RESEARCH ARTICLE OPEN ACCESS

Sensitivity of the NIH Toolbox to Detect Cognitive Change in Individuals With Intellectual and Developmental Disability

Rebecca H. Shields, MS, Aaron Kaat, PhD, Stephanie M. Sansone, PhD, Claire Michalak, Jeanine Coleman, PhD, Talia Thompson, PhD, Forrest J. McKenzie, BS, Andrew Dakopolos, PhD, Karen Riley, PhD, Elizabeth Berry-Kravis, MD, Keith F. Widaman, PhD, Richard C. Gershon, PhD, and David Hessl, PhD

Correspondence
Dr. Hessl
drhessl@ucdavis.edu

Neurology® 2023;100:e778-e789. doi:10.1212/WNL.0000000000201528

Abstract

Background and Objective

Individuals with intellectual disability (ID) experience protracted cognitive development compared with typical youth. Sensitive measurement of cognitive change in this population is a critical need for clinical trials and other intervention studies, but well-validated outcome measures are scarce. This study's aim was to evaluate the sensitivity of the NIH Toolbox Cognition Battery (NIHTB-CB) to detect developmental changes in groups with ID—fragile X syndrome (FXS), Down syndrome (DS), and other ID (OID)—and to provide further support for its use as an outcome measure for treatment trials.

Methods

We administered the NIHTB-CB and a reference standard cross-validation measure (Stanford-Binet Intelligence Scales, Fifth Edition [SB5]) to 256 individuals with FXS, DS, and OID (ages 6–27 years). After 2 years of development, we retested 197 individuals. Group developmental changes in each cognitive domain of the NIHTB-CB and SB5 were assessed using latent change score models, and 2-year growth was evaluated at 3 age points (10, 16, and 22 years).

Results

Overall, effect sizes of growth measured by the NIHTB-CB tests were comparable with or exceeded those of the SB5. The NIHTB-CB showed significant gains in almost all domains in OID at younger ages (10 years), with continued gains at 16 years and stability in early adulthood (22 years). The FXS group showed delayed gains in attention and inhibitory control compared with OID. The DS group had delayed gains in receptive vocabulary compared with OID. Unlike the other groups, DS had significant growth in early adulthood in 2 domains (working memory and attention/inhibitory control). Notably, each group's pattern of NIHTB-CB growth across development corresponded to their respective pattern of SB5 growth.

Discussion

The NIHTB-CB is sensitive to developmental changes in individuals with ID. Comparison with levels and timing of growth on the cross-validation measure shows that the NIHTB-CB has potential to identify meaningful trajectories across cognitive domains and ID etiologies. Sensitivity to change within the context of treatment studies and delineation of clinically meaningful changes in NIHTB-CB scores, linked to daily functioning, must be established in future research to evaluate the battery more completely as a key outcome measure.

From the MIND Institute and Department of Psychiatry and Behavioral Sciences (R.H.S., S.M.S., F.J.M., A.D., D.H.), University of California Davis, Sacramento; Northwestern University Feinberg School of Medicine (A.K., R.C.G.), Chicago, IL; Rush University Medical Center Departments of Pediatrics (C.M., E.B.), Neurological Sciences and Biochemistry, Chicago, IL; University of Denver Morgridge College of Education (J.C.), Denver, CO; University of Colorado School of Medicine (T.T.), Aurora; Regis University (K.R.), Denver, CO; and University of California Riverside Graduate School of Education (K.F.W.), Riverside.

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

The Article Processing Charge was funded by the NIH.

Glossary

CSS = change sensitive score; **DCCS** = Dimensional Change Card Sort; **DS** = Down syndrome; **DSM-5** = *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*; **EGCG** = epigallocatechin-3-gallate; **FICA** = Flanker Inhibitory Control and Attention; **FSIQ** = Full Scale IQ; **FXS** = fragile X syndrome; **ID** = intellectual disability; **LSWM** = List Sorting Working Memory; **NIHTB-CB** = NIH Toolbox Cognition Battery; **OID** = other ID; **ORR** = Oral Reading Recognition; **PCPS** = Pattern Comparison Processing Speed; **PDE4** = phosphodiesterase-4D; **PSM** = Picture Sequence Memory; **PV** = Picture Vocabulary; **SB5** = Stanford-Binet Intelligence Scales, Fifth Edition; **USS** = uncorrected standard score.

Intellectual disability (ID) affects approximately 1.8%–3.2% of the global population.[1,2] A diagnosis of ID affects daily life, academic achievement, and independence, and it involves substantial economic burden for the family and society.[2-4] ID is a lifelong disorder, characterized by cognitive impairment (IQ of approximately 2 or more SDs below average) plus adaptive behavior deficits. Fragile X syndrome (FXS) and Down syndrome (DS) have been at the forefront of translational research for targeted treatments of ID-associated disorders.[5-7] Numerous classes of medications have been investigated in human trials, such as mGluR5 antagonists, GABA agonists, ERK inhibitors, matrix metalloproteinase-9 activators, and a phosphodiesterase-4D (PDE4) allosteric inhibitor for FXS[8]; and a $GABA_A$ inverse agonist, memantine, rivastigmine, and the green tea extract containing epigallocatechin-3-gallate (EGCG) for DS.[9-12] Many clinical trials in FXS and DS have fallen short of expectations, possibly because of challenges in translation from animal models to humans, proof of target engagement at efficacious and safe dosages in the human brain, or inadequately sensitive and valid outcome measures.[6,13,14]

Measuring cognition in ID-associated disorders is of particular importance. Clinical trials for FXS and DS are dependent on accurate cognition assessment for screening, patient sample characterization, and detecting treatment-based changes. Full Scale IQ tests, such as the Stanford-Binet Intelligence Scales, Fifth Edition (SB5), are sometimes used. Although IQ tests are rigorous, there are some disadvantages in the clinical trial setting. Assessments of trial outcomes are repeated in relatively close intervals, and IQ tests are not designed for this purpose. They are also extensive, often requiring 2+ hours to complete; this is a disadvantage for individuals with ID, who commonly have behavioral difficulties, anxiety, or fatigue that limits the amount of accurate testing that is possible in a visit. The NIH Toolbox Cognition Battery (NIHTB-CB) is an iPad-based assessment that has potential within these groups and in the clinical trial setting. Consisting of brief memory, executive function, processing speed, and language tests, the NIHTB-CB was developed under contract issued by the NIH Blueprint for Neuroscience Research (nihtoolbox.org). Through extensive literature reviews and piloting, domains and tasks were selected that are relevant to health, are developmentally sensitive from preschool through late adulthood, and have established links to specific brain functions.[15] Compared with IQ tests, the NIHTB-CB is shorter (approximately 30–60 minutes), with a heavier emphasis on fluid reasoning and potentially more changeable aspects of executive function, and the iPad format is appealing and accessible for children and adults with ID.[16]

Because of greater attention to trial design stemming from the lessons learned in treatment studies, there have been exciting successes. For example, by combining cognitive training with EGCG vs placebo in a 12-month double-blind, trial in adults with DS, de la Torre and colleagues[9] demonstrated improvement on selected measures of memory, inhibitory control, and caregiver-reported functional academics in those treated with EGCG. In FXS, our extensive psychometric studies of the NIHTB-CB[17,18] supported its use as a key efficacy outcome in a 24-week phase-2 randomized, placebo-controlled, crossover trial of a PDE4 allosteric inhibitor (BPN14770) in 30 adult males with FXS.[19] Cognitive benefit was demonstrated based on improvement on the NIHTB-CB Crystallized Cognition Composite (composed of Oral Reading Recognition [ORR] and Picture Vocabulary [PV] tests). Benefit as assessed by visual analog caregiver rating was also clinically meaningful for language and daily functioning. Although the latter trial provided preliminary evidence of the NIHTB-CB's sensitivity to change, few studies have evaluated its capacity to detect developmental changes[20] and none yet in ID. The delineation of the natural history of specific IDs is critical for understanding whether targeted treatments can positively alter the expected trajectory of cognitive growth for a neurodevelopmental disorder.

