

Validation of a Novel Telehealth Administration Protocol for the NIH Toolbox-Cognition Battery

Alexander D. Rebchuk, MSc,¹ Halina M. Deptuck,²
Zoe R. O'Neill, BSc,² Delrae S. Fawcett, MSc,³
Noah D. Silverberg, PhD, RPsych, ABPP,^{4,5}
and Thalia S. Field, MD, FRCPC, MHSc^{1,2,6,7}

¹Faculty of Medicine, University of British Columbia, Vancouver, Canada.

²Vancouver Stroke Program, Vancouver, Canada.

³Doctors of BC, Vancouver, Canada.

⁴Division of Physical Medicine and Rehabilitation, University of British Columbia, Vancouver, Canada.

⁵Rehabilitation Research Program, Vancouver Coastal Health Research Institute, Vancouver, Canada.

⁶Division of Neurology, University of British Columbia, Vancouver, Canada.

⁷Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, Canada.

Abstract

Background: Cognition is an important outcome in many clinical trials. The NIH Toolbox-Cognition Battery (NIHTB-CB) is a computerized cognitive assessment designed for clinical research that is administered in-person. Here, we evaluated the equivalency of a novel videoconference protocol for administering the NIHTB-CB. Since our protocol required repeated assessments, we further explored the NIHTB-CB's practice effect. **Materials and Methods:** Twenty-five healthy participants completed the NIHTB-CB under two separate conditions 4 weeks apart. The standard condition followed the recommended administration protocol, whereas the videoconference condition had the examiner and participant in separate rooms but able to communicate over videoconference. A linear mixed-model analysis was performed to explore the fixed effect of testing condition and time on NIHTB-CB performance.

Results: Across all three NIHTB-CB composite scores (total, fluid, and crystallized cognition), no significant fixed effect of administration condition was found. A significant practice effect was observed for the fluid and total cognition composite scores over a 29.0 (± 2.1) day test-retest interval.

Conclusions: Our novel videoconference protocol for the NIHTB-CB is equivalent to the standard protocol in healthy participants, and may provide a solution for researchers seeking

to engage study participants at remote sites. If the NIHTB-CB is used longitudinally to monitor patients, corrections for repeated measures may be required.

Keywords: neuropsychological assessment, telemedicine, validation studies, cognitive assessment

Introduction

Cognition is an important outcome in many clinical trials.¹⁻³ Previously, no standardized cognitive battery has existed for neurological research.⁴ The NIH Toolbox-Cognition Battery (NIHTB-CB), a computerized neurocognitive assessment, was developed to provide a common metric for cognition in clinical research across the lifespan, thereby allowing for common data elements for cognition in clinical trials in the neurosciences.⁴ The NIHTB-CB is administered on a tablet device, scored automatically, and has undergone validation research in healthy controls⁵⁻⁷ and in patients with neurological diseases.⁸ The NIHTB-CB was designed for in-person assessments, with the examiner seated beside the patient. However, clinical trials may increasingly involve assessment of study participants at distributed sites. Telemedicine (medical consultation through use of videoconference technology) is commonly used for acute medical assessments as well as for elective consultations at remote sites.^{9,10} Technology-supported remote assessment may also be used to augment recruitment or minimize travel costs related to clinical trial participation and to ensure representative sampling in epidemiological research.^{11,12}

Other popular brief cognitive screens have been assessed for remote administration. Modified versions of the Montreal Cognitive Assessment (MoCA) and Mini-Mental State Examination (MMSE) have been developed for telephone administration.^{13,14} Whereas the telephone MMSE was equivalent to in-person administration,^{13,15} the MoCA, a more sensitive test for cognitive impairment,¹⁶ was not.¹⁴ If it could be administered remotely, the NIHTB-CB may be better suited for clinical trials because it measures cognition across a broader range of ability with finer gradations between ability levels (i.e., produces normally distributed scores with minimal ceiling effects).⁴⁻⁶

In this study, we explored the validity of a novel video-conference protocol for the NIHTB-CB in healthy participants. Our primary aim was to determine whether administering the NIHTB-CB over videoconference technology was equivalent to in-person administration. Since our protocol required a test-retest design, we further investigated whether there was a practice effect in healthy participants retaking the NIHTB-CB over a 4-week test-retest interval.

