8-20 min |
| PD-CRS† | 2 | 2 | N/A | 2 | 17-26 min |
| SCOPA-COG◊ | 2 | 1.25 | 2 | 3 | 10-15 min |
| Short Portable Mental Status Questionnaire▲ | 3 | N/A | N/A | N/A | 5 min |

a Screening Instrument - For initial identification of possible disorder;

b Rating Scale - For measurement of disorder severity

c Longitudinal - Sensitivity to change over time

d Diagnostic Criteria and Instrument – Categorization of patients into those with and without a disorder

† Instrument available from author– refer to Instrument Summary for information on availability

◊ Instrument available in public domain – refer to Instrument Summary for information on availability

 ▲ Free to investigators– refer to Instrument Summary for information on availability

□ Copyrighted Instrument – refer to Instrument Summary for information on availability

 **Please note**: Each of the above scales are being given a score of 1, 2, 3 for suitability (1= highest or best, 3= lowest or worst)

**Appendix A: NINDS PD CDEs – Cognitive Subgroup Psychometric Validation Scale**

Below are some musings on psychometric – clinimetric assessments for scale reliability and validity. There are other more esoteric tests (e.g., differential item function), but those are really outside the scope of these efforts. For basic definitions, reliability is defined as how free from random (and non-random) error a measure is and validity is defined as whether or not a scale measures what it was designed to measure. An important concept is that the validity coefficient (ranging from 0 to 1) can never exceed the reliability coefficient squared (ranging from 0 to 1). So you can have a reliable scale that is not valid, but you cannot have a valid scale that is not reliable.

 Below each description is a listing of whether the assessment may be influenced by patient characteristics. All of these assessments of group influence are based on psychometric theory and may break down in clinical practice.

**Reliability**

1. Internal consistency - Estimation of the correlation among the variables comprising the scale across all possible combination of variables. Typically assessed with Cronbach's alpha.

 Independent of group characteristics

2. Split-half reliability – Estimation of the correlation of two half-forms of the scale across all possible halves. Typically assessed with the Spearman-Brown coefficient or Guttman’s split-half reliability coefficient.

 Independent of group characteristics

3. Alternate form reliability – Estimation of equivalent measures across various forms of the measurement scale. Typically assessed with correlation coefficient.

 Independent of group characteristics

4. Test-retest reliability - Estimation based on the correlation between two (or more) administrations of the same item, scale, or instrument for different times, locations, or populations, when the two administrations do not differ on other relevant variables. Typically assessed using the Spearman Brown coefficient, kappa, generalized kappa for more than 2 time points.

 May be influenced by group characteristics

5. Inter(Intra)-scorer reliability - Estimation based on the correlation of scores between/among two or more examiners who examine the same patient with the same scale. Typically assessed with intra-class correlation, kappa, generalized kappa.

 May be influenced by group characteristics

**Validity**

1. Content validity – Estimation of the extent to which a measurement reflects a specific domain of interest. Typically assessed with “Face Validity” (does the measure seem to cover the domain of interest) or with a Content Validity Ratio (CVR provides a metric for assessing the agreement among raters or judges regarding how essential a particular item is to a given domain).

 Independent of group characteristics

2. Construct validity – Estimation as to the adequacy of a measure to assess a given domain. Typically assessed for both convergent and discriminate construct validity. Convergent = does the scale correlate with measures of theoretically associated domains? Discriminate = the scale does not correlate with measures of theoretically associated domains. Typically assessed with multi-trait-multi-method analysis.

2.b. Factor validity – A form of construct validity that estimates adequacy in providing domain-pure assessments. Typically assessed with exploratory factor analysis and confirmatory factor analysis.

 Construct validity may be influenced by group characteristics

3. Criterion related validity – Estimation of correlation between a measured scale and externally assessed “gold standard.” Typically assessed with correlation coefficient, Etta, receiver-operator characteristics analysis.

3.b. Predictive validity – A form of criterion validity that provides an estimation of a measure’s ability to predict criterion behaviors external to the measuring instrument. Typically assessed with correlation coefficient, Etta, logistic regression.

 Criterion validity may be influenced by group characteristics

4. Sensitivity to change – Estimation of the measure’s sensitivity to changes in the criterion behaviors. Typically assessed using the minimal clinically important change (MCIC) method with ROC analyses.

 May be influenced by group characteristics