Symposium Title: Applications of the NIH Toolbox Cognitive Battery in Neurodevelopment, Neurodegeneration and Treatment

Chair: David Hessl$^{1,2}$

1: University of California, Davis MIND Institute, Sacramento, CA, USA
2: University of California, Davis Department of Psychiatry and Behavioral Sciences, Sacramento, CA, USA

Overview: Reliable and valid measurement of cognitive function in people with neurodevelopmental and neurodegenerative disorders associated with intellectual disability across laboratories and clinic settings is challenging due to inconsistencies in test selection, behavioral non-compliance, test standardization and scoring, and unknown psychometric properties in populations of interest. The three papers in this symposium focus on various applications of the National Institutes of Health Toolbox Cognitive Battery (NIHTB-CB), a tablet-administered set of tests measuring language, processing speed, memory, and executive function. The NIHTB-CB offers a standardized, touch-screen, web-based tool with potential for use in and comparison across various cognitive disorders, and potentially for measuring cognitive response to interventions. Paper 1 (Shields et al.) begins to tackle questions about the battery’s sensitivity to change by studying 2-year developmental growth in youth with intellectual disabilities, comparing specific subgroups (fragile X, Down syndrome, other ID) and comparing NIHTB-CB scores to a well-established IQ test. Paper 2 (Berry-Kravis et al.) examines cognitive profiles using the NIHTB-CB in comparison to Wechsler IQ scales in Niemann Pick Type C1 (NPC), a neurodegenerative cholesterol storage disease. The profiles, including relatively weak performance in processing speed and executive function, may help to guide interventions aiming to slow the progress of this condition. Finally, Paper 3 (Biag et al.) represents the first steps in utilizing the NIHTB-CB as a key outcome measure in a double-blind placebo-controlled trial (metformin; fragile X syndrome; Azrieli Foundation). Although results on the efficacy of metformin on outcomes from the trial are not yet available, the baseline data reported can provide a first look at how the battery performs within a multi-center trial. In summary, the symposium will provide a brief introduction to the NIHTB-CB, including its psychometrics and recent validation in ID, and it will provide 3 specific example applications of the battery within neurodevelopmental, neurodegenerative and treatment contexts.

Paper 1 of 3

Paper Title: NIH Toolbox Cognitive Battery: Detecting Developmental Change in Domains of Cognition in Down Syndrome, Fragile X Syndrome, and Other Intellectual Disabilities

Authors: Rebecca Shields$^{1,2}$, Aaron Kaat$^3$, Elizabeth Berry-Kravis$^4$, Karen Riley$^5$, Keith Widaman$^6$, Richard Gershon$^3$, David Hessl$^{1,2}$

1: University of California, Davis MIND Institute, Sacramento, CA, USA
2: University of California, Davis Department of Psychiatry and Behavioral Sciences, Sacramento, CA, USA
3: Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
4: Rush University Medical Center, Chicago, IL, USA
5: University of Denver, Denver, CO, USA
6: University of California, Riverside, Riverside, CA, USA

Introduction: The NIH Toolbox Cognitive Battery (NIHTB-CB) is an iPad-based cognitive assessment that has been evaluated for use in individuals with neurodevelopmental disorders, including fragile X syndrome (FXS), Down syndrome (DS), and other intellectual disability (OID) (Hessl et al., 2016). The NIHTB-CB has been shown to be valid, reliable, and feasible in these individuals, with results varying somewhat by test, and with similar results to the validation study in a normative childhood sample (Zelazo et al, 2013). An important remaining question is whether the NIHTB-CB detects meaningful changes in cognition. In order to be used in treatment trials, educational assessments, or other clinical or research settings, it is necessary that the
battery is sensitive to detecting change in an individual’s performance. In fact, a lack of appropriate, psychometrically sound and objective endpoints has been cited as a potential reason for the failure of many targeted treatment trials in ID-associated disorders such as FXS (Erickson et al, 2017). To this end, we assessed participants ages 6 to 25 years at two time points, with 2 years between visits, with the goal of using the NIHTB-CB to measure developmental change in cognition. In addition, the Stanford Binet 5 (SB-5) Change Sensitive Scores (CSS) were assessed at each time point as a benchmark of cognitive change.