The overarching purpose of our program of research focused on the NIHTB-CB is to evaluate its fitness as a battery of cognitive outcome measures for treatment studies for persons with ID. Our prior work has established that each NIHTB-CB test is feasible, reliable, and valid for individuals with a mental age of 5 years or higher; furthermore, several of the tests show sound psychometric properties extending to mental ages as low as 3 years. As highlighted above, another critical criterion of measure fitness for an outcome measure is responsiveness—or sensitivity to true changes in functioning. Thus, the aim of this study was to evaluate the capacity for the NIHTB-CB to detect change in cognitive growth in groups with ID during developmental periods when gains are expected.

As the next step in evaluating the promise of the NIHTB-CB for IDs, we conducted 2-year longitudinal assessments to evaluate the sensitivity of the battery, in comparison with the SB5, to detect expected cognitive growth from childhood through early adulthood. Latent change score models evaluated levels and growth of domain-specific cognition across development and between groups—FXS, DS, and OID. We expected that, although growth would be smaller and more gradual in ID than in typical development, children with ID would show a greater rate of cognitive growth than adolescents or adults with ID. Furthermore, we hypothesized that there are distinct developmental cognitive profiles rather than a globally slow rate of growth across syndromes and domains.

Methods

Standard Protocol Approvals, Registrations, and Patient Consents

Institutional review board approval was obtained at each site before study initiation. Written consent was obtained from each guardian (or adult participant in the case of 5 individuals who were capable to give their own consent).

Participants

As part of a multisite longitudinal study, eligible participants were between 6 and 25 years at visit 1, with a diagnosis of ID or suspected ID. During visit 1, ID or borderline ID criteria were based on the DSM-5[21]—adaptive behavior deficits measured by the Vineland Adaptive Behavior Scales, Third Edition (Vineland-3)[22] and IQ < 80 on the SB5. Three groups were recruited: FXS (full mutation, with genetic confirmation), DS (with genetic confirmation if possible), and OID (with genetic confirmation of negative fragile X mutation). A minimum mental age equivalent of 3 years as measured by the SB5 was required in accordance with NIHTB-CB age limits. Participants were required to be stable with usual treatment for at least 4 weeks before each visit. Recruitment sources included research registries, flyers at local clinics, announcements through parent support foundation websites, and mailings to families registered with the California Department of Developmental Services. An additional 54 individuals were ineligible: 20 with IQ >79 and 34 with mental age below 3 years. Full protocol, details of the NIHTB-CB, and its performance at baseline in the present ID samples has been reported previously.[17,18]

Measures

The NIHTB-CB[23] includes 7 tests: Flanker Inhibitory Control and Attention (FICA), Dimensional Change Card Sort (DCCS), List Sorting Working Memory (LSWM), Pattern Comparison Processing Speed (PCPS), Picture Sequence Memory (PSM), Picture Vocabulary (PV), and ORR.[18] A published manual of standardized administration procedures for ID can be found in Ref. 24.

The SB5, which is standardized for individuals between 2-85 years, provides an overall index of intellectual ability reported as the Full Scale IQ (FSIQ). In part due to its broad developmental range, the SB5 has performed well in our prior studies of ID.[17,18,25] In addition to providing standard IQ scores, the SB5 provides change sensitive scores (CSSs).[26] The CSSs indicate performance based on the average score at a certain age. They are criterion-referenced scores on an equal-interval scale. The CSSs have a centered value of 500, which represents the average performance of a 10-year-old in the general population. The CSSs range from approximately 425, the average level of 2-year-olds, to 525, the average level of adults. This metric allows for longitudinal comparison. The CSS for FSIQ was used as the reference standard measure.

Non–age-adjusted scores were also used for Toolbox tests. For DCCS, LSWM, PCPS, PSM, PV, and ORR, uncorrected standard scores (USSs) were used, which have a mean of 100 and SD of 15. The USSs are recommended for longitudinal measurement because, like the CSSs, they are not adjusted based on age-related growth of normative peers. For FICA and DCCS, creation of alternative scores was necessary. Many participants had difficulty understanding and following task demands for FICA and DCCS. In addition, these tests involve different phases depending on performance; it was thus difficult to compare scores longitudinally when participants received different phases across visits. To address this, we created prorated scores (FICA Pro and DCCS Pro), which incorporate accuracy with reaction time and are based on the same number and type of items. Both prorated scores showed strong reliability and validity. FICA Pro replaced the USS in analyses, and DCCS Pro was used in addition to the DCCS USS. Further details about the prorated score creation and psychometrics are available in eAppendix 1 (links.lww.com/WNL/C509).

Our protocol uses mental (rather than chronologic) age to select test versions.[24] Prior studies showed that participants with ID sometimes obtain ceiling or floor scores on PSM. As this was identified mid-study, some participants received only a ceiling or floor score and were not given another age version for a more accurate assessment. For consistency, PSM analyses include only participants without a floor or ceiling score, who received the same test items at both visits (mental age at both visits fell within one version) or who received a nonfloor, nonceiling score on 2 versions, in alignment with a longitudinal change in mental age.

Model Specification

To measure developmental change in each domain, a univariate latent change score model was specified for each test using observed scores from visits 1 and 2.[27-30] Latent change score models are a type of structural equation modeling that provides estimates of change as latent variables based on 2 or more time points. Group covariates can be included to evaluate group-specific change. Change was modeled as a function of age, group, age × group interaction, sex, and sex × group interaction. Age was centered at 3 points across the sample range (10, 16, and 22 years) to represent the preadolescent period, middle of adolescence, and early adulthood. Autism is a frequent comorbidity with ID and often associated with its own cognitive traits. Therefore, parent-reported autism diagnostic status was included as an observed indicator with no regression effect on outcomes, as we did not hypothesize an effect of autism diagnosis. Models controlled for time between visits. Missing data were handled with full information maximum likelihood estimation, which is a standard recommendation to provide accurate parameter estimates in the presence of missing data.[31] Analyses included all participants with a valid score, even without completion of visit 2. For each test, the preliminary model allowed predictor variables to vary, other than autism diagnosis, whose effect was fixed to zero. As we hypothesized different group trajectories, group coefficients were free to vary. Preliminary models were modified in 2 steps to improve fit: first evaluating the effect of sex and sex × group interactions and second evaluating age × group interactions. eFigure 1 graphically presents a generic representation of this model, and eTable 1(links.lww.com/WNL/C509) gives further details of each model's specification.

Data Availability

Data are available from the NIMH Data Archive (nda.nih.gov/)—ID C3738.

Results

Descriptive Statistics

A total of 256 eligible participants were seen for visit 1, with 197 completing retesting (visit 2) by the time of the present analyses (see Table 1 for descriptive statistics of each group). Of the participants with only visit 1 data, 10 discontinued before visit 2 (5 moved away and 5 withdrew due to scheduling difficulty or challenging child behaviors at the time). At analysis, 17 expected participants have not completed visit 2; many of these remain interested but could not be seen because of the COVID-19 pandemic, or rescheduling has not yet been established. An additional 54 individuals were ineligible at visit 1 and were not included: 20 with IQ >79 and 34 with mental age below 3 years.

Data Cleaning and Scoring

Following standardized procedures developed through prior studies,[17,18,24] only scores deemed valid by the trained examiner were used. Across visits, only 7.3% of scores were excluded; the most frequent reasons for exclusion were an invalid response pattern that could not be corrected by feedback, excessive prompting, and behavioral difficulties. Table 1 provides sample sizes by group and visit, representing only data included in analyses. Full details of sample sizes by age, group, and test are available in eTable 2 (links.lww.com/WNL/C509).

