




BRIEF COMMUNICATION

Assessment of Prorated Scoring of an Abbreviated Protocol for the National Institutes of Health Toolbox Cognition Battery

Alexander D. Rebchuk¹ , Arshia Alimohammadi², Michelle Yuan^{3,4}, Molly Cairncross^{5,6}, Ivan J. Torres^{7,8} , Noah D. Silverberg^{5,6} and Thalia S. Field^{2,4,9,*} 

¹Division of Neurosurgery, University of British Columbia, Vancouver, Canada

²Faculty of Medicine, University of British Columbia, Vancouver, Canada

³Undergraduate Student, Faculty of Science, University of British Columbia, Vancouver, Canada

⁴Vancouver Stroke Program, Vancouver, BC, Canada

⁵Division of Physical Medicine & Rehabilitation, University of British Columbia, Vancouver, Canada

⁶Rehabilitation Research Program, Vancouver Coastal Health Research Institute, Vancouver, Canada

⁷Department of Psychiatry, University of British Columbia, Vancouver, Canada

⁸British Columbia Mental Health and Substance Use Services, Vancouver, Canada

⁹Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, Canada

(RECEIVED May 12, 2020; FINAL REVISION August 2, 2020; ACCEPTED September 3, 2020)

Abstract

Objective: To evaluate an abbreviated NIH Toolbox Cognition Battery (NIHTB-CB) protocol that can be administered remotely without any in-person assessments, and explore the agreement between prorated scores from the abbreviated protocol and standard scores from the full protocol. **Methods:** Participant-level age-corrected NIHTB-CB data were extracted from six studies in individuals with a history of stroke, mild traumatic brain injury (mTBI), treatment-resistant psychosis, and healthy controls, with testing administered under standard conditions. Prorated fluid and total cognition scores were estimated using regression equations that excluded the three fluid cognition NIHTB-CB instruments which cannot be administered remotely. Paired *t* tests and intraclass correlations (ICCs) were used to compare the standard and prorated scores. **Results:** Data were available for 245 participants. For fluid cognition, overall prorated scores were higher than standard scores (mean difference = +4.5, *SD* = 14.3; *p* < 0.001; ICC = 0.86). For total cognition, overall prorated scores were higher than standard scores (mean difference = +2.7, *SD* = 8.3; *p* < 0.001; ICC = 0.88). These differences were significant in the stroke and mTBI groups, but not in the healthy control or psychosis groups.

Conclusions: Prorated scores from an abbreviated NIHTB-CB protocol are not a valid replacement for the scores from the standard protocol. Alternative approaches to administering the full protocol, or corrections to scoring of the abbreviated protocol, require further study and validation.

Keywords: Neuropsychology, Validation testing, Cognition, Telemedicine, Rehabilitation, Stroke, Mild traumatic brain injury, Psychosis

INTRODUCTION

Cognition is an important outcome in research trials and clinical practice (McInnes et al., 2017; Sheffield et al., 2018; Tang et al., 2018). To provide a common metric of cognition in the context of clinical research, the NIH Toolbox Cognition Battery (NIHTB-CB) was introduced. It is a brief, tablet-based cognitive assessment that has been validated for use

in healthy populations and those with neurological and psychiatric disease (Carlozzi, Goodnight et al., 2017; Carlozzi, Tulskey et al., 2017; Weintraub et al., 2014).

The NIHTB-CB is comprised of seven instruments: two assessing crystallized cognition (Picture Vocabulary and Oral Reading Recognition) and five assessing fluid cognition (Flanker Inhibitory Control and Attention, List Sorting Working Memory, Dimensional Change Card Sort, Pattern Comparison Processing Speed, and Picture Sequence Memory) (Weintraub et al., 2013). Of these, both instruments that assess crystallized cognition and two that assess fluid cognition (List Sorting Working Memory and Picture Sequence Memory)

*Correspondence and reprint requests to: Thalia Field, MD FRCP C MHS, Associate Professor, University of British Columbia, Stroke Neurologist, Vancouver Stroke Program, S169-2211 Wesbrook Mall, Vancouver BC V6T 2B5. E-mail: thalia.field@ubc.ca

can be modified for administration without any physical contact between examinee and tablet. The other three fluid cognition instruments are scored based on accuracy and reaction time, and thus, require in-person inputs into the tablet.

