Introduction. Multiple sclerosis (MS), the most disabling chronic disorder of young adults, is associated with cognitive deficits that include problems in auditory/verbal memory, visual/spatial memory, cognitive processing speed/working memory, and executive control/function. Given that an extensive battery to assess these varied domains is not feasible, we proposed tests that are brief, relatively straightforward to score, and psychometrically sound regarding sensitivity, reliability, and validity.

This document is organized into four sections. The first goal was to identify one measure which could be included as part of the **CORE CDE for MS**. For circumstances where additional time is available or a more in-depth assessment is desired, we offer measures appropriate for assessment of each core cognitive domain affected by MS, a suggested **SUPPLEMENTAL CDE for MS**. Finally, after reviewing **CONFOUNDS** to test interpretation, we comment on topics or tasks to be regarded as **EXPLORATORY**.

**Part 1: Recommendation for Final or CORE CDE for Multiple Sclerosis.** The test that best meets our highest psychometric standards for reliability and validity is the Symbol Digit Modalities Test (SDMT). Since its initial publication, other versions of SDMT have been described, including the Rao version proposed for this MS CDE. The Rao SDMT excludes the customary written-response administration that precedes the oral-response administration in the original standardization of the test. The rationale is to limit the potential impact of confounds such as upper extremity weakness and ataxia. The Rao SDMT has excellent reliability and discriminative validity and fair to good predictive validity. It is very easy to administer by research staff and requires usually less than 5 min. It is the only test where there are data on cut-off scores that identify clinically meaningful change. For example, in recent work a 4-5 point drop was associated with clinical relapses or in patients losing capacity to work over five years. An additional reason for including the SDMT as a CORE CDE is that information processing speed has been proposed as the pivotal cognitive deficit in MS that underlies all others.

One weakness of SDMT in its current form is uncertainty regarding the equivalence of alternate test versions, which could potentially mitigate practice (ie learning) effects in controlled conditions, especially when the test-retest interval is short. Fortunately, research is underway that will determine equivalent alternate test forms [Benedict et al in progress]. Overall, we find very strong support for SDMT and recommend inclusion of SDMT as a multiple sclerosis CDE.
Part 2: Recommendation for SUPPLEMENTAL Tests that should be strongly considered when MS studies emphasize cognitive function. For the domain of auditory/verbal memory, we recommend the California Verbal Learning Test Second Edition [CVLT2]. The CVLT2 has good test-retest reliability in healthy controls and MS patients as well as good discriminative validity. While specific raw score changes that are clinically meaningful are not well defined, CVLT2 does have significant correlation with employment status. One drawback is its length, and thus we recommend that only Trials 1-5 be administered and scored if reducing time and expense are paramount in a proposed study or clinical project. There are few if any data showing superior psychometrics for the delayed recall trail. Alternatively, the Rey Auditory Verbal Learning Test [RAVLT] is an option and is possibly superior to the CVLT2, as the stimuli are not semantically related to one another. Although the RAVLT has been effectively studied in other disease groups few psychometric data are available in MS and hence it is not recommended for inclusion in the current CDE initiative. However, additional research with the RAVLT is strongly encouraged.

For the domain of visual/spatial memory, we recommend the Brief Visuospatial Memory Test Revised [BVMTR]. The BVMTR is better characterized psychometrically than the 7/24, and 10/36 Spatial Recall tasks, which are perhaps better known. BVMTR has strong psychometric properties and is easy to administer. With respect to practical application, one disadvantage is that for some patients the BVMTR will not be valid if the patient has severe UE ataxia or weakness. In such cases we suggest the copy trial of the BVMTR be administered at the end of the task and used to correct for ataxia as much as possible [see manual]. While this measure has great merit, a drawback applicable to all visual memory tests is their questionable clinical relevance as visual memory is not usually a cause of work disability and other key instrumental activities of daily living.