Materials and Methods

PARTICIPANTS

Twenty-five ($n = 25$) healthy participants were recruited using poster advertisements at a local university and teaching hospital. Participants were fluent in English, had unimpaired use of their dominant hand, and reported no history of neurological disease, learning disability, attention-deficit disorder, attention-deficit hyperactivity disorder, or an active psychiatric disorder. The experimental protocol was approved by the University of British Columbia's Clinical Research Ethics Board and Vancouver Coastal Health, and conformed to the Declaration of Helsinki. All participants provided written informed consent.

INSTRUMENTATION

Participants' demographic information was captured with a self-completed questionnaire. Options for self-identified race were taken from the NIH Toolbox, whereas options for ethnicity were in keeping with categories from the 2016 Canadian census.¹⁷

The NIHTB-CB comprises seven subtests that evaluate five cognitive domains and broadly assesses fluid and crystallized cognition.^{5,18} The fluid cognition subtests include Flanker Inhibitory Control and Attention (executive function and attention), List Sorting Working Memory (working memory), Dimensional Change Card Sort Test (executive function and shifting), Pattern Comparison Processing Speed (processing speed), and Picture Sequence Memory (episodic memory). The crystallized cognition subtests include Picture Vocabulary (language) and Oral Reading Recognition (language). The specific details of each subtest are described in detail elsewhere.¹⁸ The NIHTB-CB takes approximately 45 min to complete.

The NIHTB-CB generates raw scores that are then converted into demographically corrected *T* scores (mean = 50, standard deviation [SD] = 10) using regression-based norms that adjust for education, gender, race, and ethnicity.⁶ We analyzed demographically corrected *T* scores, where all participants were standardized as "Not Hispanic or Latino" for ethnicity (Table 1).

PROCEDURES

A randomized repeated measures crossover design study was performed to have participants serve as their own controls.

Table 1. Participant Demographic Data

Basic demographics	
Gender	
Males	12 (48)
Females	13 (52)
Age	37 years (25–64 years)
Handedness	
Right	23 (92)
Left	2 (8)
NIH Toolbox demographics	
Race ^a	
Caucasian	18 (72)
Asian	7 (28)
Middle Eastern	1 (4)
Education completed	
Participant's	16 years (15–18 years)
Mother's	16 years (12–16 years)
Additional demographics	
Ethnicity/origins ^a	
European	11 (44)
Canadian or Québécois	8 (32)
South Asian	4 (16)
East Asian	4 (16)
Middle Eastern	1 (4)
Language first learned at home	
English	19 (76)
French	1 (4)
Other ^b	7 (28)

Median (IQR) or *n* (%) is presented.

^aParticipants were permitted to give more than one response, allowing for total percentages >100%.

^bFilipino (4%), Tagalog (4%), Korean (4%), Arabic (4%), Spanish (4%), Cantonese (4%), Hungarian (4%).

Participants completed the NIHTB-CB under two conditions, standard (in-person) and videoconference, with a 4-week test-retest interval. Upon enrolment in the study, participants were randomly assigned to either a standard-then-telehealth group or telehealth-then-standard group. The standard-then-telehealth group completed the standard condition at baseline and the videoconference condition at the 4-week retest interval;

the telehealth-then-standard group completed the opposite sequence.

PROTOCOL AND MODIFICATIONS FOR VIDEOCONFERENCE ADMINISTRATION

The NIHTB-CB was administered on a tablet device (9.7" iPad Pro®; Apple, CA) in a quiet distraction-free room. A single examiner administered the NIHTB-CB. The examiner completed the NIHTB-CB training protocol at our institution and operated under the supervision of a board-certified neuropsychologist. Three examination assistants (EAs) were trained to facilitate the videoconference examinations. EA training took approximately one hour and consisted of being administered the NIHTB-CB assessment and then piloting the videoconference condition once with the examiner. All EAs underwent training to become familiar with the study protocol and NIHTB-CB, but were not trained as NIHTB-CB examiners.