The median time between visits was 25.92 months (interquartile range = 24.62–28.08). Groups did not differ on age [$F(2, 253) = 1.59$, $p = 0.21$] or the Vineland-3 Adaptive Behavior Composite [$F(2, 239) = 1.37$, $p = 0.25$]. However, FSIQ was significantly different across groups [$F(2, 252) = 34.28$, $p < 0.001$], with higher FSIQ in OID than both DS [$t(252) = 7.87$, $p < 0.001$, 95% CI: 11.67–21.66] and FXS [$t(252) = 6.13$, $p < 0.001$, 95% CI: 8.36–18.81]. FSIQ did not differ between DS and FXS.

Latent Change Score Models

Results from the latent change score models include means and variances for the latent intercept and change score at 10, 16, and 22 years. For interpretation purposes, intercepts and change scores represent points on the test's scale (USS, prorated, or CSS; see Measures). Change scores reflect latent change after 2 years. For tests using USSs, change scores are in the standard score metric (mean = 100, SD = 15), whereby a +5 change score corresponds to an increase of one-third SD in the normative sample (Cohen's $d = 0.33$). For example, an age 10 change score of 5 on PV would indicate that on average, 10-year-olds improved by 5 USS points by age 12 years.

Tables 2–4 present model parameters for OID, FXS, and DS, respectively. Results below highlight key change score results, significant group differences of change, and significant group differences of the timing of growth (i.e., change as a function of age × group interactions). In eTables 3 and 4 (links.lww.com/WNL/C509), additional results from each model are provided for covariates regressed on the change score, including coefficients for age, sex, interval, group, and age × group and sex × group interactions.

To visually illustrate developmental patterns by group, graphs were created for each domain based on mean change scores and intercepts (Figures 1 and 2). Graphs provide visual estimates of growth based on 3 key components: group-specific age 10 intercept, change scores at 10, 16, and 22 years, and the regression coefficient of age on change scores. The graphs are not representations of the complete model parameters because a visual trajectory encompassing complete model parameters was not possible with only 2 time points. The graphing method can be understood in 3 steps. First, the visit 1 intercept at age 10 years (representing the latent mean level) was plotted. Next, moving along the x-axis, the change score estimate was added to that intercept each 2 years. Finally, the regression coefficient of age on change scores was added to each point (i.e., growth is generally steeper at younger ages). For the years below 10, the same process was followed in reverse using the age 10 change score; the change score was subtracted for each 2 years below 10, minus the coefficient of age on change.

Stanford-Binet 5

The estimated change for SB5 Full Scale CSSs was significant at age 10 years for all groups, with effect sizes ranging from 0.22 to 0.40 (Cohen's d). DS growth continued to be significant at 16 years (est = 2.8, SE = 0.8, $p < 0.001$). There was a significant effect of group on change, such that DS had significantly greater growth at age 22 years than OID (b = 3.1, SE = 1.6, $p = 0.048$).

NIHTB Fluid Composite Tests

Flanker Inhibitory Control and Attention (FICA Pro)

Estimated growth on FICA Pro was significant in all groups at age 10 years (Cohen's d, 0.56–0.84) and 16 years (Cohen's d, 0.28–0.56), indicating improvement over 2 years of development. The DS group showed continued significant

Table 1 Participant Descriptive Information

	Other ID (n = 86)	Fragile X syndrome (n = 78)	Down syndrome (n = 92)
Race			
American Indian/Alaska Native	1.16%	1.28%	1.09%
Asian	2.33%	2.56%	2.17%
Black	13.95%	10.26%	5.43%
Native Hawaiian/Pacific Islander	1.16%	1.28%	1.09%
White	54.65%	80.77%	76.09%
More than one	22.09%	2.56%	10.87%
Ethnicity (% Hispanic or Latino)	29.07%	7.69%	18.48%
Sex (% male)	62.79%	73.08%	44.56%
Primary caregiver 4-year degree	56.98%	65.38%	65.22%
Autism diagnosis (parent-report)	54.65% (2 unknown)	51.28% (5 unknown)	6.52% (0 unknown)

	Other ID					
Chronologic age (y)	**Mean (SD)**	**Fragile X syndrome**	**Down syndrome**	***n* with a valid score: visit 1 and 2**		
	14.71 (5.27)	15.95 (4.93)	16.27 (5.13)			
Vineland-3 ABC	52.62 (16.29)	50.11 (18.86)	54.63 (16.38)			
SB5 Full Scale mental age (y)	6.07 (1.62)	4.91 (1.38)	4.72 (1.20)	**Other ID**	**Fragile X syndrome**	**Down syndrome**
SB5 Full Scale deviation IQ	64.25 (14.34)	50.67 (15.29)	48.13 (12.68)	86, 58	78, 66	92, 69
SB5 Full Scale CSS	477.51 (12.69)	467.82 (12.70)	466.23 (10.87)	86, 58	78, 66	92, 70
FICA incongruent prorated	0.73 (0.41)	0.56 (0.30)	0.55 (0.32)	57, 44	45,47	74, 61
DCCS USS	71.28 (24.79)	60.90 (20.34)	61.43 (18.98)	68, 47	42, 38	55, 51
DCCS prorated	0.80 (0.39)	0.67 (0.31)	0.59 (0.24)	46, 40	27, 25	37, 35
LSWM USS	67.53 (17.79)	59.78 (13.69)	52.21 (14.00)	68, 49	46, 44	58, 50
PCPS USS	79.31 (22.86)	72.36 (18.09)	62.42 (17.86)	70, 46	58, 39	60, 49
PSM USS	88.34 (14.19)	79.94 (12.62)	79.12 (11.82)	70, 51	50, 58	69, 68
PV USS	71.76 (13.22)	69.37 (12.45)	64.43 (13.09)	85, 59	75, 66	89, 69
ORR USS	78.00 (14.42)	74.27 (13.02)	73.15 (15.63)	82, 57	73, 61	87, 68

Abbreviations: ABC = Adaptive Behavior Composite standard score; CSS = Change Sensitive Score; DCCS = Dimensional Change Card Sort; FICA = Flanker Inhibitory Control and Attention; ID = intellectual disability; LSWM = List Sorting Working Memory; ORR = Oral Reading Recognition; PCPS = Pattern Comparison Processing Speed; PSM = Picture Sequence Memory; PV = Picture Vocabulary; SB5 = Stanford-Binet Intelligence Scales, Fifth Edition; USS = uncorrected standard score.

growth through age 22 years (est = 0.18, SE = 0.06, p = 0.002, d = 0.56). Matching the SB5's pattern of group comparison, at age 22 years, DS had greater 2-year growth on FICA than the OID group (b = 0.20, SE = 0.01, p = 0.007).

Dimensional Change Card Sort

For Dimensional Change Card Sort (DCCS) USSs, OID and FXS had significant growth at age 10 years (Cohen's d, 0.62 and 0.54), and this continued to be significant at age 16 years for OID (est = 6.09, SE = 3.07, p = 0.047, d = 0.27). The DS group showed no significant 2-year growth at any age point. On the DCCS Pro score, no group showed a significant 2-year change.

List Sorting Working Memory

On List Sorting Working Memory (LSWM), groups had unique growth across age. The OID group had a significant change at age 10 years (est = 11.89, SE = 2.53, $p < 0.001$) and 16 years (est = 6.78, SE = 2.08, p = 0.001), with gains no longer significant by 22 years. For FXS, 2-year growth on LSWM was significant only at age 10 years (est = 8.89, SE = 2.81, p = 0.002). The DS group, however, displayed significant growth through age 22 years (10 years: est = 7.94, SE = 3.03, $p < 0.001$; 16 years: est = 6.96, SE = 2.11, p = 0.001; 22 years: est = 5.97, SE = 2.84, p = 0.036). At 22 years, DS growth on LSWM approached a significant difference compared with FXS, with DS displaying more improvement (b = 7.30, SE = 3.72, p = 0.05).