**Methods:** Across three sites (University of Denver, Rush University Medical Center, and UC Davis MIND Institute), a total of 95 participants were seen at both time points: 29 with FXS, 41 with DS, and 25 with OID. The interval between time points was a mean of 2.12 years. Participants completed the SB-5 and NIHTB-CB at each time point. Using a latent difference score model, we assessed magnitude and significance of change of each NIHTB-CB test as well as the SB-5 in each diagnostic group, using chronological age and group as covariates. Uncorrected standard scores were used for NIHTB-CB tests, and full scale change sensitive scores (CSS) were used for the SB-5.

**Results:** We examined change at two age points, 10 years and 20 years. On the SB-5, in OID at the 10-year-old reference point, the OID group had an increase of 5.29 (SE = 1.09) points (d = 0.40) from Time 1 to Time 2. FXS showed marginally less growth than OID, (2.64 points, SE = 1.09, d = .20). The DS group showed a significant increase of 5.31 points, d = .40.

Six of the seven NIHTB-CB tests (all except Pattern Comparison) detected significant cognitive growth at age 10 in the full sample (ranging from 2.92 – 16.88 standard score points, d = .15 - .83). The executive function tests (Flanker, DCCS, and List Sorting) showed the largest improvements at this age range. All effect sizes of change in NIHTB-CB tests except Pattern Comparison were similar to or larger than the SB-5 CSS effect sizes.

Comparing diagnostic groups, some differential rates of growth were present. At age 10, FXS had significantly less growth in Oral Reading than DS (p = .02). At age 20, rates of change were much smaller and non-significant; however, OID and DS continued to significantly improve on FICA at this age (though FXS did not), and DS continued to show overall cognitive growth on SB-5 CSS. Additional participants will be added to these samples for analysis.

**Discussion:** To be able to accurately show improvement (or lack of improvement) in a treatment trial, it is necessary to have accurate and sensitive measures. These results build on the validation studies of the NIHTB-CB in intellectual disabilities, showing the sensitivity of the battery to detect changes occurring in development/over time. Importantly, the NIHTB-CB and the SB-5 show similar patterns of results, with greater change happening at age 10 and less change happening at age 20, although in general the NIHTB-CB showed larger magnitude of change. The longer/later growth of DS only is an interesting finding that justifies further research. It is possible that individuals with DS have a more protracted cognitive growth compared to other IDs, which would have impacts regarding treatments of young adults with IDs, to maximize development during this time. Our results suggest that the NIHTB-CB has potential as a measure of change in cognitive skills, and they support the use of the NIHTB-CB tests as a measurement of outcomes. Further research with larger sample sizes will add to these findings, and updated results will be presented. Use of the NIHTB-CB within controlled trials is ongoing, and this may yield additional information about its sensitivity to treatment-related change.

**References/Citations:**


**Paper 2 of 3**

**Paper Title:** Characterization of the Cognitive Profile of Niemann-Pick type C1 Using the NIH Toolbox Cognitive Battery

**Authors:** Elizabeth Berry-Kravis,1 Rebecca Jaeger,1 Audrey Thurm2

1: Rush University Medical Center, Chicago, IL, USA
2: National Institutes of Health, Bethesda, MD, USA

**Introduction:** Niemann-Pick type C1 (NPC) is a neurodegenerative storage disease characterized by cholesterol accumulation in lysosomes and late endosomes. The disease onset is insidious, with decline in the rate of development prior to clear degeneration that often leads to a profile consistent with that of an intellectual disability diagnosis. Cognitive decline appears to be uneven, with non-verbal cognition being impacted more substantially early in disease than verbal cognition on the Wechsler scales (Thurm et al. 2016), although this pattern may be impacted by motor dysfunction such as tremor or dystonia. With limited extended neuropsychological profiling of youth with NPC reported in the literature, we sought to determine the cognitive profile of NPC using the NIH Toolbox Cognitive Battery (NIHTB-CB), which may be associated with less motor demand than traditional cognitive tests. Identification of a more comprehensive cognitive profile may allow insight into the most impacted and presumably earliest cognitive signs of disease, which will be important for diagnosis, decisions about when to initiate new disease-modifying treatments currently in clinical trials, and for tracking response to treatment.