The standard condition was an in-person assessment, which followed the recommended administration protocol for the NIHTB-CB.^{5,18} The videoconference condition consisted of the examiner able to communicate over a videoconference link to the participant and EA in a separate room. A speaker (SPEAK 410; Jabra, Ballerup, Denmark) was used to improve the quality of audio connection in the participant's room. The videoconference link was hosted on the British Columbia Telehealth network, a secure, encrypted, and private network.

The standard NIHTB-CB protocol requires the examiner to be seated beside the patient to present the task instructions, monitor patient compliance, navigate through the NIHTB-CB, and, for only two subtests, to input responses into the tablet.¹⁸ For the videoconference condition, an EA was present in the examination room with the participant. For each videoconference

condition, an EA set up the tablet, and navigated through each subtest on the tablet as instructed by the examiner. *Figure 1* shows the room arrangement for the videoconference condition.

The List Sorting Working Memory and Oral Reading Recognition subtests require the examiner to score the participant in real time and input these scores into the tablet through a keyboard. For the videoconference condition, the examiner scored these subtests and communicated their scoring to the EA, over the videoconference link, who inputted the responses into the tablet. As seen in *Figure 1*, the videomonitor was arranged perpendicularly to the participant to prevent participants from observing the examiner scoring their performance. To ensure that the EA was in unison with the examiner while navigating through the test, on their monitor, the examiner was able to see the tablet screen and participant. To maximize discretion, if a question was scored as correct, the examiner raised their index finger (to make the shape of a one). If the question was scored as incorrect, they touched their index finger to thumb (shape of zero).

STATISTICAL ANALYSIS

A *priori* power analysis was performed using *G*Power*.¹⁹ Estimated values for correlation among repeated measures were based on test-retest reliability coefficient in a prior study that administered the NIHTB-CB in the conventional in-person manner two weeks apart.⁵ Twenty-five participants were required for 80% power to detect an *F*-test effect size of 0.25 using a two-sided alpha of 0.05. Our effect size was chosen to power the study to detect meaningful differences between administration formats.^{5,20}

A linear mixed-model analysis was performed to explore the fixed effect of testing condition (standard and videoconference) and time (baseline and retest) on NIHTB-CB performance in our sample. No demographic variables were included in the models given the assumption that the demographically corrected *T*-scores outputted by the NIHTB-CB would control for demographic differences between participants.⁶ A compound symmetry covariance structure had the lowest Bayesian Information Criterion (e.g., best fit), and was used by all models.

The fixed effects of condition and time were explored for all three composite scores: total cognition, crystallized condition,

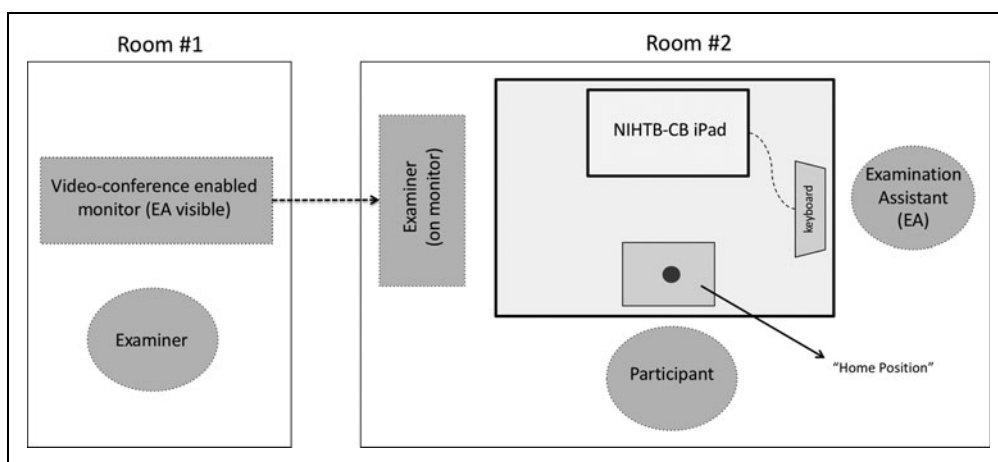


Fig. 1. Testing room arrangement for the videoconference testing condition. The examiner is displayed on a computer monitor, whereas the participant and EA are present in the room. EA, examination assistant; NIHTB-CB, NIH Toolbox-Cognition Battery.

and fluid cognition. Significance levels for the composite scores' fixed effects were set at $p=0.05$. We further explored the fixed effects for all seven NIHTB-CB subtests. For these analyses, we applied a Bonferroni correction to account for multiple comparisons, and significance was set at $p=0.007$.