Table 2 Latent Estimates for Visit 1 Score and 2-Year Change Scores in the Other ID Group

					Age 10					Age 16					Age 22				
Test	**Visit 1 variance**	**SE**	**Change score variance**	**SE**	**Visit 1 intercept**	**SE**	**Change score**	**SE**	**Change effect size[a]**	**Visit 1 intercept**	**SE**	**Change score**	**SE**	**Change effect size[a]**	**Visit 1 intercept**	**SE**	**Change score**	**SE**	**Change effect size[a]**
SB5 CSS	114.1	10.1	38.1	3.9	473.5	2.0	5.2***	1.2	0.40	478.9	1.7	1.6	0.9	0.12	484.3	2.3	−2.0	1.3	−0.15
Fica Pro	0.1	0.0	0.1	0.0	0.7	0.1	0.3***	0.1	0.84	0.7	0.0	0.1*	0.1	0.28	0.8	0.1	0.0	0.1	0.00
DCCS USS	405.4	45.1	425.3	55.8	64.9	3.6	13.8***	3.7	0.62	74.0	3.3	6.1*	3.1	0.27	83.1	3.9	−1.6	4.2	−0.07
DCCS Pro	0.1	0.0	0.1	0.0	0.7	0.1	0.0	0.1	0.00	0.8	0.1	0.0	0.0	0.00	0.9	0.1	0.0	0.1	0.00
LSWM	200.2	21.4	176.4	23.3	60.9	2.6	11.9***	2.5	0.71	66.6	2.4	6.8**	2.1	0.41	72.4	2.7	1.7	2.6	0.10
PCPS	341.4	34.7	172.3	22.6	75.1	3.4	6.9**	2.4	0.33	79.7	3.1	2.3	2.0	0.11	84.4	3.6	−2.2	2.5	−0.10
PSM	152.3	15.7	158.3	19.3	88.9	2.1	5.7**	2.1	0.42	89.7	2.0	2.7	1.9	0.20	90.6	2.3	−0.3	2.3	−0.02
PV	127.8	11.4	60.3	6.2	65.8	2.0	8.0***	1.6	0.60	72.6	1.8	4.1**	1.4	0.31	79.3	2.1	0.2	1.7	0.02
ORR	166.7	15.1	51.4	5.5	73.8	2.4	5.3***	1.4	0.36	81.1	2.0	3.2**	1.0	0.22	88.4	2.8	1.0	1.5	0.07

Abbreviations: CSS = change sensitive score; DCCS = Dimensional Change Card Sort; FICA = Flanker Inhibitory Control and Attention; LSWM = List Sorting Working Memory; ORR = Oral Reading Recognition; PCPS = Pattern Comparison Processing Speed; Pro = prorated score; PSM = Picture Sequence Memory; PV = Picture Vocabulary; SB5 = Stanford Binet Intelligence Scales, Fifth Edition, Full Scale = USS, uncorrected standard score.
*$p < .05$; **$p < .01$; and ***$p < .001$.
[a] Cohen's *d*.

Table 3 Latent Estimates for Visit 1 Score and 2-Year Change Scores in Fragile X Syndrome[a]

	Age 10							Age 16							Age 22						
	Males		Females				Change effect size[b]	Males		Females				Change effect size[b]	Males		Females				Change effect size[b]
Test	Visit 1 intercept	SE	Visit 1 intercept	SE	Change score	SE		Visit 1 intercept	SE	Visit 1 intercept	SE	Change score	SE		Visit 1 intercept	SE	Visit 1 intercept	SE	Change score	SE	
SB5 CSS	460.0	1.8	476.6	2.5	2.9**	1.0	0.22	463.7	1.5	480.3	2.4	1.2	0.8	0.09	467.4	1.7	484.0	2.6	−0.5	1.0	−0.04
FICA Pro[c]	0.5	0.1	0.5	0.1	0.2*	0.1	0.56	0.6	0.1	0.6	0.1	0.1*	0.1	0.28	0.6	0.1	0.6	0.1	0.1	0.1	0.28
DCCS USS	46.2	4.4	59.5	4.7	12.0**	4.4	0.54	55.3	3.8	68.6	4.4	4.2	3.6	0.19	64.3	4.1	77.7	4.8	−3.5	4.2	−0.16
DCCS Pro[c]	0.6	0.1	0.6	0.1	0.1	0.1	0.30	0.7	0.1	0.1	0.1	0.1	0.1	0.30	0.8	0.1	0.8	0.1	0.1	0.1	0.30
LSWM	45.8	3.0	61.2	3.3	8.9**	2.8	0.53	51.6	2.6	67.0	3.1	3.8	2.2	0.23	57.3	2.7	72.7	3.3	−1.3	2.6	−0.08
PCPS	60.8	3.4	76.4	4.3	9.6***	2.7	0.46	65.5	3.0	81.0	4.1	5.0*	2.2	0.24	70.1	3.3	85.7	4.4	0.5	2.5	0.02
PSM	76.4	2.2	90.1	3.0	5.5*	2.2	0.40	77.2	1.9	91.0	2.9	2.6	1.9	0.19	78.1	2.2	91.8	3.1	−0.4	2.3	−0.03
PV (male)[d]	59.5	1.9	—	4.8**	1.5	0.36	66.3	1.6		—	0.9	1.2	0.07	73.0	1.8		—	−3.0*	1.4	−0.23	
PV (female)[d]	—		71.7	2.7	6.5**	1.9	0.49	—		78.5	2.5	2.6	1.8	0.20	—		85.2	2.8	−1.2	2.0	−0.09
ORR	676	2.2	84.6	3.0	0.9	1.2	0.06	69.9	1.8	86.9	2.8	−0.1	1.0	0.00	72.1	2.1	89.2	3.1	−1.2	1.2	−0.08

Abbreviations: CSS = Change Sensitive Score; DCCS = Dimensional Change Card Sort; FICA = Flanker Inhibitory Control and Attention; FXS = fragile X syndrome; LSWM = List Sorting Working Memory; ORR = Oral Reading Recognition; PCPS = Pattern Comparison Processing Speed; PSM = Picture Sequence Memory; PV = Picture Vocabulary; SB5 = Stanford-Binet Intelligence Scales, Fifth Edition, Full Scale; USS = uncorrected standard score.
**p* < 0.05; ***p* < 0.01; and ****p* < 0.001.

[a] Latent variances for visit 1 and change scores are provided in this table.

[b] Cohen's *d*.

[c] One visit 1 intercept is provided across sexes; these models did not include FXS × sex interaction as a predictor of visit 1 intercept.

dSeparate change score is provided for each sex; these models included FXS × sex interaction as a predictor of change score.