In the context of the COVID-19 pandemic where strict physical distancing guidelines have been implemented, there is a strong need for remote cognitive assessments (Gostin & Wiley, 2020). Our group has previously developed and validated a protocol for administering the NIHTB-CB using telemedicine to assess participants at remote sites (Rebchuk et al., 2019). However, this protocol still requires in-person conditions for some instruments. Recent guidelines published by the NIHTB-CB developers describe an abbreviated protocol, incorporating only the four instruments that can be administered entirely remotely (*HealthMeasures Help Desk*, 2020a).

We sought to explore whether a prorated score based on this abbreviated battery could provide a valid substitute for the standard score from the full battery. We assessed the agreement between prorated fluid and total cognition scores from the abbreviated protocol versus standard scores from the full protocol. The equations we applied to estimate prorated scores were derived from published regression equations for NIHTB-CB standard scores (Casaletto et al., 2015; *HealthMeasures Help Desk*, 2020b). As much ongoing research has been modified to facilitate physical distancing, this work helps to inform the future interpretation of data collected with the abbreviated NIHTB-CB protocol.

METHODS

Data

We extracted participant-level NIHTB-CB data gathered under standard conditions by trained examiners as part of six previous or ongoing studies in individuals with neurological disease [history of stroke or mild traumatic brain injury (mTBI)] or psychosis (inpatients with treatment-resistant psychosis) and healthy controls (no history of neurological disease, learning disability, or active psychosis). See Supplementary material for details of respective studies.

For all data sets, the NIHTB-CB was administered on an iPad (Apple, California, USA), and Form A of the cognition battery was used. Participant demographic data were captured with written questionnaires. All participants were older than 18 years and provided written informed consent. The experimental protocols for the respective studies were approved previously by the University of British Columbia's Clinical Research Ethics Board, and conformed to the Declaration of Helsinki.

Statistical Analysis

We chose to report standard scores corrected for age (mean = 100, standard deviations = 15) and not other demographic variables because education levels may not be equivalent across regions where our data were collected (Vancouver, Canada) and where the NIHTB-CB was normed

(United States) (Chevalier et al., 2016). As well, several of our participants identified with race(s) that the NIHTB-CB race/ethnicity options failed to capture.

Prorated fluid (Equation 1) and total (Equation 2) cognition scores were derived from appropriate regression equations provided with the NIHTB-CB (Casaletto et al., 2015; *HealthMeasures Help Desk*, 2020b). The prorated fluid cognition score included instruments (List Sorting Working Memory and Picture Sequence Memory) that can be administered remotely without the examinee having direct access to the tablet.

- (1) Prorated Fluid Cognition = $100 + 15 * [((\text{Mean of List Sort \& Pic Seq Mem Age-corrected Scores}) - 100.15)/10.10]$
- (2) Prorated Total Cognition = $100 + 15 * [((\text{Mean of Age-corrected Prorated Fluid Composite Score \& Crystallized Composite Scores}) - 100.02)/12.93]$

Data were separated into healthy controls and disease-specific groups (stroke, mTBI, and psychosis). Demographic data between groups were compared using one-way analysis of variance for parametric data and chi-square test for categorical data.

Paired *t* tests were used to compare the standard and prorated fluid cognition score within each group; data met assumptions of normality (Meyers et al., 2013). Prediction error was determined for the difference between standard and prorated scores for each participant, as well as mean prediction error for each group. Intraclass correlation (ICC) values between standard and prorated fluid cognition group-level scores were generated using two-way mixed effects, absolute agreement, and multiple measurements model (Koo & Li, 2016). Data met assumptions of normality and equality of variance for ICC analyses. All analyses were repeated for the prorated total cognition scores. We operationalized a clinically meaningful discrepancy as 0.5 standard deviations (or 7.5 standard score points), and calculated the frequency of participants with prorated–standard discrepancies exceeding this magnitude (Silverberg & Millis, 2009). A prediction error of zero reflects equal standard and prorated scores. Chi-square tests were used to compare observed frequencies of participants with clinically significant prediction errors (i.e., exceeding ± 0.5 SD difference between total and prorated score) between groups. Data met the assumptions of chi-square testing.

Given the exploratory nature of the study, we did not correct for multiple comparisons. Significance was set a priori at 0.05. Statistical analyses were performed using IBM SPSS Statistics (Version 19.0; IBM Corp., Armonk, NY).

RESULTS

Data were available for 245 participants: 77 (31.4%) healthy controls, 66 (26.7%) individuals with mTBI, 63 (25.7%) with a history of stroke, and 39 (15.9%) with active psychosis. Almost half (48.6%) were female, mean age was 41.8 years ($SD = 11.9$), and mean duration of education was 14.9 years ($SD = 2.6$). Subgroup characteristics are shown in Table A1 in the Appendix.