For the domain of processing speed/working memory we reviewed the SDMT and the Paced Auditory Serial Addition Test [PASAT]. The versions of both tests are included in the Brief Repeatable Battery (BRB) and are reliable and have good validity. There was consensus that the SDMT is superior to the PASAT based on the very strong and documented reliability, sensitivity, and validity findings. For example, correlations with cerebral neuroimaging abnormalities were highest for SDMT. However, the PASAT is better studied than the SDMT within the context of clinical trials of disease modifying therapies.

In the area of executive control/function, we identified three key tests, the Controlled Oral Word Association Test [COWAT], variants of the Stroop conflict paradigm, and the Sorting Test from the Delis Kaplan Executive Function System [DKEFS Sorting Test]. Other tests were discussed. For example, the Wisconsin Card Sorting Test has appeared in many studies and is used frequently in other diseases, but it is a one solution task, and is obviously invalid after the first administration. There are other executive function tasks/tests with the same caveat, for example Tower of London [or other tower tests] and Booklet Category Test. COWAT is recommended because it is reliable and sensitive in most studies [cf Benedict for an exception] and was thus included in the original Rao Screening Battery for MS. It is also very brief and easy to administer. It is limited by language constraints. There are many versions of the Stroop task which hinders recommendation of a specific, validated test. The Denny and Lynch group have published widely on a computerized version and the Amato group has an Italian version, whereas the more frequently utilized versions in US clinical practice are the Golden Stroop Color/Word Test and the DKEFS Color/Word Interference Test. Several studies have shown that the speed, not response inhibition,
aspect of the Stroop task is most relevant in MS.\textsuperscript{41} Because of this lack of firm psychometric foundation for any one version of the Stroop paradigm, we are cautious about recommending it for the CDE. Finally Beatty\textsuperscript{44-45} and Benedict's work with the DKEFS Sorting Test has strongly supported its validity, but there are little data on test-retest and alternate-form reliability. There are not enough data to make firm conclusions either way, but the meager data available are not particularly encouraging.\textsuperscript{37}

Based on these considerations we have proposed a brief core battery that includes the best available tests with respect to psychometric properties for each of the major cognitive domains affected by MS. The battery, presented in the recommended order of administration, takes approximately 55 minutes to administer:

<table>
<thead>
<tr>
<th>Test</th>
<th>Domain</th>
<th>Time</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVLT2 Learning Trials 1-5</td>
<td>auditory/verbal memory</td>
<td>10 min</td>
<td>Primary</td>
</tr>
<tr>
<td>BVMTR Learning Trials 1-3</td>
<td>visual/spatial memory</td>
<td>05 min</td>
<td>Primary</td>
</tr>
<tr>
<td>Rao PASAT 3.0 ISI</td>
<td>processing speed and working memory</td>
<td>05 min</td>
<td>Secondary</td>
</tr>
<tr>
<td>Rao SDMT</td>
<td>processing speed and working memory</td>
<td>05 min</td>
<td>Primary</td>
</tr>
<tr>
<td>DKEFS Sorting Test</td>
<td>executive function</td>
<td>10 min</td>
<td>Secondary</td>
</tr>
<tr>
<td>CVLT2 Learning Delayed Recall</td>
<td>auditory/verbal memory</td>
<td>10 min</td>
<td>Primary</td>
</tr>
<tr>
<td>BVMTR Learning Delayed Recall &amp; Copy</td>
<td>visual/spatial memory</td>
<td>05 min</td>
<td>Primary</td>
</tr>
<tr>
<td>COWAT</td>
<td>executive function</td>
<td>05 min</td>
<td>Primary</td>
</tr>
</tbody>
</table>

An important consideration for all of the proposed tests is normalization. These tests were created by various authors and have been modified by some in the development of optimal research batteries. Fortunately, while some of the test manuals are out of date [eg SDMT manual from 1982], all of the tests were employed in a recent regression-based normalization study,\textsuperscript{46} thus providing one source for norms that will aid interpretation, at least for English speaking subjects.