Table 4 Latent Estimates for Visit 1 Score and 2-Year Change Scores in Down Syndrome[a]

	Age 10					Age 16					Age 22				
Test	Visit 1 intercept	SE	Change score	SE	Change effect size[b]	Visit 1 intercept	SE	Change score	SE	Change effect size[b]	Visit 1 intercept	SE	Change score	SE	Change effect size[b]
SB5 CSS	462.2	1.7	4.5***	1.0	0.34	465.9	1.4	2.8***	0.8	0.21	469.6	1.7	1.1	1.0	0.08
FICA Pro	0.5	0.1	0.3***	0.1	0.84	0.5	0.0	0.2***	0.0	0.56	0.6	0.1	0.2**	0.1	0.56
DCCS USS	52.9	3.8	3.2	4.8	0.14	62.0	3.2	1.6	3.4	0.07	71.1	3.6	0.1	4.7	0.00
DCCS Pro	0.5	0.1	0.0	0.1	0.00	0.5	0.1	0.0	0.1	0.00	0.6	0.1	0.0	0.1	0.00
LSWM	44.3	2.6	7.9**	3.0	0.47	50.1	2.2	7.0**	2.1	0.42	55.9	2.5	6.0*	2.8	0.36
PCPS	56.2	3.2	11.2***	2.7	0.53	60.8	2.9	6.6**	2.3	0.31	65.5	3.2	2.1	2.6	0.10
PSM	78.8	2.1	7.4***	2.1	0.54	79.6	1.8	4.5*	1.8	0.33	80.4	2.0	1.5	2.1	0.11
PV	58.9	2.0	3.6*	1.6	0.27	63.8	1.5	2.9*	1.2	0.22	68.6	2.1	2.2	1.7	0.17
ORR	71.6	2.1	3.4**	1.2	0.23	73.9	1.8	2.3*	1.0	0.16	76.2	2.1	1.2	1.2	0.08

Abbreviations: CSS = Change Sensitive Score; DCCS = Dimensional Change Card Sort; FICA = Flanker Inhibitory Control and Attention; LSWM = List Sorting Working Memory; ORR = Oral Reading Recognition; PCPS = Pattern Comparison Processing Speed; PSM = Picture Sequence Memory; PV = Picture Vocabulary; SB5 = Stanford-Binet Intelligence Scales, Fifth Edition, Full Scale; USS = uncorrected standard score.
**p* < .05; ***p* < .01; and ****p* < .001.
[a] Latent variances for visit 1 and change scores are provided in this table.
[b] Cohen's *d*.

Pattern Comparison Processing Speed

Growth estimates on Pattern Comparison Processing Speed (PCPS) showed that all groups improved significantly at age 10 years (Cohen's *d*, 0.33-0.53). Two-year growth continued to be significant at age 16 years in FXS (*d* = 0.24) and in DS (*d* = 0.31).

Picture Sequence Memory—Episodic Memory

For Picture Sequence Memory (PSM), all groups had significant growth estimates at age 10 years (Cohen's *d*, 0.40–0.54). At age 16 years, change was only significant for DS (b = 4.47, SE = 1.76, *p* = 0.011). No group had a significant change score at age 22 years.

NIHTB Crystallized Composite Tests

Picture Vocabulary

For Picture Vocabulary (PV), all groups showed significant 2-year growth at age 10 years (Cohen's *d*, 0.27–0.60), and growth was still significant at age 16 years for OID and DS (Cohen's *d*, 0.31 and 0.22). At 10 years, the OID group, which had a higher starting level on PV than DS (b = 6.90, SE = 2.37, *p* = 0.004), also had a larger change score than DS (b = 4.38, SE = 1.94, *p* = 0.024). There was a significant DS × age interaction effect on PV change scores (b = 0.53, SE = 0.23, *p* = 0.021), indicating a more stable, flattened profile in DS compared with the other groups. At age 22 years, no groups improved significantly; in fact, males with FXS had a significant negative PV change score at this age (b = −3.00, SE = 1.38, *p* = 0.03).

Oral Reading Recognition

Change score estimates on Oral Reading Recognition (ORR) were significant in the OID and DS groups at age 10 years (Cohen's *d*, 0.36 and 0.23) and age 16 years (Cohen's *d*, 0.22 and 0.16). The FXS group did not have significant 2-year change at any age. There was a significant group effect on ORR change scores, such that FXS had a less change than OID at age 10 years (b = −4.37, SE = 1.81, *p* = 0.016) and age 16 years (*b* = −3.31, SE = 1.34, *p* = 0.014).

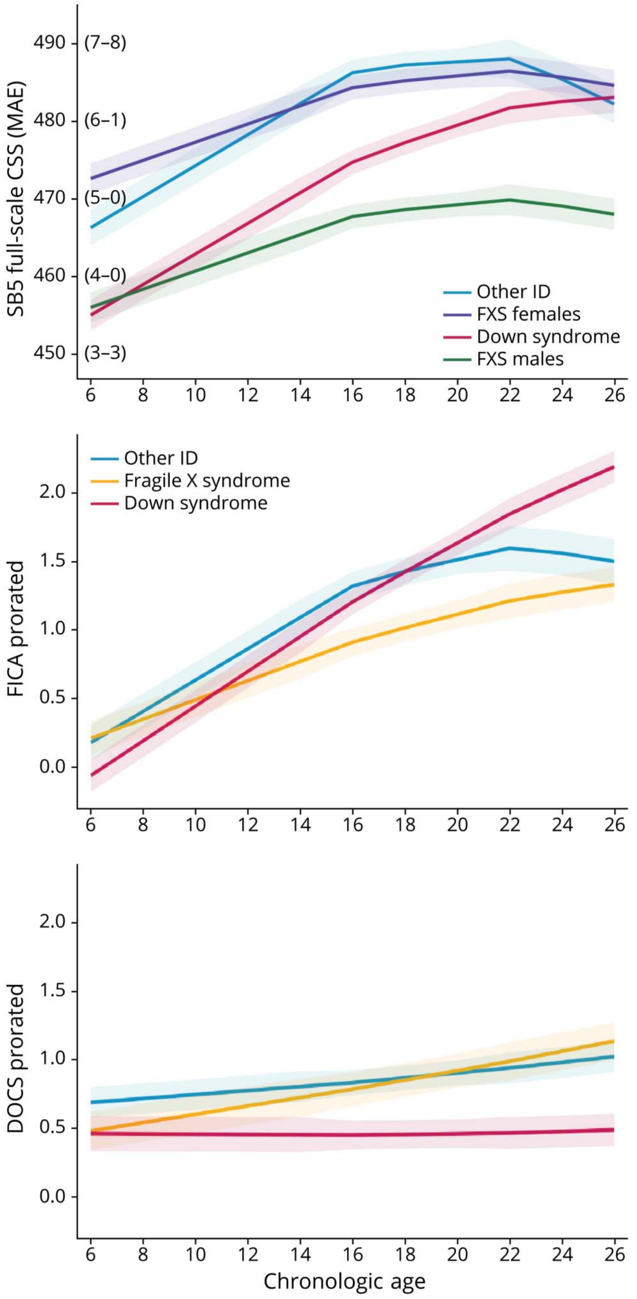
Effect of Sex on Latent Intercepts and Change Scores

There was a significant effect of sex on visit 1 intercept, with males having a lower starting level than females, for the SB5, PCPS, PSM, and ORR tests. For all models including sex as a predictor of visit 1 score (all tests except FICA Pro and DCCS Pro), there was a significant sex × group interaction; males with FXS had a lower starting level than females with FXS, whereas the other groups had no difference by sex. For the PV model, which included sex and sex × group as predictors of change scores, these effects did not significantly predict change scores. Detailed results for main effects of sex and sex × group interaction on intercepts and change scores are available in eAppendix 2 (links.lww.com/WNL/C509).

