

Neurocognitive outcomes in neurofibromatosis clinical trials

Recommendations for the domain of attention

Karin S. Walsh, PsyD
Jennifer Janusz, PsyD
Pamela L. Wolters, PhD
Staci Martin, PhD
Bonita P. Klein-Tasman, PhD
Mary Anne Toledo-Tamula, MA
Heather L. Thompson, PhD, CCC-SLP
Jonathan M. Payne, DPsych
Kristina K. Hardy
Peter de Blank, MD, MSCE
Claire Semerjian, MA
Laura Schaffner Gray, MPhil
Sondra E. Solomon, PhD†
Nicole Ullrich, MD
For the REiNS International Collaboration

Correspondence to
Dr. Walsh:
kwalsh@childrensnational.org

Supplemental data
at Neurology.org

ABSTRACT

Neurofibromatosis type 1 (NF1) is associated with neurocognitive deficits that can impact everyday functioning of children, adolescents, and adults with this disease. However, there is little agreement regarding measures to use as cognitive endpoints in clinical trials. This article describes the work of the Neurocognitive Committee of the Response Evaluation in Neurofibromatosis and Schwannomatosis (REiNS) International Collaboration. The goal of this committee is to identify standardized and specific cognitive assessment tools for use in NF clinical trials. The committee first identified cognitive domains relevant to NF1 and prioritized attention as the first domain of focus given prior and current trends in NF1 cognitive clinical trials. Performance measures and behavioral rating questionnaires of attention were reviewed by the group using established criteria to assess patient characteristics, psychometric properties, and feasibility. The highest rated tests underwent side-by-side comparison. The Digit Span subtest from the Wechsler scales was given the highest ratings of the performance measures due to its good psychometrics, feasibility, utility across a wide age range, and extensive use in previous research. The Conners scales achieved the highest ratings of the behavioral questionnaires for similar reasons. Future articles will focus on other cognitive domains, with the ultimate goal of achieving agreement for cognitive endpoints that can be used across NF clinical trials. *Neurology*® 2016;87 (Suppl 1):S21-S30

GLOSSARY

ADHD-RS = ADHD Rating Scale; **CAARS** = Conners Adult ADHD Rating Scale; **COA** = clinical outcome assessment; **DS** = Digit Span; **FDA** = Food and Drug Administration; **FSIQ** = full-scale IQ; **MATRICES** = Measurement and Treatment Research to Improve Cognition in Schizophrenia; **NF1** = neurofibromatosis type 1; **ObsRO** = observer-reported outcome; **Perfo** = performance outcome; **REiNS** = Response Evaluation in Neurofibromatosis and Schwannomatosis; **TEA** = Test of Everyday Attention; **TEA-Ch** = Test of Everyday Attention for Children.

Neurocognitive sequelae in neurofibromatosis type 1 (NF1) have been well-documented and result in significant morbidity and dysfunction in children, adolescents, and adults with the disease.¹⁻⁷ Research has documented phenotypic patterns of cognitive dysfunction in NF1, as well as the functional impact these deficits have on individuals in naturalistic settings such as school and work.⁸⁻¹¹

Despite the high prevalence and significant morbidity of cognitive impairments, few intervention trials have targeted cognitive dysfunction as a primary endpoint. It is vital to establish standards to evaluate therapies for the treatment of cognitive deficits, and to work towards acceptance of cognition as a therapeutic target. Behavioral interventions to prevent or remediate cognitive deficits would benefit from consensus on cognitive endpoints as well. With these goals in mind, the Response Evaluation in Neurofibromatosis and Schwannomatosis (REiNS) Neurocognitive Subcommittee was formed to identify standardized, specific cognitive endpoints for clinical trials.

While human clinical trials targeting cognition in NF1 have been relatively slow to emerge, murine models of cognitive dysfunction have been ongoing since the original work of Silva

†Deceased.

From the Children's National Health System (K.S.W., K.K.H.), The George Washington School of Medicine, Washington, DC; Children's Hospital Colorado (J.J.), University of Colorado School of Medicine, Aurora; Pediatric Oncology Branch (P.L.W., S.M.), National Cancer Institute, Bethesda, MD; University of Wisconsin-Milwaukee (B.P.K.-T.); Clinical Research Directorate/Clinical Monitoring Research Program (M.A.T.-T.), Leidos Biomedical Research, Inc., NCI at Frederick, MD; California State University (H.L.T.), Sacramento; Murdoch Children's Research Institute (J.M.P.), Victoria, Australia; Rainbow Babies & Children's Hospital (P.d.B.), Cleveland, OH; Roosevelt University (C.S.), Chicago, IL; Boston Children's Hospital (L.S.G., N.U.), MA; and University of Vermont (S.E.S.), Burlington.

REiNS International Collaboration members are listed on the *Neurology*® Web site at Neurology.org.

Go to Neurology.org for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

et al.¹² Animal models have provided guidance in the development of clinical trials, providing pathway level hypotheses for translation to human trials.¹³ However, the translation from mouse behavior to human cognitive models has been challenging due to the range and complexity of human cognition and the lack of consensus regarding how best to measure cognition for clinical trials endpoints.

The diversity of approaches to measurement of cognitive outcomes is illustrated when reviewing the cognitive clinical trials that have been completed or are ongoing. As of 2016, there have been 9 clinical trials targeting cognitive outcomes in NF1, each employing a different battery of cognitive tests despite targeting the same primary constructs of attention, working memory, visual memory, intelligence, and emotional/behavioral functioning (table 1). The lack of evaluative and analytic consistency and standards significantly hinders our ability to adequately assess therapeutic efficacy, to combine smaller samples to yield higher power to detect effects, and limits the generalizability of findings.