\textbf{Part 3: Confounds to interpretation of cognitive testing results and potential covariates in clinical trials.}

Sensory deficits are obvious confounds to valid interpretation of neuropsychological test results. Visual acuity is important for performance on the SDMT, BVMTR and other visually presented tests.\textsuperscript{47} In 2002, the consensus MACFIMS opinion paper\textsuperscript{48} recommended a near visual acuity screen, corrected at 20/70. Color vision impairments are also common in MS. The Stroop conflict tasks clearly require color vision, as does the DKEFS Sorting Test. Therefore, examiners must ensure that all colored stimuli can be readily distinguished by the patient.

Likewise motor tone, strength and coordination deficiencies are potential impediments to neuropsychological testing. The BVMTR requires the rendering of rather course or easily copied figures,
and the copy trial is used to adjust for tremulous drawings on the recall trials. The DKEFS Sorting Test requires that only a card be moved, and therefore only patients with severe UE motor dysfunction will be unable to complete the test. The Rao SDMT, PASAT, CVLT2 and COWAT rely to some extent on oral coordination, and dysarthria, which is common in MS, may hinder data interpretation.

In clinical studies, other tests that quantify some of these confounds can be employed and utilized as screens or covariates in data analysis. For UE function we have used the Holyan 9-Hole Peg Test (9HPT)\textsuperscript{49} which requires the participant to insert and then remove nine pegs from holes in a pegboard as quickly as possible. We record the average number of seconds required to complete the task twice with each hand, and for all four trials. For articulation of speech we have employed the Maximum Repetition Rate of Syllables and Multisyllabic Combinations Test \textsuperscript{50} in which the respondent must repeat phonemes (e.g., “ba–ta–ka”) as rapidly as possible for 6 seconds. This is essentially a quantification of dysarthria which has modest association with these cognitive tests.\textsuperscript{51}

MS-related cognitive decline can be associated with depression.\textsuperscript{52-55} The nature of this association is unclear: depression might cause cognitive dysfunction, or it might result from an underlying disease process that also causes cognitive impairment.\textsuperscript{56} We have used the Beck Depression Inventory Fast Screen for Medical Patients (BDIFS)\textsuperscript{57} to quickly assess depression in MS. The Center for Epidemiologic Studies Depression Scale (CESD)\textsuperscript{58-59} is another, more popular option. Both are simple paper and pencil surveys. BDIFS presents 7 items under which the most accurate self statement is endorsed. For the CESD respondents are asked to rate on a four point scale the frequency with which mood-related depressive symptoms have occurred over the past week (e.g., ‘I felt lonely’ and ‘I had crying spells’). We have experience with both tests which are valid in MS,\textsuperscript{60-62} but the CESD may be a better choice as it is used widely in Europe,\textsuperscript{63} has excellent construct validity,\textsuperscript{64} and is commonly used in other diseases.\textsuperscript{65}

\textbf{Part 4: EXPLORATORY Aspects of the MS CDE: Other considerations and research needs.}

(a) The supplemental and core tests recommended have good test-retest reliability, but more research is needed to validate parallel or alternate test forms. Alternate forms may mitigate, but are unlikely to eliminate practice effects. Research is also needed to determine optimal test-retest intervals for clinical trials.

(b) Recommendations for individual tests incorporated consideration of the psychometric properties of each measure. There are other interesting, potentially valuable ‘tasks’ or self-report measures of depression that should be considered for further research but that are either new or not yet standardized enough to be considered for the core battery in its present form. For example, in the area of cognitive processing speed, computer administered tests have great potential and it would be useful to collect psychometric data with extant tests including CogniStat, Headminder, ANAM and Neurotrax. The n-back procedure widely used in fMRI studies in MS is another measure in which further psychometric data should be collected. For depression, we are aware of the PROMIS initiative but to the best of our knowledge the new depression scale is not yet validated in prospective MS samples.