Discussion

The results of this study demonstrate that the NIHTB-CB detects significant 2-year growth in a variety of domains of cognition in youth with ID. As expected, these gains tended to be steeper in early school-aged children compared with those observed in adolescence and young adulthood. However, it is important to emphasize that some significant gains were seen even in mid-adolescence and early adulthood (e.g., attention/

Figure 1 Mean Developmental Growth Estimates for SB5, FICA Pro, and DCCS Pro



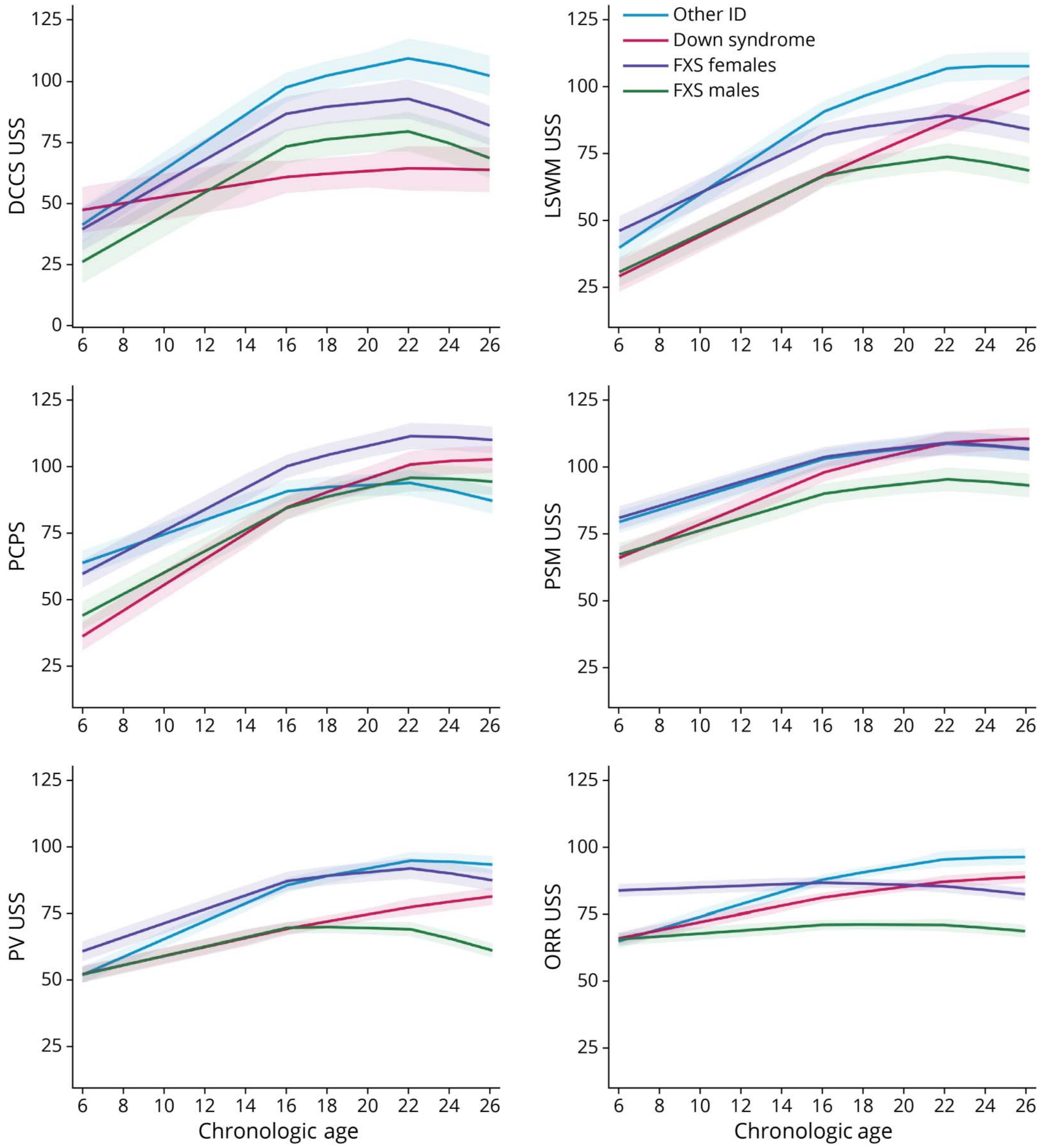
Graphs represent approximate mean trajectories of developmental growth based on each test's latent change score model. The approximate mean change is plotted by using the group-specific intercept at age 10 years, adding the group-specific change score per 2 years (centered at 10, 16, and 22 years) and adding the regression outcome of age on change scores. Shaded bands represent 95% CIs for the change score around each plotted level (every 2 years). SB5 change sensitive scores (CSSs) are a component of SB5 scoring and allow for precise measurement and raw comparison over time (i.e., without age adjustment). The CSSs are based on a centering constant of 500, which indicates the mean performance of a 10-year-old in the general population (i.e., a mental age equivalent of 10 years).[22] The y-axis shows mental age equivalents corresponding with CSSs (years-months). The FICA prorated score represents correct items per second in the fish portion. The DCCS prorated score represents correct items per second in the mixed shape and color test portion. DCCS = Dimensional Change Card Sort; FICA = Flanker Inhibitory Control and Attention; MAE = mental age equivalent; SB5 = Stanford-Binet Intelligence Scales, Fifth Edition.

inhibitory control in all groups; episodic memory in DS; working memory, receptive vocabulary, and oral reading in OID and DS; and processing speed in FXS and DS). Effect sizes of NIHTB-CB growth were comparable to, or exceeded, gains detected by the reference standard measure, the SB5 Full Scale CSS. Notably, in addition to effect sizes of growth, trajectories also aligned, as group-specific patterns of NIHTB-CB growth paralleled the SB5. For example, DS significant growth was detected by nearly all measures up to age 16 years (FICA Pro, LSWM, PCPS, PSM, PV, and ORR), whereas the SB5 also detected growth to this age. Similarly, in FXS, the timing of growth aligned with the SB5 for DCCS USS, LSWM, PSM, and PV (significant growth at age 10 years, but not 16 or 22 years). Exceptions to this parallel were the continued significant growth of FICA and LSWM through age 22 years in DS, continued significant growth of FICA through age 16 years in FXS, and the significant 2-year decline in PV in FXS. Furthermore, group comparisons in growth on the SB5 mirrored the FICA Pro—both measures detected that DS had significantly more growth in early adulthood than the OID group. Regarding sex differences in starting levels, results largely fit with expected lower functioning in males with FXS compared with females with FXS.

The present results, using only 2 time points, already demonstrate important growth profiles through young adulthood. For example, DS displays a less steep yet prolonged growth estimate on several tests (FICA Pro, LSWM, and PV) compared with other groups. In fact, DS was the only group to have significant growth at 22 years; on LSWM (Cohen's $d = 0.36$) and FICA Pro ($d = 0.56$), young adults with DS show clear continued gains. This highlights the stage of natural cognitive maturation as a potential target for cognitive intervention in DS.

Although the number of time points per participant is small, group comparisons of trajectories suggest different developmental cognitive profiles not only across syndromes but also across cognitive domains, with comparatively slower growth in receptive vocabulary and oral reading and more rapid growth in processing speed in DS and comparatively slower growth in inhibitory control and attention, working memory, and oral reading in FXS. The relatively flat developmental profile of the DCCS prorated score in all groups, with no significant change observed over 2 years at any age, may reflect an especially detrimental impact of ID on the development of cognitive flexibility (as measured here) or possibly a lack of sensitivity to change of the DCCS test in this population. As mentioned earlier, DCCS involves the challenging process of learning multiple matching rules and when to switch between rules; this task may be too complex for some participants. Thus, it may not be that DCCS fails to detect cognitive flexibility but rather that the earliest aspects of cognitive flexibility are not tapped by the test or have not yet developed in some participants. Regarding repeated visits, it is also possible that different types of items administered across levels of difficulty, which can vary across test administrations

Figure 2 Mean Developmental Growth Estimates for NIHTB-CB USSs



Graphs represent approximate mean trajectories of developmental growth based on each test's latent change score model. The approximate mean change is plotted by using the group-specific intercept at age 10 years, adding the group-specific change score per 2 years (centered at 10, 16, and 22 years) and adding the regression outcome of age on change scores. Shaded bands represent 95% CIs for the change score around each plotted level (every 2 years). The uncorrected standard scores (USSs) provided by the NIHTB-CB have a mean of 100 and SD of 15 in the normative sample; the USSs are not age adjusted and thus allow for clear longitudinal comparison (i.e., without age adjustment).[19] NIHTB-CB = NIH Toolbox Cognition Battery.

within participants, could affect reliability and sensitivity of this test. As additional assessments are collected in these samples, more precise profiles of cognitive growth as measured by the NIHTB-CB will emerge.