Other disease groups have developed consensus and standards for evaluating cognitive outcomes to support the development of new therapies.^{14,15} Perhaps the most successful work in this area has been the work of the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) collaboration.^{15–17} The MATRICS group posited that the lack of approved treatments for cognitive dysfunction in schizophrenia was directly related to the use of diverse assessment approaches across clinical trials. This problem significantly limited the ability to adequately evaluate a drug's effectiveness in treating cognitive deficits, especially for governmental organizations tasked with oversight.¹⁵ The MATRICS group identified specific criteria critical for determining appropriate assessment tools for endpoints in clinical trials of schizophrenia, including (1) good test-retest reliability, (2) high utility for repeated assessment, (3) a relationship with functional outcomes, (4) sensitivity to a pharmacologic agent, and (5) practicality.¹⁵ Co-primary endpoints that assess the impact of cognitive functioning on everyday living skills were also emphasized. These measures

complement the primary outcome of cognitive performance and provide essential evidence of therapeutic utility.

Following the example of other disease-specific groups, this article is the first of its kind to provide recommendations for cognitive outcomes for clinical trials in NF1. This will support the promotion of cognition as a target for interventions in NF1 and encourage the development of novel compounds and behavioral therapies targeting these impairments.

CONSIDERATIONS AND CHALLENGES There are challenges associated with efficiently and effectively developing a standardized cognitive test battery for use in clinical trials. First, there are many cognitive domains that could conceivably be appropriate for clinical trials endpoints and within many of these domains are subdomains of skills, each of which has the potential to be a specific therapeutic target (table 2). For example, full-scale IQ (FSIQ) is composed of multiple distinct functional areas including verbal and nonverbal reasoning, working memory/attention, and processing speed. Given the multifactorial nature of intellect, global measures such as FSIQ will likely lack the sensitivity to be ideal primary endpoints, particularly when there are discrepancies between the various functions assessed, as is common in NF1.

Second, many tests are available to assess each cognitive domain. Each test varies in its reliability and validity for specific populations, as well as its feasibility for use in clinical trials, including use for shorter test-retest periods. Developmental factors are another significant challenge for cognitive researchers. Few cognitive tests or symptom checklists are intended to span broad age ranges, with most meant for specific use with preschoolers, school-age children/adolescents, or adults. It will be important to include young children in clinical trials; however, preschool measures are sparse compared to those available for other age ranges, and will require careful consideration. Finally, there is virtually no literature about cognitive effects of aging in NF1, such that it is unclear which measures are most effective at characterizing cognitive functioning in older adulthood. Based on these considerations, we prioritized the domain of attention, a highly prevalent problem in NF1, which has been evaluated in recent or ongoing clinical and preclinical trials. Further, the current work will focus on school-aged children (≥ 6 years) through adults. We will review and make recommendations for the preschool age group, as well as offer methodologic and statistical approaches to longitudinal designs across age groups (e.g., preschool into school age, and school age into adulthood), in future publications.

Table 1 Clinical trials targeting cognitive outcomes in neurofibromatosis type 1

Study reference ^a	Intervention/design	No.	Primary construct of interest and outcome measures	Other outcome measures	Results
34	Methylphenidate (single-arm)	93	Attention: TOVA; CBCL	Not applicable	Significant improvement on TOVA and improvement after 1 year on CBCL
35	Simvastatin (randomized placebo controlled)	62	Long-term visual memory: RCFT (Delayed Recall); attention: Cancellation Test (speed)	RCFT (Copy); Cancellation Test (SD); Stroop; Block Design; Object Assembly; VMI; JLO	No significant improvement following treatment with the exception of object assembly (specifically in the group with lower baseline functioning on the task)
36	Single-arm lovastatin (phase I/II; on treatment 3 months)	24	Short- and long-term visual and verbal memory: WRAML-2 (Story, Design, and Picture Memory), CVMT, CVLT-C; attention: TEA-Ch (Score), Sky Search, Sky Search Dual Task, Creature Counting)	JLO; Grooved Pegboard	Some improvements in memory with 13%–39% exceeding base rate change (80% CI) from pretreatment to posttreatment evaluation, indicative of clinically meaningful change following treatment for at least this subgroup of participants
37	Methylphenidate (placebo-controlled)	80	Psychological: STAI-C; CDI		No results published or presented
38	Randomized placebo-controlled simvastatin (on treatment 12 months)	84	Global functioning/intelligence: WISC-III-NL (full-scale intelligence); attention and emotional: CBCL (Attention Problems and Internalizing Problems scales)	RCFT (Delayed Recall); Stroop; Grooved Pegboard; TRF; YSR; CHQ (Parent)	No significant improvement following treatment
39	Methylphenidate (placebo-controlled, crossover)	39	Attention: CPRS-3	CTRS-3	Significant improvement in symptoms on CPRS
40	Lovastatin (placebo-controlled)	44	Nonverbal declarative memory: BVMT; working memory: Letter-Number Sequencing	Digit Cancellation; HVL; WISC-III Object Assembly; CBCL/YASR	Significant improvement on Letter-Number Sequencing, HVL, and categorical verbal generation; significant improvement in adult patients on internalizing symptoms; all other primary and secondary outcomes not significant
27	Randomized placebo-controlled lovastatin (STARS)	146	Visual working memory: CANTAB (Paired Associates Learning); attention: TEA-Ch (Score)	CANTAB (Spatial Working Memory, Stockings of Cambridge, Stop Signal Task, Motor); TEA-Ch (Sky Search, Sky Search Dual Task, Creature Counting); JLO; WISC-III (Object Assembly); CPT-II; COWAT	Results pending
41	Lamotrigine (randomized placebo-controlled)	Active	Intellectual: WISC-III (performance IQ)	WISC-III (full IQ); SA-Dots; BRIEF; ADHD Questionnaire; PAL (CANTAB); VMI-6; MVPT-3; Grooved Pegboard	Active