Table 1. Standard and prorated age-corrected standard scores (mean, *SD*) for fluid cognition and total cognition in healthy participants and those with stroke, psychosis, and mTBI. ICCs (95% CIs) between standard and prorated scores are given

	Fluid cognition			
	Standard score	Prorated score	<i>p</i>	ICC (95% CIs)
All subjects	95.4 (19.8)	99.8 (21.8)	<0.001	0.86 (0.80, 0.90)
Healthy control	109.3 (15.0)	112.1 (17.7)	0.109	0.73 (0.58, 0.83)
Stroke	93.0 (17.9)	99.0 (22.0)	0.002	0.82 (0.68, 0.90)
Psychosis	76.1 (19.6)	75.9 (15.9)	0.939	0.86 (0.74, 0.93)
mTBI	93 (14.7)	100.6 (16.9)	<0.001	0.74 (0.46, 0.86)
	Total cognition			
	Standard score	Prorated score	<i>p</i>	ICC (95% CIs)
All subjects	103.4 (16.9)	106.2 (18.1)	0.001	0.88 (0.83, 0.91)
Healthy control	115.7 (11.1)	117.5 (13.3)	0.08	0.86 (0.77, 0.91)
Stroke	100.4 (13.7)	104.1 (16.0)	0.001	0.89 (0.80, 0.94)
Psychosis	84.2 (17.3)	84.3 (15.3)	0.915	0.95 (0.90, 0.97)
mTBI	103.3 (13.0)	107.8 (14)	<0.001	0.89 (0.73, 0.94)

Group-level Comparisons

Overall, fluid cognition prorated scores were higher than standard fluid cognition scores (mean difference +4.5, *SD* = 14.3; $p < 0.001$). These differences were significant in the stroke and mTBI groups, but not in the healthy or psychosis groups. This resulted in overall prorated scores for total cognition also being higher than standard total cognition scores (mean difference +2.7, *SD* = 8.3; $p < 0.001$). Again, these differences were only significant in the stroke and mTBI groups (see Table 1). Overall agreement between prorated and total scores as per the ICC was moderate-to-good for fluid cognition only, and good-to-excellent for total cognition.

Individual-level Comparisons

Clinically significant fluid cognition prediction errors (greater than ± 0.5 *SD* difference between total and prorated scores) were present in 62.9% of participants; 42.9% were overestimated and 20.0% were underestimated. For total cognition, 40.4% of participants had a prediction error; 28.6% were overestimated and 11.8% were underestimated (Figure 1). The psychosis group had the lowest percentage (59.0%) of fluid prediction errors, followed by healthy control (60.0%), mTBI (65.2%), and stroke (69.8%) groups. These numerical differences did not meet statistical significance ($p = 0.425$). For total cognition, healthy controls had the lowest percentage (33.8%) of prediction errors greater than ± 0.5 *SD*, followed by those with psychosis (35.9%), mTBI (42.4%), and stroke (49.2%). Again, these differences were not statistically significant ($p = 0.275$).

DISCUSSION

The aim of this exploratory study was to assess the validity of a prorated score, based on a proposed abbreviated NIHTB-CB

protocol, against the standard score for the usual protocol (*HealthMeasures Help Desk*, 2020a). Particularly during COVID-19-related physical distancing measures, the potential advantage of an abbreviated protocol is its ability for remote administration without personnel alongside the examinee. Beyond the COVID-19 pandemic, advantages of a fully remote protocol could include greater participation by those with mobility restrictions or in isolated communities, and fewer losses to follow-up (Berge et al., 2016).

Overall, we found that prorated scoring for the abbreviated protocol overestimated fluid and total cognition standard scores. However, differences were noted between testing groups, with no group-level differences seen between prorated and standard scores in healthy individuals or in those with treatment-resistant psychosis.

It is uncertain as to whether these significant differences in group-level performance represent true differences related to domain-specific deficits from lesional injuries in the stroke or mTBI participant groups, random error, or insufficient statistical power to detect between-group differences in the healthy control group or, in particular, the psychosis group, which has the fewest participants (McInnes et al., 2017; Nys et al., 2007; O'Brien et al., 2003). The instruments included within our prorated scores include measures of working memory and episodic memory, and fail to capture processing speed, attention, and executive function (Mungas et al., 2014). It may be that anatomic lesions or functional deficits (e.g., frontal lobe injury, motor deficits, and fatigue) in the stroke and mTBI cohorts result in worse performance in executive function and timed tasks, in particular, and hence lead to the overestimation of prorated scores with exclusion of instruments assessing these specific domains. The data were collected as part of six separate studies, and unmeasured confounders specific to study conditions may also play a role.