**Methods:** Participants included patients with NPC1 tested at baseline with the NIHTB-CB when entering a treatment trial or a natural history study. Testing was not attempted with individuals under age 4, or those functioning with a mental age of less than 4. Patients for whom testing was attempted but practice testing was unsuccessful and the patient could not get a score on half or more of the subtests were excluded. Twenty-one patients (age 16.4 ± 8.7, range 4-31) were able to complete testing. The NIHTB-CB subtests administered were Picture Vocabulary (PV), Oral Reading Recognition (ORR), List Sorting Working Memory (LSWM), Pattern Comparison Processing Speed (PCPS), Picture Sequence Memory (PSM), Flanker Inhibitory Control and Attention (FIC), and Dimensional Change Card Sort (DCCS). Composite scores for Fluid Cognition, Crystallized Cognition and Total Cognition were generated from the NIHTB-CB program. Age-adjusted standard scores were used to evaluate performance on each subtest of NIHTB-CB relative to the norm of 100 ±15 for each test. Test performance on subtests was compared using paired student’s t-test.

**Results:** All 21 patients were able to get scores on ORR, PCPS, FIC, and DCCS. One patient failed to get a score on PV, 2 on PSM, and 5 on LSWM, suggesting the working memory task was most difficult to pass the practice items for this group of patients. Mean NIHT-CB subtest standard scores for the group were as follows: PV 86.5±15.4, ORR 90.2±15.6, LSWM 76.6±16.7, PCPS 55.6±24.5, PSM 82.4±17.2, FIC 72±24.5, DCCS 72.5±21.2. Performance on PCPS was significantly worse than on all other tests (p ranging from 0.002 (FIC) to 0.0000005 (ORR)). Performance on ORR was numerically better than all other tests and significantly better than PCPS, FIC, DCCS, and LSWM (p ranging 0.017 (LSWM) to 0.0000005 (PCPS)). Performance on PV was significantly better than PCPS (p=0.00001), FIC and DCCS (both p=0.017). PSM was significantly different only from PCPS (better) and LSWM was significantly different only from ORR (worse) and PCPS (better). The composite score for crystallized cognition was much better than the fluid cognition composite (87.4±14.6 vs 57.2±23.7, p=0.00005). The results were nearly identical when data analysis was restricted to only the 16 patients who had a score on every subtest. There were strong correlations NIHTB-CB scores and comparative constructs from the WISC/WAIS for total cognition/FSIQ (r=0.79, p=0.0008), PV/VIQ (r=0.73, p=0.0003), crystallized cognition/VIQ (r=0.76, p=0.0001), PCPS/Processing Speed (r=0.80, p=0.00001), and a weaker correlation between LSWM/working memory (r=0.57, p=0.03). Mean standardized performance on NIHTB-CB Total Cognition was 67.2±21.4 and FSIQ...
67.5±16.5, while the Verbal Comprehension Index on the WISC/WAIS at 79.1±18.5 was somewhat lower than NIHTB-CB
Crystallized Cognition, the Working Memory Index at 70.5±13.8 was somewhat lower than LSWM, and the Processing Speed
Index at 65.4±15.5 was somewhat higher than PCPS.

Discussion: These preliminary results suggest there is a specific pattern of performance on NIHTB-CB subtests in patients with
early mild-moderate symptoms of NPC1, with clear areas of strength in reading and vocabulary and weakness in processing
speed, inhibitory control and flexibility. This pattern suggests that areas of weakness on NIHTB-CB could be followed in pre-
symptomatic patients to screen for evidence of very early clinical symptom onset, to guide initiation of treatments. Overall in this
population NIHTB-CB displays good construct validity. Although there was generally good correspondence, reasons for mild
disparities in mean standard scores between NIHTB-CB and the Wechsler tests are not clear but could have to do with difference
in the specific motor demands of the test, age issues or other factors. Additional work on an expanded NPC cohort is needed to
evaluate impact of motor skills on relative performance on NIHTB-CB versus standard tests, and to look at longitudinal
performance on the NIH-TCB.

References/Citations:

• Thurm, A., Farmer, C., Farhat, N.Y., et al. (2016) Cohort study of neurocognitive functioning and adaptive behavior in
children and adolescents with Niemann–Pick Disease type C1. Developmental Medicine and Child Neurology, 58, 262-
269.