Abbreviations: ADHD = attention-deficit/hyperactivity disorder; BRIEF = Behavior Rating Inventory for Executive Function; BVMT = Brief Visuospatial Memory Test-Revised; CANTAB = Cambridge Neuropsychological Test Automated Battery; CBCL = Child Behavior Checklist; CDI = Children's Depression Inventory; CHQ = Child Health Questionnaire; CI = confidence interval; COWAT = Controlled Oral Word Association Test; CPRS/CTRS = Conners' Parent Rating Scale/Conners' Teacher Rating Scale; CPT-II = Conners Continuous Performance Test, Second Edition; CVLT-C = California Verbal Learning Test for Children; CVMT = Continuous Visual Memory Test; HVL = Hopkins Verbal Learning Test; JLO = Judgment of Line Orientation; MVPT-3 = Motor-Free Visual Perception Test, Third Edition; PAL = Paired Associative Learning; RCFT = Rey Complex Figure Test; SA-Dots = Sustained Attention Dots of the Amsterdam Neuropsychological Tasks; STAI-C = State Trait Anxiety Inventory for Children; STARS = A Randomized Placebo-Controlled Study of Lovastatin in Children With Neurofibromatosis Type 1; TEA-Ch = Test of Everyday Attention for Children; TOVA = Test of Variables of Attention; TRF = Teacher Report Form; VMI = Beery Test of Visual-Motor Precision; WISC = Wechsler Intelligence Scale for Children; WRAML = Wide Range Assessment of Memory and Learning, Second Edition; YSR/YASR = Youth Self-Report/Young Adult Self Report.

^aLiterature search included PubMed and Clinicaltrials.gov.

METHODS Goals of the REiNS Neurocognitive Committee. The goals of this working group are to (1) critically and scientifically review standardized cognitive assessment tools for use in NF1 clinical trials, (2) provide recommendations for a core test bank of neurocognitive measures across cognitive domains to be utilized in NF1 clinical trials to enhance consistency across trials and sites, and (3) function as a resource for researchers in the development of cognitive protocols. By providing this guidance, we hope to support the development and evaluation of novel therapeutics for the treatment of cognitive deficits in children and adults with NF1.

Food and Drug Administration (FDA) clinical outcome assessment (COA). To select clinically useful cognitive endpoints,

we have been guided by recommendations regarding COAs.¹⁸ These COAs must be appropriately defined and reliable for the evaluation of a specific condition of interest in a specific context of use, based on available evidence. Qualification as a COA indicates that a tool can be relied upon to measure a distinct concept and have a specific interpretation and application in drug development. The FDA has deemed 2 areas of specific relevance: performance outcome (PerFO) and observer-reported outcome (ObsRO) assessments. PerFO fall into 2 categories: (1) paper-pencil tests and (2) computerized tasks/batteries. ObsRO include behavior rating questionnaires completed by caretakers (parents, guardians) and teachers, which assess a child's behavioral/emotional functioning or cognitive functions including attention, impulsivity, executive function, and adaptive function, in their everyday environment. Within each of the

Table 2 Cognitive domains relevant to neurofibromatosis clinical trials

Cognitive domain	Description	Subdomains
Attention	The ability to select and focus on relevant information while filtering out irrelevant information	Vigilance (engagement, focus); sustained attention; attentional shift (re-engagement)
Executive function	An umbrella term for higher-order cognitive processes involved in managing cognitive and behavioral resources to achieve a goal	Initiation; inhibition; monitoring (cognitive and behavioral); working memory; strategic planning; organization; flexibility; emotional regulation
Memory	The ability to retain and recall information	Encoding; consolidation; retrieval; (verbal and visual; short- and long-term)
Processing speed	The ability to complete tasks quickly, automatically, and fluently	
Visual-spatial/visual-motor	Visual-spatial skills include the ability to mentally manipulate 2D and 3D objects, as well as determine spatial relationships between objects; visual-motor skills are the ability to coordinate motor movements with visual input	Mental rotation; visual closure; spatial orientation; object localization; form/pattern discrimination; drawing/copying
Language	Receptive language skills are the ability to understand language heard or read; expressive language skills are the ability to put thoughts into words and sentences in a way that makes sense and is grammatically correct	Phonology—phonemic awareness, decoding, encoding (spelling and reading); semantics—receptive (aural and written—to include reading comprehension); expressive (oral and written—to include spelling and composition); morphology—understanding and producing grammatical aspects of language; pragmatics—understanding social aspects of language; using social language appropriately; syntax—understanding and using appropriate sentence structure
Adaptive	The ability to handle common life demands	Community living; health and safety; self-care; leisure skills; coping skills; social skills; functional communication; functional academics; work skills; self-direction
Behavioral/emotional	An individual's mood and behavioral presentation	Depression; anxiety; conduct problems; aggression; withdrawal; atypical behaviors
Academics	School-related skills	Reading; reading comprehension; spelling; written expression; math calculation; math problem-solving

cognitive domains being evaluated, we provide recommendations for both performance and observational measures, which may serve as co-primary outcome tools in NF1 clinical trials to evaluate their impact on everyday function, as has been recommended by the FDA.¹⁹

Current clinical trials. We completed a literature search for past and current pharmacologic cognitive clinical trials in NF1 in order to summarize the literature and document the diversity of assessment instruments utilized for the same or very similar primary outcome constructs. We performed several searches to include PubMed.gov and ClinicalTrials.gov. Using the search terms “neurofibromatosis type 1 cognitive clinical trial” and “neurofibromatosis type 1 methylphenidate” identified 8 current or complete/published trials. We were made aware of a ninth pharmacologic trial that was accepted for publication but ahead of print, which was included.

Identification of cognitive domains and tools for review.

