Test-Retest Reliability and Practice Effects on the NIH Toolbox Cognition Battery in Duchenne Muscular Dystrophy

Submitter Aaron Kaat

Affiliation Feinberg School of Medicine, Northwestern University, Chicago

SUBMISSION DETAILS

I agree to provide poster pdf for attendee download. Yes

Poster PDF for download <blank>

Methodological Issue Being Addressed Does the NIH Toolbox Cognition Battery have appropriate evidence for use as a performance-based clinical outcome assessment in Duchenne Muscular Dystrophy (DMD)?

Introduction The NIH Toolbox Cognition Battery (NIHTB-CB) is a low-burden assessment battery measuring both fluid and crystalized cognitive skills with strong validity evidence in the general population. It is composed of three composites (Total, Fluid, and Crystalized Cognition) and seven tests (Fluid: Dimensional Change Card Sorting, Flanker Inhibitory Control, Picture Sequence Memory, List Sorting Working Memory, Pattern Comparison Processing Speed; Crystallized: Picture Vocabulary and Oral Reading and Recognition). The purpose of this study was to evaluate the reliability of the NIHTB-CB in a sample of boys with DMD—a rare genetic disorder characterized by progressive skeletal and cardiac muscle weakness and cognitive impairments, notably in executive functioning areas (Thangarajh et al., 2019).

Methods The NIHTB-CB was administered to 29 participants with DMD between 4.5 and 27 years old. The sample was recruited as and is a subsample of a larger psychometric readiness study. The NIHTB-CB was administered at baseline and after 2 to 6 weeks. Feasibility was indexed as the proportion of the sample capable of completing the measure on both assessment occasions. Then, we used a mixed effect model with a random intercept for participant to estimate test-retest reliability and to estimate practice effects (on the raw test scale and as a standardized mean difference [SMD]). Test-retest reliability was calculated both for the raw score (or unadjusted standard score for the composites) and for the age-adjusted standard score.

Results Feasibility of the individual tests was high (minimum 83%), but many individuals failed to complete at least one measure on at least one occasion, reducing feasibility of the composites (51% for Fluid and Total Composites; 55% for the Crystallized Composite). Reliability was appropriate (all raw score ICCs > 0.70). In general, the raw scores exhibited higher reliability than the age-adjusted scores (median ICC=0.85 vs 0.68); and the reliability of the three composites was also higher than the reliability of the individual measures (median 0.89 vs 0.75 for raw scores), which is as expected given that more information generally improves reliability evidence though these results remains limited due to the poor feasibility of the composites. On the individual tests, practice effects were negligible-to-small and nonsignificant in all cases. The median standardized mean difference between assessment occasions was SMD=0.07 (maximum SMD=0.23). Full results

are provided in the attached Table 1.

Conclusion DMD clinical trials predominantly have focused on motor outcomes, but cognitive outcomes are increasingly being recognized as an important secondary treatment target. Subtle executive function deficits in particular are relevant. But in order for future clinical trials to evaluate cognitive performance, reliable and fit-for-purpose measures are necessary. This study provides some of the evidence needed to support that assessment—at least for individual tests—and should encourage ongoing evaluation of the fitness of the NIHTB in DMD. The executive function measures are likely most relevant for future clinical trials; they had some of the best feasibility with good reliability and near-zero practice effects. Across the battery, test-retest reliability was high for the raw scores, and there were minimal practice effects. Norm-referenced scores exhibited slightly lower reliability, but these scores are less applicable for the context of use of outcomes measurement (Farmer et al., 2023). Feasibility of some tests were lower, but this was inconsistent within or across participants, such that only the composites were significantly affected by test feasibility. This suggests that the individual measures may also be more appropriate for the clinical trial context than composites, which is also consistent with the DMD profile of broadly low average overall cognitive performance but more subtle specific deficits in executive function subdomains.

Citations:
Farmer, C., Thurm, A., Troy, J.D., and Kaat A.J. (2023), Comparing ability and norm-referenced scores as clinical trial outcomes for neurodevelopmental disabilities: a simulation study. J Neurodevelop Disord 15: 4. https://doi.org/10.1186/s11689-022-09474-6
Thangarajh, M., Kaat, A.J., Bibat, G., Mansour, J., Summerton, K., Gioia, A., Berger, C., Hardy, K.K., and Wagner, K.R. (2019), The NIH Toolbox for cognitive surveillance in Duchenne muscular dystrophy. Ann Clin Transl Neurol, 6: 1696-1706. https://doi.org/10.1002/acn3.50867

Co-Authors

* Presenting Author

First Name	Last Name	Affiliation
Aaron *	Kaat *	Feinberg School of Medicine, Northwestern University, Chicago
Mathula	Thangarajh	Virginia Commonwealth University

Keywords

Keywords
Duchenne Muscular Dystrophy
NIH Toolbox
Cognition
Reliability

Guidelines I have read and understand the Poster Guidelines

Disclosures The authors report no conflicts of interest for this work. Research reported in it was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number R21TR004007. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Related Tables and Supporting Materials Table 1 - revised.docx