Results

Twenty-five participants (median age=37 years, IQR 25–64 years) completed the NIHTB-CB under two unique conditions, with a 4-week (mean=29.0 days; SD=2.1 days) test–retest interval. The sample was predominately right handed (92%), Caucasian (72%) or Asian (28%), and learned English as their first language (76%). Participants had completed a median 16 years (IQR 15–18 years) of education and reported heterogeneous ethnicities (Table 1).

We found no significant fixed effect of administration condition on composite scores for total cognition ($p=0.417$), fluid cognition ($p=0.363$), or crystallized cognition ($p=0.896$) (Fig. 2a). We also found no effect of condition when each subtest was examined separately (Fig. 2b).

A significant fixed effect for time was found for fluid cognition ($p<0.001$) and total cognition ($p<0.001$) composite scores. No significant effect for time was found for crystallized cognition ($p=0.076$) composite score (Fig. 2c). When each subtest was analyzed in isolation, a significant fixed effect for time was found for Flanker Inhibitory Control and Attention ($p<0.001$), and Pattern Comparison ($p<0.001$). All other subtests did not reach statistical significance for the fixed effect of time (Fig. 2d).

Discussion

The aim of this study was to investigate the validity of administering the NIHTB-CB over telehealth technology. Healthy adults completed the NIHTB-CB twice, once under standard conditions and once through videoconference, in a randomized counterbalanced order. Our primary finding was that performance for each NIHTB-CB composite score and subtest was equivalent between the videoconference and standard administration conditions.

This finding suggests that the NIHTB-CB in-person or videoconference administration protocols produce comparable scores. Their equivalency should be confirmed in an independent study in patients with neurological disease. We suspect that the control and videoconference conditions were equivalent because of the NIHTB-CB's use of examinee responding on the tablet and automated scoring for most subtests. Only two subtests (Oral Reading Recognition and List Sorting Working Memory) require scoring by the examiner. For these two subtests, a stable high-quality audio connection is imperative as the examiner must hear the participant's response

for each item. For the other five subtests, any momentary audio or video disruptions are likely of little consequence.

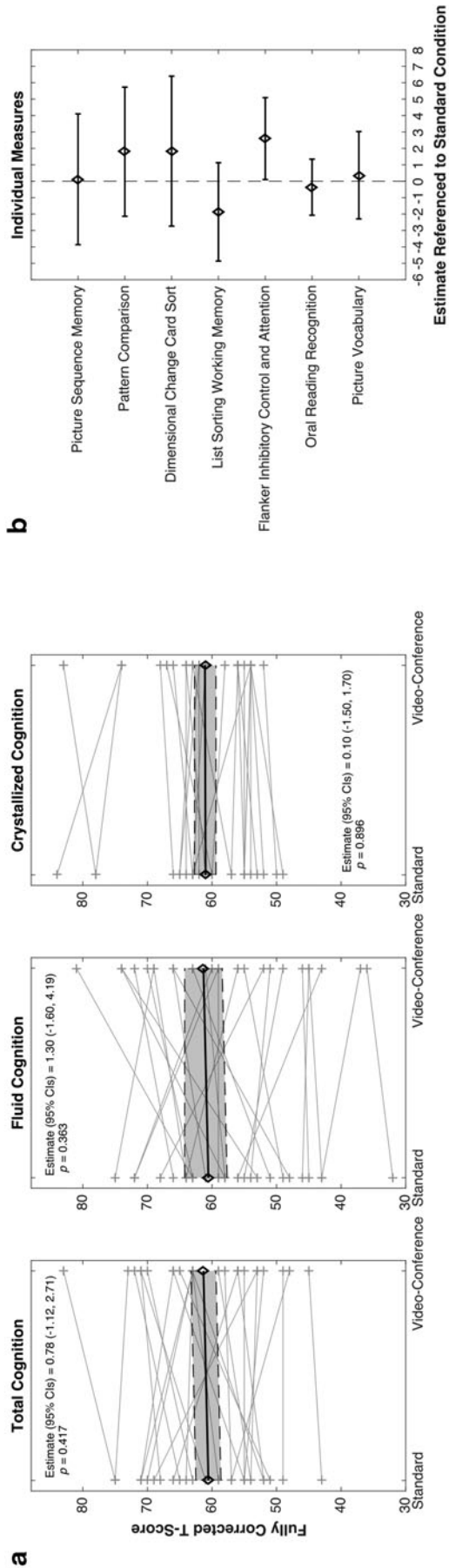
Our findings demonstrate the suitability of the NIHTB-CB for use in clinical trials with recruitment and assessment of participants in geographically widespread areas. Previous attempts at designing neurocognitive evaluations for remote administration required modifications to the content of the original test.^{13–15} We have shown that the NIHTB-CB can be administered over telehealth without any modifications to content or standardized instructions, thereby permitting direct comparison between scores obtained in-person and remotely. The video link permits item-level response speed to be measured accurately, which is not possible for telephone-based cognitive assessment.²¹ Although our study exclusively investigated the NIHTB-CB, our remote administration protocol using videoconference technology appears to have potential as a valid template for remote administration of other neurocognitive tests.