Although exclusion of processing speed, attention, and executive function tests from prorated scores failed to

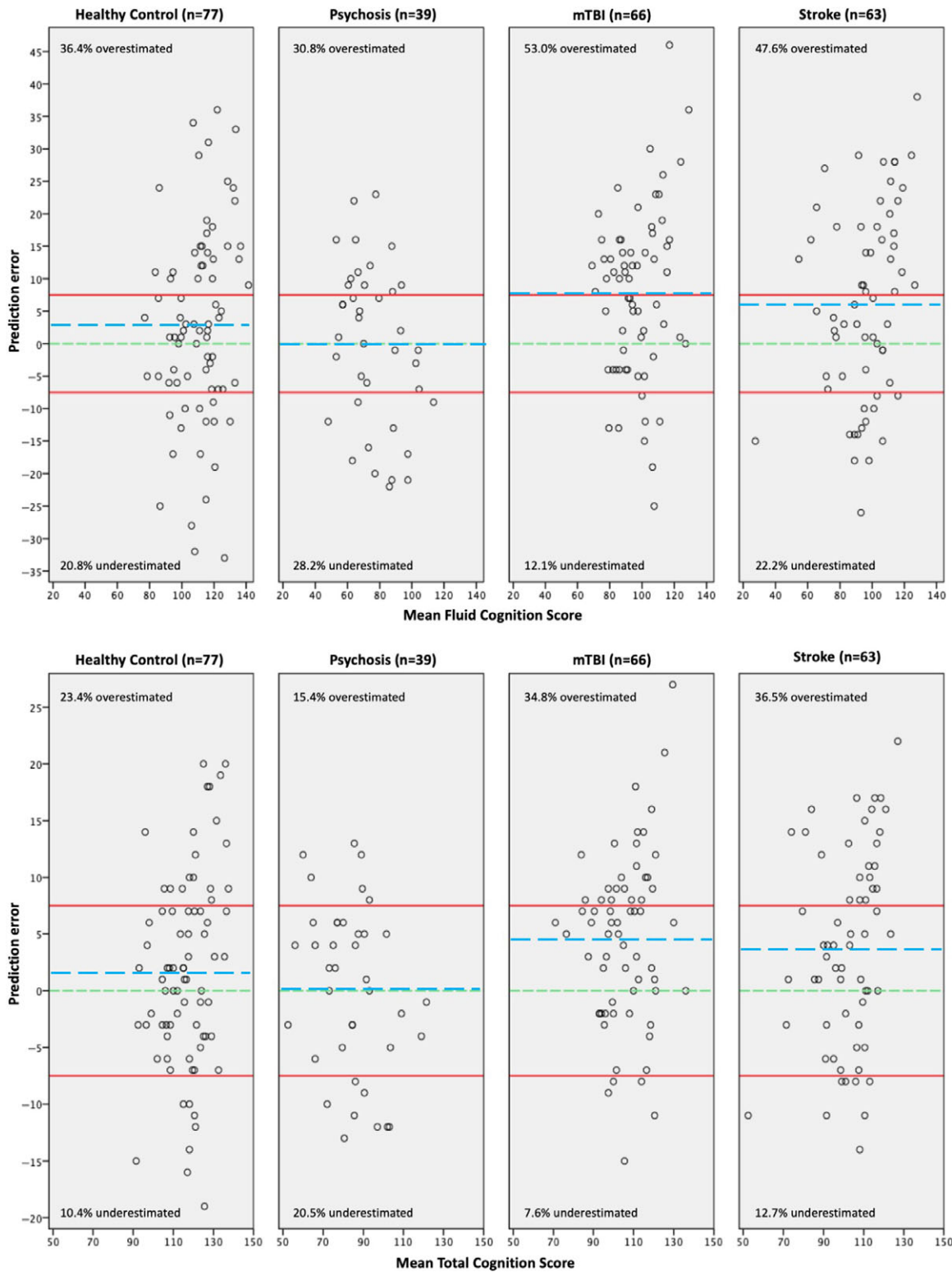


Fig. 1. Bland–Altman plot for fluid cognition (top row) and total cognition (bottom row) prediction errors for healthy controls, stroke, psychosis, and mTBI groups, including mean group difference (blue dotted line). Participant level data are represented by circles. Threshold for acceptable prediction error was set ± 0.5 SD (red lines) from zero (green dotted line). A prediction error of zero indicates equal standard and prorated scores.