These longitudinal data and change metrics may provide useful information with which to compare treatment study results. For example, treatment with a PDE4 allosteric inhibitor over a 24-week period in adult patients with FXS yielded significant gains of 2.8 and 5.8 USS points in ORR and PV, respectively, compared with placebo.[19] In this study's FXS sample, there were no significant gains on these tests during early adulthood, but at age 10 years, there was significant 2-year growth of 4.8 points on PV for males with FXS (and 6.5 points for females with FXS). Given the high test reliability and lack of any practice effects on these tests in FXS,[18] the magnitude of gains observed in the active treatment group can be better appreciated in the context of normative FXS development. The PV gains associated with PDE4 inhibition in adults with FXS are a striking contrast to the 2-year developmental decline in PV in adulthood shown in this study. In fact, the PV change with 24 weeks of PDE4 inhibition was greater than the amount of change shown in pre-adolescent children with FXS across 2 years of development in the present study.

There are several important limitations of this study. First, developmental changes in cognition reported here are based on 2 assessments per participant spaced approximately 2 years apart. Additional assessments will improve the accuracy of these profiles and growth estimates. Participants in the lower and upper ranges were more difficult to recruit, and there were more assessment challenges in the youngest participants with significant ID. Although the NIHTB-CB has lower age limits and better feasibility than many alternatives, assessments valid at the lowest mental ages remain a critical need in the context of ID and clinical trials. Test stage differences and scoring methods are another limitation. These may induce artificially large increases or decreases within participants over time (DCCS and FICA), potentially magnifying or obscuring the degree of change. If these tests are used in treatment studies, investigators need to be aware of these factors and consider using alternative scoring methods. Similarly, PSM uses different item sets (picture stories) according to age. Therefore, over time, an individual may receive different story sequences that could reduce reliability and precision in the measurement of change. An updated version of the NIHTB-CB is anticipated for 2023, which introduces several new tests, includes modifications to some tests (e.g., to reduce test staging effects or modified scoring procedures for DCCS and FICA), and reweights scores to the 2020 US Census. Of note, the current results may not generalize entirely to upcoming versions of the battery, and further validation may be needed. Regarding feasibility challenges for PCPS, Speeded Matching is a newly developed instrument with much better feasibility for ID, improved internal consistency, and a strong correlation with other processing speed measures.[32] Finally, groups were not ideally matched on several variables (e.g., ethnicity, race, and sex), and some group differences could be affected by these factors. For example, comparisons in growth could potentially reflect aspects of autism or differences in overall IQ rather than syndrome-specific factors. Recruitment of fully matched participants in these ID groups is challenging, but well-matched groups are essential for learning about syndrome-specific profiles and how they compare with the larger population of persons with ID.

Although the NIHTB-CB continues to show promise as a battery of cognitive outcomes, important knowledge gaps remain. Most notably, the degree to which growth in test performance represents, or perhaps predicts, changes in real-life daily functioning is not yet known. In ID, we might expect gains in some domains of adaptive behavior to follow from or track with expanded cognitive capacity. A related knowledge gap is: what constitutes a clinically meaningful change in NIHTB-CB scores? The meaningful change threshold refers to the level of difference in scores in a domain of interest which patients (or perhaps caregivers, in ID) perceive as beneficial. These are especially important considerations for understanding clinical response to treatment and drug approval decisions. Given that the NIHTB-CB is increasingly chosen as a primary or secondary outcome measure for clinical trials targeting cognition in ID conditions, the present results provide important new information about the battery's utility in the field and an initial description of the natural history of cognitive growth based on these tests in 2 common syndromes, FXS and DS.

Acknowledgment

The authors extend their appreciation to the families who gave their time and effort to participate in the study and to Leonard Abbeduto, LeAnn Baer, Ruth McClure Barnes, Kyle Bersted, Mikayla Brown, Ana Candelaria, Erin Carmody, Darian Crowley, Suzanne Delap, Andrea Drayton, Randi Hagerman, Anne Hoffmann, Londi Howard, Paige Landau, Caroline Leonczyk, Michael Nelson, Jacklyn Perales, Lacey Pomerantz, Shanelle Rodriguez, Melanie Rothfuss, Ryan Shickman, Andrea Schneider, Haleigh Scott, Laurel Snider, Rachel Teune, Denny Tran, Jamie Woods, and members of the MIND IDDRC Clinical Translational Core for their contributions to the study.

Study Funding

This study was funded by the NICHD (R01HD076189), the Health and Human Services Administration of Developmental Disabilities (90DD0596), and the MIND Institute Intellectual and Developmental Disabilities Research Center (P50HD103526).

Disclosure

R.H. Shields, A. Kaat, S.M. Sansone, C. Michalak, J. Coleman, T. Thompson, A. Dakopolos, F. McKenzie, and K. Riley: no relevant disclosures; E. Berry-Kravis has received funding from the following, all of which are directed to RUMC in support of rare disease programs, and she receives no personal funds and has no relevant financial interest in any of the commercial entities listed: Acadia, Alcobra, Anavex, Biogen, BioMarin, Cydan, Fulcrum, GeneTx, GW, Ionis, Lumos, Marinus, Neuren, Neurotrope, Novartis, Orphazyme, Ovid, Roche, Seaside Therapeutics, Tetra, Ultragenyx, Yamo, and Zynerba to consult on trial design and development strategies and/or to conduct clinical studies in FXS or other NNDs or neurodegenerative disorders; Vtesse/Sucampo/Mallinckrodt Pharmaceuticals to conduct clinical trials in NPC; and Asuragen Inc to develop testing standards for FMR1 testing; R. Gershon led the original development team for the NIH Toolbox and continues to oversee maintenance and distribution of the measures as a volunteer and/or as funded by the NIH. All license fees collected by Apple for use of the NIH Toolbox are used for support and maintenance and no other relevant disclosures; K.F. Widaman: no relevant disclosures; D. Hessl has received funding from the following, all of which are directed to the UC Davis, in support of fragile X treatment programs, and he receives no personal funds and has no relevant financial interest in any of the commercial entities listed: Autifony, Ovid, Tetra, Healx, and Zynerba pharmaceutical companies to consult on outcome measures and clinical trial design. Go to Neurology.org/N for full disclosures.

Publication History

Received by *Neurology* February 11, 2022. Accepted in final form September 21, 2022. Submitted and externally peer reviewed. The handling editor was Renee Shellhaas, MD, MS.

Appendix Authors

Name	Location	Contribution
Rebecca H. Shields, MS	University of California Davis Health, Sacramento, CA	Authored the manuscript, performed statistical analysis, and coordinated the study
Aaron J. Kaat, PhD	Northwestern University, Chicago, IL	Directed NIHTB-CB activities for the study, advised on the protocol and analysis, and authored portions of the manuscript
Stephanie M. Sansone, PhD	University of California Davis Health, Sacramento, CA	Assisted in developing the protocol and coordinated the study
Claire Michalak, BS	Rush University Medical Center, Chicago, IL	Coordinated the study at Rush University and critically reviewed the manuscript
Jeanine Coleman, PhD	University of Denver, Denver, CO	Coordinated the study at the University of Denver and critically reviewed the manuscript
Talia Thompson, PhD	University of Colorado, Aurora, CO	Conducted assessments and coordinated the study at the University of Denver and critically reviewed the manuscript
Forrest J. McKenzie, BS	University of California Davis Health, Sacramento, CA	Conducted assessments and coordinated the study at the University of California Davis
Andrew Dakopolos, PhD	University of California Davis Health, Sacramento, CA	Authored portions of the manuscript
Karen Riley, PhD	Regis University, Denver, CO	PI at the University of Denver and critically reviewed the manuscript
Elizabeth Berry-Kravis, MD, PhD	Rush University Medical Center, Chicago, IL	PI at Rush University and critically reviewed the manuscript
Richard C. Gershon, PhD	Northwestern University, Chicago, IL	PI and director of the NIH Toolbox and critically reviewed the manuscript
Keith F. Widaman, PhD	University of California Riverside, Riverside, CA	Supervised analyses and critically reviewed the manuscript
David Hessl, PhD	University of California Davis Health, Sacramento, CA	Designed the study, obtained funding, directed the multisite study, and authored portions of the manuscript