Although the need for an EA for the remote assessment is more resource intensive than for a telephone follow-up assessment, and requires financially compensating two individuals (i.e., examiner and EA), using a centralized examiner would be feasible for a minimally staffed distributed study team conducting in-person baseline or follow-up cognitive assessments. The EA position requires minimal training time and experience, making it a feasible and cost-effective option for a remote site, as compared with a fully trained examiner. An alternative future approach for consideration by the programmers of the NIHTB-CB app would be to allow for remote control of the tablet and NIHTB-CB by an examiner at a site separate from the study participant.

Our observed practice effect for the total cognition and fluid cognition composite scores over a 4-week test–retest period was consistent with a prior study that used a shorter test–retest window (15.5–4.8 days).⁵ Because the same version of the NIHTB-CB was repeated (no validated alternative form is available), practice effects for fluid cognition subtests may reflect familiarization with both the test procedures and the item content.²² Based on our findings and those of Heaton et al.,⁵ total and fluid cognition NIHTB-CB composite scores appear to have a practice effect that persists for at least 4 weeks. Further work is required to establish the durability of the practice effect. This would allow researchers to apply appropriate corrections (e.g., regression-based norms for practice effects) when the NIHTB-CB is administered in a longitudinal study.⁵

This study was designed to assess the validity of a videoconference administration protocol for the NIHTB-CB. We used a single examiner who was unblinded to participants' results. To mitigate any potential scoring bias, no interim statistical analyses were performed. Our small sample size ($n=25$) provided adequate power to answer our research question but may limit the