significantly affect the assessment of healthy controls and psychosis cohorts at the group level, we cannot confidently conclude that prorated scores are equivalent to standard scores in these groups. Amongst healthy controls, 60.0%

of prorated scores were overestimated or underestimated by a clinically significant margin, and amongst psychosis patients, the rate was 59.0%. Given the significant variability in patient-level performance, these two methods should not

be considered equivalent when considering individual-level data.

Our study has limitations. Our findings are limited to healthy individuals and those with stroke, mTBI, or treatment-resistant psychosis. Future studies should explore whether there may be groups in which an abbreviated protocol may be appropriate. Additionally, we only reported age-corrected scores, which do not control for sex, education, and ethnicity of participants; these factors may influence NIHTB-CB performance (Casaletto et al., 2015).

At this point in time, we are simply comparing in-person testing with prorated versus standard scoring in advance of considering entirely remote adaptations of the NIHTB-CB protocol. We have not prospectively validated an abbreviated remote protocol as we are limited by current physical distancing recommendations related to the COVID-19 pandemic.

In conclusion, an abbreviated NIHTB-CB protocol is a pragmatic solution in the context of physical distancing requirements, but does not constitute a valid replacement for the standard protocol. Our preliminary findings suggest that prorated scores excluding the Flanker Inhibitory Control and Attention, Dimensional Change Card Sort, and Pattern Comparison Processing Speed instruments may tend to overestimate Fluid Composite scores. Thus, a fully remote version of the NIHTB-CB should include adapted versions of the timed instruments. We provide empirical evidence in support of newly updated guidelines by the NIHTB developers, which now state that prorated scores may not be comparable to standard scores (Salesforce, 2020). Still, remote administration of the current abbreviated protocol warrants further validation of the nontimed instruments. These individual instruments, administered remotely, may still benefit continuity of research measuring crystallized cognition and working and episodic memory.

ACKNOWLEDGEMENTS

The authors thank Leah Kuzmuk, Halina Deptuck, Zoe O'Neill, Hadley Pearce, Tasha Klotz, and Hires Gindwani for their assistance in data collection. This article was discussed with the Department of Medical Social Sciences at Northwestern University, which governs the scientific activity of NIH Toolbox, prior to submission. At this time, neither the authors' group nor theirs would recommend an abbreviated NIHTB-CB protocol with prorated scoring as a replacement for the standard protocol.

FINANCIAL SUPPORT

NDS reports salary support from the Michael Smith Foundation for Health Research. TSF is supported by a Heart and Stroke Foundation of Canada National New Investigator Award, a Michael Smith Health Professional Investigator Award, and a Vancouver Coastal Health Research Institute Clinician-Scientist Award.

CONFLICT OF INTEREST

IJT has received consulting fees or sat on advisory boards for Lundbeck Canada, Sumitomo Dainippon, and Community Living British Columbia (CLBC). TSF receives study medication from Bayer Canada. The other authors report no relevant conflicts.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <https://doi.org/10.1017/S1355617720001010>