References

1. Maulik PK, Mascarenhas MN, Mathers CD, Dua T, Saxena S. Prevalence of intellectual disability: a meta-analysis of population-based studies. *Res Dev Disabilities*. 2011;32(2):419-436. doi:10.1016/j.ridd.2010.12.018
2. Olusanya BO, Wright SM, Nair MKC, et al. Global burden of childhood epilepsy, intellectual disability, and sensory impairments. *Pediatrics*. 2020;146(1):e20192623. doi:10.1542/peds.2019-2623
3. McGrath RJ, Stransky ML, Cooley WC, Moeschler JB. National profile of children with Down syndrome: disease burden, access to care, and family impact. *J Pediatr*. 2011;159(4):535-540.e2. doi:10.1016/j.jpeds.2011.04.019
4. López-Bastida J, Oliva-Moreno J. Cost of illness and economic evaluation in rare diseases *Adv Exp Med Biol*. 2010;686:273-282. doi:10.1007/978-90-481-9485-8_16
5. Hagerman RJ, Berry-Kravis E, Hazlett HC, et al. Fragile X syndrome. *Nat Rev Dis Primers*. 2017;3:17065. doi:10.1038/nrdp.2017.65
6. Esbensen AJ, Hooper SR, Fidler D, et al. Outcome measures for clinical trials in Down syndrome. *Am J Intellect Dev Disabil*. 2017;122(3):247-281. doi:10.1352/1944-7558-122.3.247
7. Antonarakis SE, Skotko BG, Rafii MS, et al. Down syndrome. *Nat Rev Dis Primers*. 2020;6(1):9.doi:10.1038/s41572-019-0143-7
8. Berry-Kravis EM, Lindemann L, Jønch AE, et al. Drug development for neurodevelopmental disorders: lessons learned from fragile X syndrome. *Nat Rev Drug Discov*. 2018;17(4):280-299. doi:10.1038/nrd.2017.221
9. De La Torre R, De Sola S, Hernandez G, et al. Safety and efficacy of cognitive training plus epigallocatechin-3-gallate in young adults with Down's syndrome (TESDAD): a double-blind, randomised, placebo-controlled, phase 2 trial. *Lancet Neurol*. 2016; 15(8):801-810. doi:10.1016/s1474-4422(16)30034-5
10. Boada R, Hutaff-Lee C, Schrader A, et al. Antagonism of NMDA receptors as a potential treatment for Down syndrome: a pilot randomized controlled trial. *Translational Psychiatry*. 2012;2(7):e141. doi:10.1038/tp.2012.66
11. Spiridigliozzi GA, Hart SJ, Heller JH, et al. Safety and efficacy of rivastigmine in children with Down syndrome: a double blind placebo controlled trial. *Am J Med Genet A*. 2016;170(6):1545-1555. doi:10.1002/ajmg.a.37650
12. Duchon A, Gruart A, Albac C, et al. Long-lasting correction of in vivo LTP and cognitive deficits of mice modelling Down syndrome with an α5-selective GABA. *Br J Pharmacol*. 2020;177(5):1106-1118. doi:10.1111/bph.14903
13. Erickson CA, Davenport MH, Schaefer TL, et al. Fragile X targeted pharmacotherapy: lessons learned and future directions *J Neurodevelopmental Disord*. 2017;9(1):7. doi:10.1186/s11689-017-9186-9
14. Budimirovic DB, Berry-Kravis E, Erickson CA, et al. Updated report on tools to measure outcomes of clinical trials in fragile X syndrome *J Neurodevelopmental Disord*. 2017;9(1):14. doi:10.1186/s11689-017-9193-x
15. Weintraub S, Bauer PJ, Zelazo PD, et al. I. NIH Toolbox Cognition Battery (CB): introduction and pediatric data. *Monogr Soc Res Child Dev*. 2013;78(4):1-15. doi:10.1111/mono.12031
16. Thompson T, Coleman JM, Riley K, et al. Standardized assessment accommodations for individuals with intellectual disability. *Contemp Sch Psychol*. 2018;22(4):443-457. doi:10.1007/s40688-018-0171-4
17. Hessl D, Sansone SM, Berry-Kravis E, et al. The NIH Toolbox Cognitive Battery for intellectual disabilities: three preliminary studies and future directions. *J Neurodevelopmental Disord*. 2016;8(1):35.doi:10.1186/s11689-016-9167-4
18. Shields RH, Kaat AJ, McKenzie FJ, et al. Validation of the NIH Toolbox cognitive battery in intellectual disability. *Neurology*. 2020;94(12):e1229-e1240. doi:10.1212/WNL.0000000000009131
19. Berry-Kravis EM, Harnett MD, Reines SA, et al. Inhibition of phosphodiesterase-4D in adults with fragile X syndrome: a randomized, placebo-controlled, phase 2 clinical trial. *Nat Med*. 2021;27(5):862-870. doi:10.1038/s41591-021-01321-w
20. Gershon R, Nowinski C, Peipert JD, et al. Use of the NIH Toolbox for assessment of mild cognitive impairment and Alzheimer's disease in general population, African-American and Spanish-speaking samples of older adults. *Alzheimer's Demen*. 2020; 16(S6).doi:10.1002/alz.043372
21. Association AP. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. Author; 2013.
22. Sparrow SS, Balla DA, Cicchetti DV. *Vineland Adaptive Behavior Scales, Third Edition*. AGS Publishing; 2016.
23. Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ. NIH toolbox for assessment of neurological and behavioral function. *Neurology*. 2013; 80(Issue 11, Supple):S2-S6. doi:10.1212/wnl.0b013e3182872e5f
24. McKenzie FJ, Drayton A, Shields RH, et al. *National Institutes of Health Toolbox Cognitive Battery Supplemental Administrator's Manual for Intellectual and Developmental Disabilities: A Guide on Administration and Scoring Standards*. UC Davis MIND Institute Translational Psychophysiology and Assessment Laboratory; 2019. nihtoolbox.force.com/s/article/nih-toolbox-cognitive-battery-supplemental-manual.
25. Sansone SM, Schneider A, Bickel E, Berry-Kravis E, Prescott C, Hessl D. Improving IQ measurement in intellectual disabilities using true deviation from population norms. *J Neurodevelopmental Disord*. 2014;6(1):16. doi:10.1186/1866-1955-6-16
26. Roid GH. *Stanford-Binet Intelligence Scales. 5 ed*. Riverside; 2003.
27. Team RCR: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2014.
28. Yves R. Lavaan: an R package for structural equation modeling. *J Stat Softw*. 2012; 48(2):1-36. doi:10.18637/jss.v048.i02
29. McArdle JJ. Latent variable modeling of differences and changes with longitudinal data. *Annu Rev Psychol*. 2009;60:577-605. doi:10.1146/annurev.psych.60.110707.163612
30. Ghisletta P, McArdle JJ. Latent curve models and latent change score models estimated in R. *Struct Equation Model A Multidisciplinary J*. 2012;19(4):651-682. doi:10.1080/10705511.2012.713275
31. Widaman KF. Best practices in quantitative methods for developmentalists: III. Missing data: what to do with or without them. *Monogr Soc Res Child Dev*. 2006;71(3):42-64.
32. Kaat AJ, McKenzie FJ, Shields RH, et al. Assessing processing speed among individuals with intellectual and developmental disabilities: a match-to-sample paradigm. *Child Neuropsychol*. 2021:1-13. doi:10.1080/09297049.2021.1938987