Because of the extensive number of cognitive domains and associated assessment tools, it would be prohibitive to complete an exhaustive review of all published tests. In order to increase the feasibility of achieving the goals of the working group, a 2-step process was implemented. The first phase involved identifying all cognitive domains relevant to NF1 and prioritizing those domains, with the highest importance associated with the known or likely targets for cognitive clinical trials now and in the near future. The second phase included surveying experienced NF1 cognitive researchers and clinicians, who were asked to list the assessment instruments within each cognitive domain that they use most consistently and consider to be reliable and sensitive tools with the NF1 population. The list of potential assessment tools was summarized and discussed to establish a final pool of instruments to be reviewed in order to establish a core set of test instruments. This manuscript provides recommendations for

paper-pencil tests (PerfO) and behavioral questionnaires (ObsRO) of attention in NF1. Because computerized test batteries evaluate a range of cognitive domains, recommendations for computerized test batteries that include tests of attention will be included in a future article.

Establishing evaluation criteria. Based on a systematic review process established by the REiNS Patient-Reported Outcomes Committee, we developed the Cog-RATE form as a means of reviewing cognitive tests.²⁰ The form enabled the committee to rate each measure on 6 key criteria: patient characteristics (age range for measure; use with specific populations), use in published studies, domains assessed, availability of standard scores, psychometric characteristics, and feasibility for clinical trials. The Neurocognitive Committee decided to give greater consideration to each test's psychometric properties, patient characteristics, and feasibility of use in clinical trials, which were considered the most relevant for clinical trials. The process utilized by our committee in reviewing each tool is provided in table 3.

RESULTS We reviewed 5 performance-based measures and 5 observation measures of attention available for school-aged children through adults (age 6 and up; table 4). Of the 5 performance measures considered, 2 tasks, Spatial Span from the Wechsler scales and Spatial Span from the Stanford-Binet, did not remain under consideration given their limited use in previous NF1 clinical studies and limitations in specificity.^{21,22}

Three measures were sufficiently rated to remain under consideration: Digit Span (DS) from the Wechsler scales (Wechsler Intelligence Scale for

Table 3 Review and rating process of cognitive tests and rating scales

Step 1: Identify cognitive measures to review per domain
Survey neurofibromatosis researchers and clinicians about cognitive measures they use
Identify other possible measures from personal knowledge or literature review
Review of a measure by one member helps decide if a full group review is needed
Step 2: Prepare for the review of a measure
Members use the Cog-RATE form to systematically review cognitive measures
One member is identified to serve as a lead reviewer for a measure
The lead reviewer and group chair review the literature about the measure
Relevant articles describing the measure and its use in validation/descriptive studies and clinical trials are posted on the group's document-sharing site 1-2 weeks prior to the call
Members review the scale and articles and rate each Cog-RATE criteria
Step 3: Group review and rating of a measure by conference call
Cog-RATE forms may be e-mailed to the chair if members cannot attend a call
Members provide and discuss their rating from 0 (lowest rating/unacceptable) to 3 (highest rating) for each of the 6 Cog-RATE criteria; ratings are recorded
Discrepancies are discussed to make sure they are not based on lack of familiarity with the measure or not understanding all the information
The mean of the ratings for each of the 6 criteria is computed; the overall score is the total mean of the 6 criteria
Strengths and weaknesses are summarized; notes are taken regarding any outstanding issues
Members discuss whether the measure could be recommended as a primary or secondary outcome measure
The group determines if any additional information is needed to reach a final decision and whether a measure needs to be brought up for a second review
For newly developed measures, a future review is planned to consider new information
The top closely rated measures may be compared side-by-side in another meeting
Step 4: Post-call activities
The group chair provides notes to summarize the conference call
Final notes and the Cog-RATE form ratings are uploaded to the document-sharing site to document the group's decision
Final recommendations are reviewed with the entire REiNS collaborative group at yearly meetings

Abbreviations: REiNS = Response Evaluation in Neurofibromatosis and Schwannomatosis.

Children and Wechsler Adult Intelligence Scale), the Test of Everyday Attention (TEA), and the TEA for Children (TEA-Ch).²³⁻²⁶ The TEA-Ch and DS have both been used in previous published clinical trials with NF1 and demonstrated adequate psychometric properties. As the TEA-Ch includes 9 different subtests, we evaluated this measure in several ways: as a complete battery including all 9 subtests, as a screening instrument of 4 subtests (Sky Search, Score!, Creature Counting, and Sky Search Dual Task), and as individual subtests. We reviewed the TEA as a complete battery as well as by individual subtests, particularly as they relate to continuity with the TEA-Ch. These measures were then compared across the primary evaluative factors.

Although the TEA-Ch (and the TEA, the adult version of the test) is a feasible and sensitive assessment of attention, it lacks normative data in individuals 16 and 17 years of age. In prior studies, this gap has resulted in arbitrary truncation of inclusion criteria related to age.²⁷ In addition, there is a lack of

parallel subtests between the TEA-Ch and TEA, significantly limiting the use of these tools across age ranges in clinical trials. For these reasons, the committee could not recommend the tools as primary outcome measures in NF1 clinical trials.

DS can be used as a combined performance score, or separated into Forward, Backward, and Sequencing components. The committee ultimately chose DS as the recommended PerfO measure of attention given its utility in a wide range of ages (6 years through adult), good feasibility (easily administered in sites that have a Wechsler instrument), as well as good psychometric properties, particularly test-retest reliability. Because DS Backward and Sequencing tap into aspects of executive function including working memory, the committee recommends the specific use of DS Forward for the measurement of attention. When using any performance-based test in preintervention and postintervention designs, the risk of practice effects must always be considered. Previous intervention trials using DS have found limited