Paper 3 of 3

Paper Title: Baseline Toolbox Data from a Multicenter Double-Blind, Placebo-Controlled Trial of Metformin in Individuals with
Fragile X Syndrome

Authors: Hazel Maridith B. Biag, MD1,2; Tracy Pham3; Rebecca Shields, MS1,4; Andrea Schneider, PhD1,2; David Hessl, PhD1,4; Randi
J. Hagerman, MD1,2
1: MIND Institute, UC Davis Medical Center, Sacramento, CA
2: Department of Pediatrics, UC Davis Medical Center, Sacramento, CA
3: Stollery Children’s Hospital, University of Alberta, Edmonton, AB
4: Department of Psychiatry and Behavioral Sciences, UC Davis School of Medicine, Sacramento, CA

Introduction: Fragile X syndrome (FXS) is the most common inherited form of intellectual disability (ID) and is typically diagnosed
between 2 and 3 years of age at the time language delays and behavioral manifestations emerge (Bailey et al., 2011). Studies
utilizing the Drosophila model and knock out mouse model of fragile X syndrome (FXS) have found metformin, a drug commonly
used in individuals with type 2 diabetes, obesity, and impaired glucose tolerance, to rescue memory, social novelty deficits, and
neuroanatomical abnormalities (Monyak et. al, 2017). These studies provided preliminary evidence that metformin could be
used as a targeted treatment for the cognitive and behavioral problems associated with FXS. A multicenter randomized, double-
blind, placebo-controlled trial of metformin in children, adolescents, and young adults with FXS is ongoing and the baseline data
for the NIH Toolbox Cognitive Battery (NIHTB-CB) assessment will be presented here.

Methods: Participants include 22 males and 1 female with a molecular genetic confirmation of full FMR1 mutation or mosaicism
fragile X syndrome. Inclusion and exclusion criteria along with other measurements utilized in the trial will be discussed at the
presentation. Cognition is 1 of 4 domains measured by the NIH Toolbox for the Assessment of Neurological and Behavioral
Function (NIHTB-CB) and can be used in those as young as 3 years old (Weintraub et al., 2013). This trial utilizes 8 measures
within this domain to assess different areas of cognition. The Flanker Inhibitory control and Attention tests measures inhibition
and visual attention, Dimensional Change Card Sort Test (DCCS) measures cognitive flexibility, List Sorting Working Memory Test
(LS) requires immediate recall and sequencing of different visually and orally presented stimuli, Pattern Comparison Processing
Speed Test (PC) measures speed of processing, Picture Vocabulary Test (PVT) measures receptive vocabulary, Oral Reading Recognition Test (OR) (Zelazo et al., 2013), Picture Sequence Memory (PSM) measures episodic memory (Bauer et al., 2013) and a newly added measure Speeded Matching (SM) also measures cognitive flexibility. Mental age was calculated via Leiter-3 non-verbal (NV) raw scores.

**Results:** Thus far, 33 participants aged 6 to 25 have been screened, 9 failed to meet inclusion or exclusion criteria, 3 were unable to do any portion of testing due to behavioral noncompliance, and 2 were below the mental age of 3 but testing was attempted regardless. Feasibility, as measured by the percent of participants who were able to pass practice and received a valid score, varied depending on the subtest. Preliminary baseline results show that the Flanker had a feasibility of 80.0%, PVT 100%, DCCS 90%, PSM 90%, LS 45%, SM 92.9%, PC 70%, and OR 95%. In regard to the sum of Leiter-3 raw scores, it was shown that the PVT (theta score) is correlated moderately with Leiter NV total raw score (n=20, r=.70, p<.001) and the Leiter-3 Figure Ground (FG) raw score was moderately correlated with Flanker computed score (n=15, r=.46, p=.08).

**Discussion:** Preliminary analyses from this trial suggests that the NIHTB-CB has relatively good feasibility for data acquisition in those aged 6-25 with FXS. The measure with the lowest feasibility was LS, suggesting that those in this population have more difficulty with immediate recall and sequencing of different stimuli presented visually and orally. The measure with the highest feasibility was PVT. Literature has suggested that receptive vocabulary strongly correlates with “g”/general intelligence; the moderate correlation between the Leiter NV scores and the PVT (theta score) suggests that this measure aligns well with the overall cognitive functioning level in this population and supports its usefulness for measuring an important cognitive domain via a brief and highly feasible iPad test. Though the moderate correlation between the FG raw score (a task related to visual attention) and the Flanker computed score approaches statistical significance, the calculations are limited by a smaller sample size; statistical significance is expected to strengthen as more participant scores are analyzed. Analyses are ongoing and we anticipate presenting on 40 participants. We will also be presenting further on correlative data between Leiter-3, Aberrant Behavior Checklist-Community (ABC-C), and SNAP-IV scores. Updated results will be given at the conference.

**References/Citations:**