NIH Toolbox - Cognitive Battery Administration Condition Effect Estimate



NIH Toolbox - Cognitive Battery Administration Time Effect Estimate

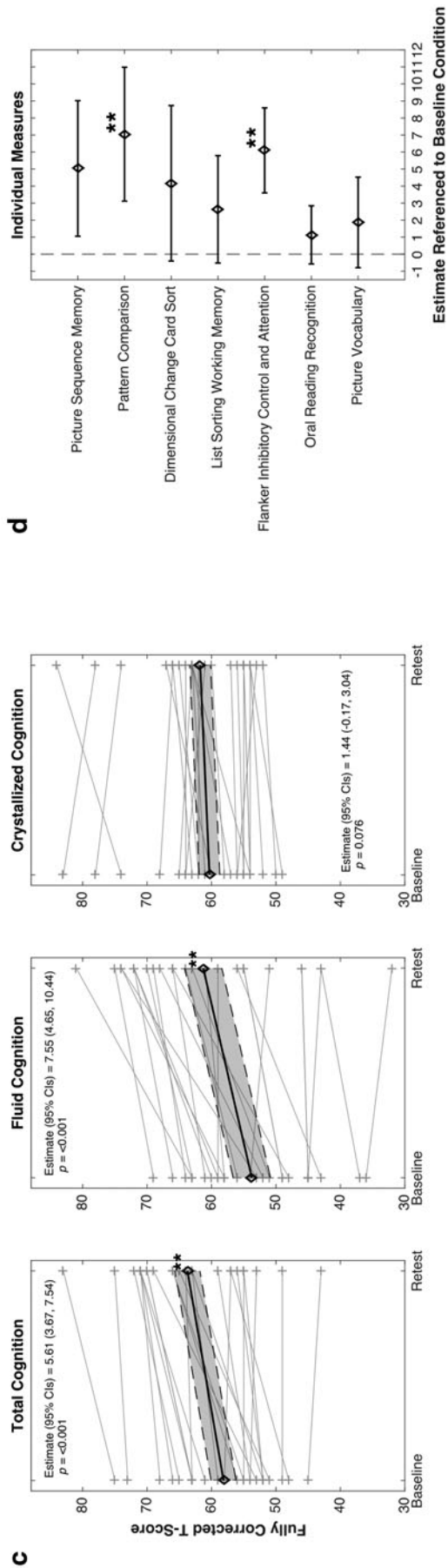


Fig. 2. Estimates of the fixed effect of condition and time on NIHTB-CB performance. Top left **(a)** fixed effect of condition for the three composite scores, with 95% CIs shown in shading. Top right **(b)** fixed effect estimates of condition for all seven subsets, referenced to the standard condition. Bottom left **(c)**, fixed effect of time for the three composite scores, with 95% CI shown in shading. Bottom right **(d)**, fixed effect estimates of time for all seven subsets, referenced to the baseline condition. For NIHTB-CB composite score plots **(a, c)**, fully corrected T-scores (y-axis) are presented. For NIHTB-CB subtest plots **(b, d)**, asterisks indicate statistical significance ($p < 0.007$) and error bars represent 95% confidence intervals. CI, confidence interval.

generalizability of our results. Specifically, the high level of education in our sample (median education = 16 years) may have captured participants more comfortable with test-taking or videoconferencing technology. Our findings require independent confirmation in a larger more diverse sample. Our videoconferencing condition relied upon a fast and stable internet connection, which we recognize is not available in all settings.

In conclusion, our novel videoconferencing protocol for administering the NIHTB-CB appears equivalent to the standard in-person administration protocol in healthy participants. No difference in NIHTB-CB performance for any of the three main composite scores (total, fluid, or crystallized cognition) or for any of the seven subtests was found between standard and videoconferencing administration conditions. Although we observed no learning effect for the crystallized cognition composite score for a 29-day test-retest interval, the fluid and total cognition composite scores demonstrated a significant learning effect over this interval. If the NIHTB-CB is used longitudinally to monitor patients, corrections for repeated measures may be required. Our videoconferencing protocol provides a potential solution for researchers seeking to engage study participants at remote sites.

Acknowledgments

All NIH Toolbox-related materials are copyright of 2006–2018 Northwestern University and the National Institutes of Health. The Department of Medical Social Sciences at Northwestern University, which governs the scientific activity of NIH Toolbox, approved the videoconferencing modification of the protocol. This work was supported by the UBC Summer Student Research Program and Canadian Stroke Trials for Optimized Results (CaSTOR), an initiative funded by the Canadian Institutes of Health Research (CIHR). Both N.D.S. and T.S.F. are supported by both a Health Professional Investigator Award from the Michael Smith Foundation for Health Research and a Clinician-Scientist Career Development Award from the Vancouver Coastal Health Research Institute.

Disclosure Statement

No competing financial interests exist.