Table 4 Performance-based measures reviewed

Measures	Normative data, n	Age range, year:month	Reliability (test-retest)	Validity studies	Pros	Cons	Cog-RATE overall group rating
Digit Span WISC-IV ²³	2,200	6:0-16:11	No. = 243; DS: $r = 0.81$; DSF: $r = 0.79$; DSB: $r = 0.76$; no. = 43 ⁴² ; DS: $r = 0.74$	Concurrent: WISC-III: $r = 0.77$; WAIS-III: $r = 0.73$; construct: CMS attention/concentration: $r = 0.72$	Widely used in descriptive and clinical studies, including NF1 cohorts	Total score is a composite of DSF and DSB, which measure different constructs ^{43,44}	2.62
WAIS-IV ²⁴	2,200	16:0-90:11	No. = 293; DS: $r = 0.83$; DSF: $r = 0.77$; DSB: $r = 0.71$	Construct: CMS attention/concentration; $r = 0.76$; RBANS attention; $r = 0.65$ (WAIS-IV technical manual)			
TEA-Ch ²⁶	293	6:0-15:11	No. = 55 ⁴⁵ ; $r = 0.65-0.87$	Structural equation model reveals 3 factors: selective, switching, and sustained attention (regression coefficients: 0.44-0.79); TEA-Ch subtest correlations with IQ are generally weak (all $r < 0.31$)	Used in clinical studies, including NF1; numerous subtests that allow assessment across different attentional capacities	Normative data limited to 6-15 years, 11 months	2.44
TEA ²⁵	154	18:0-80:0	No. = 118 ²⁵ ; $r = 0.59-0.86$	Construct: $r = 0.42-0.63$ with other established measures of attention	Numerous subtests that allow assessment across different attentional capacities	Has not been used in NF1 clinical studies	2.18
Spatial Span (WISC-IV Integrated) ²¹	2,200	6:1-16:11	No. = 174 ²¹ ; SPF: $r = 0.79$; SPB: $r = 0.81$	Discriminative: SSF and SSB scores for ADHD and ASD standardization samples are significantly lower than matched control group		Total score is a composite of SPF and SPB, and they appear to measure different constructs ⁴⁶ ; floor effects in younger children; not used in NF1 clinical studies	2.26
Block Span (Stanford Binet-V) ²²	4,800	2:0-85:0	No. = 356 ²² ; $r = 0.66-0.93$	Construct: $r = 0.66-0.85$ with the Wechsler Scales (WPPSI-R, WISC-III, WAIS-III)		Test specificity in question since other factors can impact performance (e.g., visual scanning and problems with motor control); not used in NF1 clinical studies	2.42

Abbreviations: ADHD = attention-deficit/hyperactivity disorder; ASD = autism spectrum disorder; CMS = Children's Memory Scale; DS = Digit Span; DSB = Digit Span Backward; DSF = Digit Span Forward; NF1 = neurofibromatosis type 1; RBANS = Repeatable Battery for the Assessment of Neuropsychological Status; SPB = Spatial Span Backward; SPF = Spatial Span Forward; SSB = spatial span backward; SSF = spatial span forward; TEA-Ch = Test of Everyday Attention for Children; WAIS = Wechsler Adult Intelligence Scale; WISC = Wechsler Intelligence Scale for Children; WPPSI = Wechsler Preschool and Primary Scale of Intelligence.

Table 5 Observer-rated measures reviewed

Measures	Normative data	Age range, year: month	Reliability	Validity studies	Pros	Cons	Cog-RATE overall group rating
Conners 3 ³²	3,400	6:0-16:0 (parent/teacher); 8:0-18:0 (self-report)	Test-retest $r = 0.85$ (parent); $r = 0.85$ (teacher); $r = 0.79$ (self-report); interrater $r = 0.81$ (parent); $r = 0.73$ (teacher)	Confirmatory factor analyses revealed parent and self-reports have adequate fit; teacher version was slightly lower; acceptable between-informant correlations: parent to teacher $r = 0.60$; parent to self-report $r = 0.56$; teacher to self-report $r = 0.48$; convergent and divergent validity established with other measures of child and adolescent psychopathology; moderate to high discriminant validity reported in the manual for the different scoring version of the Conners 3; probability of accurately predicting group membership (e.g., ADHD vs non-ADHD, impulsive vs hyperactive vs combined) ranged from 0.66 to 0.95	Previous version (CRS-R) has been used in 16 clinical trials (PubMed search term: Conners' Rating Scales-Revised); CRS-R has been used in many cross-sectional studies, including NF1 ⁴⁷ ; Scoring options for DSM-5 and DSM-IV-TR	Current version has limited use in published clinical trials to date	2.65
CAARS ³³	2,000	18+ (observer and self-report)	Test-retest $r = 0.80-0.91$ (self-report); $r = 0.85-0.95$ (observer) ³³	Confirmatory factor analyses for both self-report and observer forms met criteria for good fit; was able to discriminate between ADHD and control group with an overall correct classification rate of 85%; construct validity examined using the WURS, $r = 0.37-0.67$	Used in many cross-sectional and large clinical trials	Not used in NF1 cohorts to date	2.30
ADHD Rating Scale-IV ³¹	2,000	6:0-18:0 (parent/teacher)	Not available	Consistent with the 2-factor model of ADHD (inattention and hyperactive/impulsive); differentiated children with impulsive and combined subtypes from unaffected controls	Used in many cross-sectional and clinical studies, including NF1 cohorts	Weak evidence of convergent and discriminant validity reported using both teacher and parent versions ⁴⁸	2.65
Adult ADHD Symptom Scale, Self-Report ⁴⁹	720	17:0-50+	Not available	Not available	Have been used in multiple descriptive studies	Not used in NF1 cohorts to date	2.41
Vanderbilt ADHD Diagnostic Parent Rating Scale ⁵⁰	243	5:0-9:11	Not available	Consistent with the 2-factor model of ADHD (inattention and hyperactive/impulsive); moderate concurrent validity with DISC-IV $r = 0.79$		Not widely used in studies; no studies identified within NF1 samples	1.72

Abbreviations: ADHD = attention-deficit/hyperactivity disorder; CAARS = Conners Adult ADHD Rating Scale; CSR-R = Conners Rating Scale-Revised; DISC-IV = Diagnostic Interview Schedule for Children, 4th edition; DSM = *Diagnostic and Statistical Manual of Mental Disorders*; NF1 = neurofibromatosis type 1; WURS = Wender Utah Rating Scale.

practice effects in as short as a 6 week test-retest time span, suggesting that DS can be used even over short intervals.^{28,29} However, practice effects should always be managed through study methodology (e.g., inclusion of a control group) or statistical approaches, such as the use of reliable change indices, which take into account psychometric issues, practice effects, and other sources of variance to ensure that change demonstrated on tests is meaningful.³⁰ In addition, the DS, like many tests, has a truncated range of scores that may limit the ability to detect change for individuals who are not impaired at baseline. Researchers will need to take these considerations into account when using this measure as a clinical trial endpoint.