(c) Much more information is needed regarding the definition of clinically relevant impairment with respect to poor performance on psychometric tests. Similarly, further clarification is needed to determine the extent to which change over time on test performance is clinically relevant. One approach is to identify the relation between incremental changes in raw test scores of reliable, valid tests and clinically relevant measures of quality of life as has been done with the SDMT.\textsuperscript{13-15, 28} Conversely, one can standardize tests
with greater face validity that can be administered in a controlled setting. Examples include a driver simulation test\textsuperscript{66} or a test of activities of daily living\textsuperscript{(ADL)} such as a kitchen preparation or cooking task\textsuperscript{67}.

(d) The panel feels strongly that as clinical trials always assess change in response to an intervention, it is imperative to accumulate normative data in healthy samples at multiple time points.

(e) The recommended battery was developed for MS research with primary or secondary outcomes that provide an overall assessment of cognitive functioning. It is expected that targeted interventions of specific cognitive abilities might require other measures including more experimentally based tasks.

(f) Our committee recognizes that there are other domains that could be included under the umbrella of neuropsychology, broadly defined, not addressed herein. These other domains include fatigue, quality of life, employment, and affective disorders.
The investigation of cognitive functioning in children and adolescents with MS is limited to a small number of studies conducted over the past decade by only a few groups. The results to date suggest that as in adult-onset MS, children and adolescents with MS may demonstrate deficits in processing speed, attention/executive functions, episodic memory, and visual-spatial abilities, though other domains of function, including general intelligence and language can also be impaired. Some domains of function, such as attention, have not been well defined and will require further investigation in this population (e.g. sustained attention vs. divided attention).