REFERENCES

- Lautenschlager NT, Cox KL, Flicker L, et al. Effect of physical activity on cognitive function in older adults at risk for Alzheimer disease: A randomized trial. *JAMA* 2008;300:1027–1037.
- Pearce LA, McClure LA, Anderson DC, et al. Effects of long-term blood pressure lowering and dual antiplatelet treatment on cognitive function in patients with recent lacunar stroke: A secondary analysis from the SP3 randomised trial. *Lancet Neurol* 2014;13:1177–1185.
- Tuszynski MH, Thal L, Pay M, et al. A phase 1 clinical trial of nerve growth factor gene therapy for Alzheimer disease. *Nat Med* 2005;11:551–555.
- Gershon RC, Wagster MV, Hendrie HC, et al. NIH toolbox for assessment of neurological and behavioral function. *Neurology* 2013;80(11 Suppl 3):S2–S6.
- Heaton RK, Akshoomoff N, Tulsky D, et al. Reliability and validity of composite scores from the NIH Toolbox Cognition Battery in adults. *J Int Neuropsychol Soc* 2014;20:588–598.
- Casaletto KB, Umlauf A, Beaumont J, et al. Demographically corrected normative standards for the English version of the NIH Toolbox Cognition Battery. *J Int Neuropsychol Soc* 2015;21:378–391.
- Weintraub S, Dikmen SS, Heaton RK, et al. The cognition battery of the NIH toolbox for assessment of neurological and behavioral function: Validation in an adult sample. *J Int Neuropsychol Soc* 2014;20:567–578.
- Carlozzi NE, Goodnight S, Casaletto KB, et al. Validation of the NIH toolbox in individuals with neurologic disorders. *Arch Clin Neuropsychol* 2017;32:555–573.
- Blacquiere D, Lindsay MP, Foley N, et al. Canadian stroke best practice recommendations: Telestroke best practice guidelines update 2017. *Int J Stroke* 2017;12:886–895.
- Demaerschalk BM, Berg J, Chong BW, et al. American Telemedicine Association: Telestroke Guidelines. *Telemed J E Health* 2017;23:376–389.
- Berge E, Stapf C, Al-Shahi Salman R, et al. Methods to improve patient recruitment and retention in stroke trials. *Int J Stroke* 2016;11:663–676.
- Feldman WB, Kim AS, Chiong W. Trends in recruitment rates for acute stroke trials, 1990–2014. *Stroke* 2017;48:799–801.
- Kennedy RE, Williams CP, Sawyer P, et al. Comparison of in-person and telephone administration of the Mini-Mental State Examination in the University of Alabama at Birmingham Study of Aging. *J Am Geriatr Soc* 2014;62:1928–1932.
- Pendlebury ST, Welch SJ, Cuthbertson FC, et al. Telephone assessment of cognition after transient ischemic attack and stroke. *Stroke* 2013;44:227–229.
- Roccaforte WH, Burke WJ, Bayer BL, Wengel SP. Validation of a telephone version of the mini-mental state examination. *J Am Geriatr Soc* 1992;40:697–702.
- Pendlebury ST, Cuthbertson FC, Welch SJ, et al. Underestimation of cognitive impairment by Mini-Mental State Examination versus the Montreal Cognitive Assessment in patients with transient ischemic attack and stroke. *Stroke* 2010;41:1290–1293.
- Statistics Canada, "Census of Population (appendix 5.1)," Ottawa, 2016.
- Weintraub S, Dikmen SS, Heaton RK, et al. Cognition assessment using the NIH Toolbox. *Neurology* 2013;80(11 Suppl 3):S54–S56.
- Faul F, Erdfelder E, Lang AG, Buchner A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007;39:175–191.
- Cohen J. A power primer. *Psychol Bull* 1992;112:155.
- Gavett BE, Crane PK, Dams-O'Connor K. Bi-factor analyses of the Brief Test of Adult Cognition by Telephone. *NeuroRehabilitation* 2013;32:253–265.
- Benedict RH, Zgaljardic DJ. Practice effects during repeated administrations of memory tests with and without alternate forms. *J Clin Exp Neuropsychol* 1998;20:339–352.

Address correspondence to:

Thalia S. Field, MD, FRCPC, MHS
Vancouver Stroke Program
S169–2211 Wesbrook Mall
V6T 2B5 Vancouver
Canada

E-mail: thalia.field@ubc.ca

Received: January 26, 2018

Revised: February 14, 2018

Accepted: February 21, 2018

Online Publication Date: June 21, 2018