Five observation attention scales were reviewed (table 5). The 2 measures with the highest ratings were the Conners scales and the ADHD Rating Scale (ADHD-RS).³¹ The Conners scales include 2 versions: the Conners Parent Rating Scale-3 (ages 6 through 18) and the Conners Adult ADHD Rating Scale (CAARS; age 18 and older).^{32,33} Similarly, the ADHD-RS has child and adult rating forms. The Vanderbilt ADHD Rating Scale was also considered, but limitations related to a severely truncated age range, limited use in studies generally, regionally limited normative data, and lack of age-based standard scores removed the instrument from further consideration.

In comparing the Conners scales and the ADHD-RS, both received similar ratings. However, the committee believed that the superior psychometric information available and the extensive use of the prior version of the Conners scales in clinical trials were significant advantages and the Conners scales were therefore chosen for recommendation. In addition, the Conners scales have options for multiple informants for both the child and adult scales, adding to their reliability in multiple settings. For school-aged children, options include a parent, teacher, and self-report questionnaire. For adults, options include a self-report and an observer report. Additionally, the Conners scales have extensive psychometric data available and provide subscales beyond attention-deficit symptomatology that are of interest and importance in NF, including learning problems, executive function, and peer relations (table 4). However, caution is warranted in using the executive function subscale as a primary outcome variable as it only correlated significantly with the Plan/Organize scale of the Behavior Rating Inventory of Executive Function.³³

DISCUSSION AND FUTURE DIRECTIONS The establishment of recommendations for a rigorously reviewed core test bank of cognitive tests for use in NF clinical trials is of vital importance given the

morbidity these deficits cause in individuals with NF1 and the increased focus from the scientific community on the development of targeted therapies to ameliorate these impairments. To date, there has been no systematic review of the psychometric properties of available cognitive assessment tools specifically for use in clinical trials in NF1. As we work to promote the development of novel therapies targeting cognition in this population, it is imperative that we prioritize this step. To that end, the REiNS Neurocognitive Committee has taken a systematic, detailed approach to reviewing available tests of cognition in order to be able to make recommendations for performance- and observer-based instruments in clinical trials targeting cognitive deficits. We have taken guidance from the FDA in our approach, with the identification of and regulatory approval of cognitive enhancing agents for NF1 in mind. In this article, we focused on paper-pencil tools assessing attention, and our extensive review led us to recommend DS from the Wechsler scales as the performance-based outcome measure, and the Conners scales (Conners 3 for children; CAARS for adults) as the observer-rated outcome measure. The recommendations being made are to establish a core set of tests for use in clinical trials, but are not meant to preclude individual researchers' ability to include other tests relevant to the specific research hypotheses in the battery of outcome measures for any given trial. Future articles will provide recommendations for computerized batteries (including tests of attention) and tests in other key domains, including executive function, memory, academic skills, and others.

No single instrument will capture all aspects of a complex domain or function ideally for all trial designs. However, the DS and Conners scales represent the ObsRO and PerFO tools that have been reviewed so far with the best combination of psychometric properties, patient characteristics, and feasibility for NF1 clinical trials currently available. Adoption of these measures in clinical trials of cognitive outcomes in NF1 will support the development of novel therapeutics for cognitive dysfunction in NF1 and lead to improved impact through the use of consistent, reliable, and sensitive cognitive outcome measures.

AUTHOR CONTRIBUTIONS

Karin Walsh: design and conceptualization of the study, collection and interpretation of the data, drafting and revising the manuscript. Jennifer Janusz: design and conceptualization of the study, collection and interpretation of the data, drafting and revising the manuscript. Pamela Wolters: study concept, collection and interpretation of the data, drafting and revising the manuscript. Staci Martin: study concept, collection and interpretation of the data, drafting and revising the manuscript. Bonita Klein-Tasman: design and conceptualization of the study, collection and interpretation of the data, revising the manuscript. Mary Anne Toledo-Tamula: design and conceptualization of the study, collection and interpretation of the data, drafting and revising the manuscript.

Heather Thompson: design and conceptualization of the study, collection and interpretation of the data, revising the manuscript. Jonathan Payne: design and conceptualization of the study, collection and interpretation of the data, revising the manuscript. Kristina Hardy: design and conceptualization of the study, collection and interpretation of the data, revising the manuscript. Peter de Blank: design and conceptualization of the study, collection and interpretation of the data, revising the manuscript. Claire Semerjian: design and conceptualization of the study, collection and interpretation of the data. Laura Schaffer Gray: design and conceptualization of the study, collection and interpretation of the data. Sondra Solomon: design and conceptualization of the study, collection and interpretation of the data, revising the manuscript. Nicole Ullrich: design and conceptualization of the study, collection and interpretation of the data, revising the manuscript.

STUDY FUNDING

No targeted funding reported.