References

- Berge, E., Stapf, C., Al-Shahi Salman, R., Ford, G.A., Sandercock, P., van der Worp, H.B., ... ESO Trials Network Committee (2016). Methods to improve patient recruitment and retention in stroke trials. *International Journal of Stroke: Official Journal of the International Stroke Society*, 11(6), 663–676.
- Carlozzi, N.E., Goodnight, S., Casaletto, K.B., Goldsmith, A., Heaton, R.K., Wong, A.W.K., ... Tulsy, D.S. (2017). Validation of the NIH toolbox in individuals with neurologic disorders. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 32(5), 555–573.
- Carlozzi, N.E., Tulsy, D.S., Wolf, T.J., Goodnight, S., Heaton, R.K., Casaletto, K.B., ... Heinemann, A.W. (2017). Construct validity of the NIH toolbox cognition battery in individuals with stroke. *Rehabilitation Psychology*, 62(4), 443–454.
- Casaletto, K.B., Umlauf, A., Beaumont, J., Gershon, R., Slotkin, J., ... Heaton, R.K. (2015). Demographically corrected normative standards for the English version of the NIH toolbox cognition battery. *Journal of the International Neuropsychological Society: JINS*, 21(5), 378–391.
- Chevalier, T.M., Stewart, G., Nelson, M., McInerney, R.J., & Brodie, N. (2016). Impaired or not impaired, that is the question: navigating the challenges associated with using Canadian normative data in a comprehensive test battery that contains American tests. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 31(5), 446–455.
- Gostin, L.O., & Wiley, L.F. (2020). Governmental public health powers during the COVID-19 pandemic: stay-at-home orders, business closures, and travel restrictions. *JAMA: The Journal of the American Medical Association*, 323(21), 2137–2138. <https://doi.org/10.1001/jama.2020.5460>
- HealthMeasures Help Desk. (2020a). Retrieved May 5, 2020, from <https://nihtoolbox.force.com/s/article/Coronavirus-COVID-19>
- HealthMeasures Help Desk. (2020b). Retrieved May 2, 2020, from <https://nihtoolbox.force.com/s/article/calculating-scores>
- Koo, T.K., & Li, M.Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- McInnes, K., Friesen, C.L., MacKenzie, D.E., Westwood, D.A., & Boe, S.G. (2017). Mild Traumatic Brain Injury (mTBI) and chronic cognitive impairment: a scoping review. *PLoS One*, 12(4), e0174847. <https://doi.org/10.1371/journal.pone.0174847>
- Meyers, J.E., Zellinger, M.M., Kockler, T., Wagner, M., & Miller, R.M. (2013). A validated seven-subtest short form for the WAIS-IV. *Applied Neuropsychology: Adult*, 20(4), 249–256.

- Mungas, D., Heaton, R., Tulskey, D., Zelazo, P.D., Slotkin, J., Blitz, D., . . . Gershon, R. (2014). Factor structure, convergent validity, and discriminant validity of the NIH Toolbox Cognitive Health Battery (NIHTB-CHB) in adults. *Journal of the International Neuropsychological Society: JINS*, 20(6), 579–587.
- Nys, G.M.S., van Zandvoort, M.J.E., de Kort, P.L.M., Jansen, B.P.W., de Haan, E.H.F., & Kappelle, L.J. (2007). Cognitive disorders in acute stroke: prevalence and clinical determinants. *Cerebrovascular Diseases*, 23(5–6), 408–416.
- O'Brien, J.T., Erkinjuntti, T., Reisberg, B., Roman, G., Sawada, T., Pantoni, L., . . . DeKosky, S.T. (2003). Vascular cognitive impairment. *Lancet Neurology*, 2(2), 89–98.
- Rebchuk, A.D., Deptuck, H.M., O'Neill, Z.R., Fawcett, D.S., Silverberg, N.D., & Field, T.S. (2019). Validation of a novel telehealth administration protocol for the NIH toolbox-cognition battery. *Telemedicine Journal and E-Health: The Official Journal of the American Telemedicine Association*, 25(3), 237–242.
- Salesforce. (2020). Retrieved July 21, 2020, from https://nihtoolbox.my.salesforce.com/sfc/p/#2E000001H4ee/a/2E000000MeFL/gIwGRi9AseGx1M4X5V_wB3hmJw.h0CzmsggMZIs_mDc
- Sheffield, J.M., Karcher, N.R., & Barch, D.M. (2018). Cognitive deficits in psychotic disorders: a lifespan perspective. *Neuropsychology Review*, 28(4), 509–533.
- Silverberg, N.D., & Millis, S.R. (2009). Impairment versus deficiency in neuropsychological assessment: Implications for ecological validity. *Journal of the International Neuropsychological Society: JINS*, 15(1), 94–102.
- Tang, E.Y., Amiesimaka, O., Harrison, S.L., Green, E., Price, C., Robinson, L., . . . Stephan, B.C. (2018). Longitudinal effect of stroke on cognition: a systematic review. *Journal of the American Heart Association*, 7(2), e006443. <https://doi.org/10.1161/JAHA.117.006443>
- Weintraub, S., Dikmen, S.S., Heaton, R.K., Tulskey, D.S., Zelazo, P.D., Bauer, P.J., . . . Gershon, R.C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, 80(11 Suppl 3), S54–S64.
- Weintraub, S., Dikmen, S.S., Heaton, R.K., Tulskey, D.S., Zelazo, P.D., Slotkin, J., . . . Gershon, R. (2014). The cognition battery of the NIH toolbox for assessment of neurological and behavioral function: validation in an adult sample. *Journal of the International Neuropsychological Society: JINS*, 20(6), 567–578.