One difficulty in interpreting the nature and severity of cognitive dysfunction in pediatric MS is that the data are too sparse and the measures too varied to permit adequate data pooling across studies. Neuropsychological testing in pediatric MS presents other unique challenges compared to adults given differences in developmental trajectories that typically correspond to different sets of tests and/or normative groups for different ages. At this point relative to adult neuropsychological studies in MS, the research and understanding of the cognitive complications of pediatric MS are still in their infancy and specific recommendations for a battery of CDEs would be premature. With respect to individual measures, this committee concludes that comparable to adult studies in MS, the Symbol Digit Modalities Test appears to hold particular promise in children and adolescents (age 8 and older) with MS, and initial studies suggest that the SDMT may also be sensitive to longitudinal change ¹. In addition to the SDMT, Table 1 shows that many measures applied to children and adolescents with MS have strong psychometric properties, show promise for the assessment of cognitive function in this population, and are currently used by national and international research groups.
<table>
<thead>
<tr>
<th>Cognitive Domain</th>
<th>Measure</th>
<th>Estim. time</th>
<th>Age span</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Measure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Processing speed</td>
<td>Rao SDMT (oral version)</td>
<td>05 min</td>
<td>8 years and older</td>
</tr>
<tr>
<td><strong>Other Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal learning and memory</td>
<td>-CVLT-C</td>
<td>20 min</td>
<td>5 – 16 (children)a</td>
</tr>
<tr>
<td></td>
<td>-Selective Reminding Test</td>
<td>15 min</td>
<td>5-15 (children) a</td>
</tr>
<tr>
<td></td>
<td>-TOMAL-2 subtests (Stories, Word List Learning)</td>
<td>12 min / test</td>
<td>5 - 59</td>
</tr>
<tr>
<td></td>
<td>-WRAML-2 subtests</td>
<td>15 min</td>
<td></td>
</tr>
<tr>
<td>Visual learning and memory</td>
<td>-BVMT-R</td>
<td>10 min</td>
<td>6 and older</td>
</tr>
<tr>
<td></td>
<td>-TOMAL-2 subtests (Abstract Visual Memory, Faces)</td>
<td>10 min</td>
<td>5 - 59</td>
</tr>
<tr>
<td></td>
<td>-WRAML-2 subtests</td>
<td>10 min / test</td>
<td>5 - 90</td>
</tr>
<tr>
<td>Processing speed / attention</td>
<td>-Trail Making Test – Sequencing (DKEFS)</td>
<td>05 min / test</td>
<td>8 – 89 (DKEFS norms)</td>
</tr>
<tr>
<td></td>
<td>-Rapid Picture Naming and Visual Matching (both from Woodcock-Johnson III Tests of Cognitive Abilities)</td>
<td>25 min</td>
<td>8 - 90</td>
</tr>
<tr>
<td>Language</td>
<td>-Vocabulary (from WASI)</td>
<td>10 min</td>
<td>6 – 89</td>
</tr>
<tr>
<td></td>
<td>-Expressive One Word Picture Vocabulary Test</td>
<td>10 min</td>
<td>2 and older</td>
</tr>
<tr>
<td></td>
<td>-Verbal Fluency (DKEFS)</td>
<td>10 min</td>
<td>8 – 89 (DKEFS norms)</td>
</tr>
<tr>
<td></td>
<td>-Picture Vocabulary (from Woodcock-Johnson III)</td>
<td>6 min</td>
<td>5 – 89 (DKEFS norms)</td>
</tr>
<tr>
<td></td>
<td>-CELF-III (Listening to Paragraphs)</td>
<td></td>
<td>8 – 89 (DKEFS norms)</td>
</tr>
<tr>
<td>Visuomotor/visual spatial function</td>
<td>-Beery VMI</td>
<td>5 min</td>
<td>2 – 100</td>
</tr>
<tr>
<td></td>
<td>-Block Design (from WASI)</td>
<td>15 min</td>
<td>6 – 89</td>
</tr>
<tr>
<td>Psychomotor function</td>
<td>-Grooved Pegboard</td>
<td>5 min</td>
<td>5 and older</td>
</tr>
<tr>
<td></td>
<td>-Grip Strength</td>
<td>2 min</td>
<td>5 and older</td>
</tr>
<tr>
<td></td>
<td>-9-Hole Peg Test</td>
<td>3 min</td>
<td></td>
</tr>
<tr>
<td>Executive Function</td>
<td>-Trail Making Test – Shifting (DKEFS)</td>
<td>8 min</td>
<td>8 – 89 (DKEFS norms)</td>
</tr>
<tr>
<td></td>
<td>-WCST</td>
<td>25 min</td>
<td>7 – 89</td>
</tr>
<tr>
<td></td>
<td>-CNT</td>
<td>20 min</td>
<td>5 – 14</td>
</tr>
<tr>
<td></td>
<td>-Color Word Interference (DKEFS)</td>
<td>15 min</td>
<td>8 – 89 (DKEFS norms)</td>
</tr>
<tr>
<td></td>
<td>-Verbal Fluency (Switching) (DKEFS)</td>
<td>05 min</td>
<td>8 – 89 (DKEFS norms)</td>
</tr>
<tr>
<td>General Intelligence</td>
<td>-WASI</td>
<td>30 min</td>
<td>6 – 89</td>
</tr>
<tr>
<td></td>
<td>-WISC-IV</td>
<td>60 min</td>
<td>6-16</td>
</tr>
</tbody>
</table>
NINDS Multiple Sclerosis Common Data Elements (CDE) Recommendations
Neuropsychology/Cognition Subgroup

Note: CVLT-C = California Verbal Learning Test- Childrens; CVLT-II = California Verbal Learning Tests – II; SDMT = Symbol Digit Modalities Test; TOMAL= Test of Memory and Learning; WRAML=Wide Range Assessment of Memory and Learning; WASI – Wechsler Abbreviated Scale of Intelligence; COWAT = Controlled Oral Word Association Test; DKEFS = Delis Kaplan Executive Function System; CELF = Clinical Evaluation of Language Fundamentals; Beery VMI= Beery Visuomotor Integration Test; WCST = Wisconsin Card Sorting Test; CNT= Contingency Naming Test; WISC-IV = Wechsler Intelligence Scale for Children – IV.

a Adult versions also available.

Pediatric MS References


Adult MS References