DISCLOSURE

K. Walsh received research support from the Children's Tumor Foundation and the Gilbert Neurofibromatosis Institute. J. Janusz reports no disclosures relevant to the manuscript. P. Wolters received funding from the Childhood Brain Tumor Foundation and Neurofibromatosis Therapeutics Acceleration Program and holds stock in Bristol-Meyers-Squibb. S. Martin received funding for a trip from the Children's Tumor Foundation. B. Klein-Tasman received funding from the NF Midwest and NF Northeast. M. Toledo-Tamula received funding from the federal funds from the National Cancer Institute, National Institute of Health, under contract no. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade name, commercial products, or organization imply endorsement by the US government. H. Thompson received funding through the Canadian Institutes of Health Research Doctoral Research Awards Program and the Ontario Barbershoppers' Harmonize for Speech Fund. She also received funding from the British Columbia Neurofibromatosis Foundation (BCNF) for travel to the BCNF Annual Conference. J. Payne received research support from the Children's Tumor Foundation, the Department of Defense, and the Murdoch Children's Research Institute. K. Hardy received research support from the Children's Tumor Foundation and the Gilbert Neurofibromatosis Institute. P. de Blank received support from the Neurofibromatosis Therapeutic Acceleration Program. C. Semerjian and L. Schaffer Gray report no disclosures relevant to the manuscript. S. Solomon is deceased; she reported no disclosures prior to her death. N. Ullrich received research support from the Department of Defense, NIH/National Cancer Institute, Neurofibromatosis Inc., and the Children's Tumor Foundation. Go to Neurology.org for full disclosures.

Received November 18, 2015. Accepted in final form April 6, 2016.

REFERENCES

1. Lehtonen A, Howie E, Trump D, Huson SM. Behaviour in children with neurofibromatosis type 1: cognition, executive function, attention, emotion, and social competence. *Dev Med Child Neurol* 2013;55:111–125.
2. Levine TM, Materek A, Abel J, O'Donnell M, Cutting LE. Cognitive profile of neurofibromatosis type 1. *Semin Pediatr Neurol* 2006;13:8–20.
3. Rosser TL, Packer RJ. Neurocognitive dysfunction in children with neurofibromatosis type 1. *Curr Neurol Neurosci Rep* 2003;3:129–136.
4. North K, Hyman S, Barton B. Cognitive deficits in neurofibromatosis 1. *J Child Neurol* 2002;17:605–612; discussion 627–629, 646–651.
5. Kayl AE, Moore BD. Behavioral phenotype of neurofibromatosis, type 1. *Ment Retard Dev Disabil Res Rev* 2000;6:117–124.

6. Ozonoff S. Cognitive impairment in neurofibromatosis type 1. *Am J Med Genet* 1999;89:45–52.
7. Acosta MT, Gioia GA, Silva AJ. Neurofibromatosis type 1: new insights into neurocognitive issues. *Curr Neurol Neurosci Rep* 2006;6:136–143.
8. Pride NA, Crawford H, Payne JM, North KN. Social functioning in adults with neurofibromatosis type 1. *Res Dev Disabil* 2013;34:3393–3399.
9. Oates EC, Payne JM, Foster SL, Clarke NF, North KN. Young Australian adults with NF1 have poor access to health care, high complication rates, and limited disease knowledge. *Am J Med Genet A* 2013;161A:659–666.
10. Pride NA, Payne JM, North KN. The impact of ADHD on the cognitive and academic functioning of children with NF1. *Dev Neuropsychol* 2012;37:590–600.
11. Coudé FX, Mignot C, Lyonnet S, Munnich A. Academic impairment is the most frequent complication of neurofibromatosis type-1 (NF1) in children. *Behav Genet* 2006;36:660–664.
12. Silva AJ, Frankland PW, Marowitz Z, et al. A mouse model for the learning and memory deficits associated with neurofibromatosis type I. *Nat Genet* 1997;15:281–284.
13. Shilyansky C, Lee YS, Silva AJ. Molecular and cellular mechanisms of learning disabilities: a focus on NF1. *Annu Rev Neurosci* 2010;33:221–243.
14. Wefel JS, Vardy J, Ahles T, Schagen SB. International Cognition and Cancer Task Force recommendations to harmonise studies of cognitive function in patients with cancer. *Lancet Oncol* 2011;12:703–708.
15. Green MF, Nuechterlein KH, Gold JM, et al. Approaching a consensus cognitive battery for clinical trials in schizophrenia: the NIMH-MATRICES conference to select cognitive domains and test criteria. *Biol Psychiatry* 2004;56:301–307.
16. Lystad JU, Falkum E, Mohn C, et al. The MATRICES consensus cognitive battery (MCCB): performance and functional correlates. *Psychiatry Res* 2014;220:1094–1101.
17. Green MF, Harris JG, Nuechterlein KH. The MATRICES consensus cognitive battery: what we know 6 years later. *Am J Psychiatry* 2014;171:1151–1154.
18. Drug Development Tools Qualification Programs: Clinical Outcome Assessment Qualification Program. Available at: <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284077.htm>. Accessed June 26, 2015.
19. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. 2007. Available at: <http://www.fda.gov/downloads/Drugs/Guidances/ucm071590.pdf>. Accessed July 23, 2013.
20. Wolters PL, Martin S, Merker VL, et al. Patient-reported outcomes in neurofibromatosis and schwannomatosis clinical trials. *Neurology* 2013;81(suppl 1):S6–S14.
21. Wechsler D, Kaplan E, Fein D, et al. Wechsler Intelligence Scales for Children. 4th ed. San Antonio, TX: Harcourt Assessment; 2004.
22. Roid GH. Stanford-Binet Intelligence Scales: Technical Manual. 5th ed. Itasca, IL: Riverside Publishing; 2003.
23. Wechsler D. Wechsler Intelligence Scale for Children. 4th ed. San Antonio, TX: Pearson; 2003.
24. Wechsler D. Wechsler Adult Intelligence Scale. 4th ed. San Antonio, TX: Pearson; 2008.

25. Robertson IH, Ward T, Ridgeway V, Nimmo-Smith I. The Test of Everyday Attention. Bury St. Edmunds, UK: Thames Valley Test Company; 1994.
26. Manly T, Robertson IH, Anderson V, Nimmo-Smith I. The Test of Everyday Attention for Children: Manual. Bury St. Edmunds, UK: Thames Valley Test Company; 1999.
27. A randomized placebo-controlled study of lovastatin in children with neurofibromatosis type 1—full text view ClinicalTrials.gov. Available at: <https://www.clinicaltrials.gov/ct2/show/NCT00853580?term=STARS+neurofibromatosis&rank=1>. Accessed June 26, 2015.
28. Løhaugen GCC, Antonsen I, Håberg A, et al. Computerized working memory training improves function in adolescents born at extremely low birth weight. *J Pediatr* 2011;158:555–561.
29. Hellwig-Brida S, Daseking M, Keller F, Petermann F, Goldbeck L. Effects of methylphenidate on intelligence and attention components in boys with attention-deficit/hyperactivity disorder. *J Child Adolesc Psychopharmacol* 2011;21:245–253.
30. Iverson GL, Green P. Measuring improvement or decline on the WAIS-R in inpatient psychiatry. *Psychol Rep* 2001;89:457–462.
31. DuPaul GJ, Anastopoulos AD, Power TJ, Reid R, Ikeda MJ, McGoey KE. Parent ratings of attention-deficit/hyperactivity disorder symptoms: factor structure and normative data. *J Psychopathol Behav Assess* 1998;20:83–102.
32. Conners CK. Conners. 3rd ed. Toronto: Multi-Health Systems; 2008.
33. Conners CK, Erhardt D, Sparrow EP. Conners' Adult ADHD Rating Scales (CAARS) Technical Manual. North Tonawanda, NY: Multi-Health Systems; 1999.
34. Mautner VF, Kluwe L, Thakker SD, Lark RA. Treatment of ADHD in neurofibromatosis type 1. *Dev Med Child Neurol* 2002;44:164–170.
35. Krab LC, de Goede-Bolder A, Aarsen FK, et al. Effect of simvastatin on cognitive functioning in children with neurofibromatosis type 1: a randomized controlled trial. *J Am Med Assoc* 2008;300:287–294.
36. Acosta MT, Kardel PG, Walsh KS, Rosenbaum KN, Gioia GA, Packer RJ. Lovastatin as treatment for neurocognitive deficits type 1: phase I study. *Pediatr Neurol* 2011;45:241–245.
37. Lion-Francois L, Kemlin I. Comportemental and Neuropsychologic Study of Children With Neurofibromatosis Type 1 Treated by Methylphenidate: A Double-blind Randomised Study: Methylphenidate Versus Placebo. Bethesda, MD: National Library of Medicine; 2016.
38. Van der Vaart T, Plasschaert E, Rietman AB, et al. Simvastatin for cognitive deficits and behavioural problems in patients with neurofibromatosis type 1 (NF1-SIMCODA): a randomised, placebo-controlled trial. *Lancet Neurol* 2013;12:1076–1083.
39. Lion-François L, Gueyffier F, Mercier C, et al. The effect of methylphenidate on neurofibromatosis type 1: a randomised, double-blind, placebo-controlled, crossover trial. *Orphanet J Rare Dis* 2014;9:142.
40. Bearden CE, Helleman GS, Rosser T, et al. A randomized placebo-controlled lovastatin trial for neurobehavioral function in neurofibromatosis type 1. *Ann Clin Trans Neurol* 2016;3:266–279.
41. Elgersma Y, Moll HA. The Effect of Lamotrigine on Cognitive Deficits Associated With Neurofibromatosis Type 1: A Phase II Randomized Controlled Multi-centre Trial (NF1-EXCEL) (NLM identifier: NCT02256124). In: ClinicalTrials.gov [Internet]. Available at: <https://clinicaltrials.gov/ct2/show/NCT02256124?term=neurofibromatosis+lamotrigine&rank=1>. Accessed March 7, 2016.
42. Ryan JJ, Glass LA, Bartels JM. Stability of the WISC-IV in a sample of elementary and middle school children. *Appl Neuropsychol* 2010;17:68–72.
43. Reynolds CR. Forward and backward memory span should not be combined for clinical analyses. *Neuropsychol Rev* 1997;12:29–40.
44. Rosenthal EN, Riccio CA, Gsanger KM, Jarratt KP. Digit Span components as predictors of attention problems and executive functioning in children. *Arch Clin Neuropsychol* 2006;21:131–139.
45. Manly T, Anderson V, Nimmo-Smith I, Turner A, Watson P, Robertson IH. The differential assessment of children's attention: the Test of Everyday Attention for Children (TEA-Ch), normative sample and ADHD performance. *J Child Psychol Psychiatry* 2001;42:1065–1081.
46. Wilde N, Strauss E. Functional equivalence of WAIS-III/WMS-III digit and spatial span under forward and backward recall conditions. *Clin Neuropsychol* 2002;16:322–330.
47. Isenberg JC, Templer A, Gao F, Titus JB, Gutmann DH. Attention skills in children with neurofibromatosis type 1. *J Child Neurol* 2012;28:45–49.
48. Gomez R, Burn GL, Walsh JA, de Moura MA. A multitrait-multisource confirmatory factor analytic approach to the construct validity of ADHD rating scales. *Psychol Assess* 2003;15:3–16.
49. Barkley RA, Murphy KR. Attention-Deficit Hyperactivity Disorder: A Clinical Workbook. 2nd ed. New York: The Guilford Press; 1998.
50. Wolraich ML, Lambert W, Doffing MA, Bickman L, Simmons T, Worley K. Psychometric properties of the Vanderbilt ADHD diagnostic parent rating in a referred population. *J Pediatr Psychol* 2003;28:559–568.