Received: 15 May 2023 | Revised: 14 October 2023 | Accepted: 19 October 2023

DOI: 10.1002/dad2.12498

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

Reliability of remote National Alzheimer's Coordinating Center Uniform Data Set data

Viktorija Smith MS | **Kyan Younes MD** | **Kathleen L. Poston MD, MS** | **Elizabeth C. Mormino PhD** | **Christina B. Young PhD**

Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Palo Alto, California, USA

Correspondence
Christina B. Young, PhD, and Victoria Smith, MS, Department of Neurology and Neurological Sciences, Stanford University School of Medicine, 453 Quarry Road, Palo Alto, CA 94304, USA.
Email: cbyoung@stanford.edu and vpsmith@stanford.edu

Funding information
National Institutes of Health, Grant/Award Numbers: K99AG071837, P30AG066515, R01NS115114, K01AG051718; Alzheimer's Association, Grant/Award Number: AARFD-21-849349

Abstract

INTRODUCTION: The National Alzheimer's Coordinating Center (NACC) Uniform Data Set (UDS) neuropsychological battery is being used to track cognition in participants across the country, but it is unknown if scores obtained through remote administration can be combined with data obtained in person.

METHODS: The remote UDS battery includes the blind version of the Montreal Cognitive Assessment (MoCA), Number Span, Semantic and Phonemic Fluency, and Craft Story. For these tests, we assessed intraclass correlation coefficients (ICCs) between in-person and remote scores in 3838 participants with both in-person and remote UDS assessments, and we compared annual score changes between modalities in a subset that had two remote assessments.

RESULTS: All tests exhibited moderate to good reliability between modalities (ICCs = 0.590–0.787). Annual score changes were also comparable between modalities except for Craft Story Immediate Recall, Semantic Fluency, and Phonemic Fluency.

DISCUSSION: Our findings generally support combining remote and in-person scores for the majority of UDS tests.

KEYWORDS
cognition, intraclass correlation coefficients, NACC, remote assessment, telehealth, teleneuropsychology

1 | BACKGROUND

The coronavirus disease of 2019 (COVID-19) pandemic rapidly accelerated the adoption of remote at-home neuropsychological assessments in research and clinical practice.[1,2] Longitudinal research projects, including those tracking cognition in almost 45,000 participants (≈13,000 currently active) in Alzheimer's Disease Research Centers (ADRCs) across the United States, switched to remote neuropsychological assessments to minimize in-person contact. Previously, gold-standard in-person tests were used, and it is uncertain if switching to remote tests introduced a systematic bias in test scores. In other words, it is unclear if remote scores can be combined with in-person scores, especially because remote assessments were conducted in uncontrolled environments in the participants' homes.

Prior research comparing in-person and remote assessments has generally demonstrated moderate to high consistency, with some variability across tests[3-5]. Given that different interpretations of intraclass correlation coefficients (ICCs) exist, we followed recommendations by Koo and colleagues[6] and interpreted ICCs less than 0.5, between 0.5 and 0.75, between 0.75 and 0.90, and greater than 0.90 as poor, moderate, good, and excellent reliability, respectively. The Montreal Cognitive Assessment (MoCA) has yielded moderate to excellent ICCs ranging from 0.59 to 0.93, and there have been no significant performance differences between remote and in-person modalities.[5,7–11] For memory assessments, most studies have focused on the Hopkins Verbal Learning Test – Revised, and only one study included the Craft Story test showing moderate to good reliability across immediate and delayed recall (ICC = 0.7–0.79).[12–14] For attention and working memory tasks, such as the Digit Span Forward and Backward, most studies have shown moderate reliability (ICC = 0.59–0.75) with no significant differences in mean performance across modalities.[15–17] One study showed better performance with in-person administration of Digit Span Forward, although the authors concluded that the difference was not clinically meaningful as it was within an acceptable test–retest range and the ICC was good (0.75).[13] Phonemic Fluency tasks have shown good to excellent reliability (ICC = 0.83–0.93), with consistent scores across modalities in several studies with varying sample sizes and diagnostic groups.[13–15,18] In contrast, Semantic Fluency tasks have shown mixed results, with some reporting poor to moderate validity metrics (ICC = 0.45–0.74)[12,16] but small effect size differences between modalities.[16] Finally, there is limited examination of executive functioning tests as many of these tests require paper–pencil responses that are difficult to capture remotely. One study showed significantly faster in-person performance on Oral Trails A, a processing speed measure, although the difference was reported to be small and clinically insignificant and the ICC was good (0.83); Oral Trails B, a measure of attention switching, did not show significant differences between modalities with also good ICC (0.79).[13]

Although these studies generally support remote assessments, there are several limitations to consider. First, these studies included relatively small sample sizes (11 to 119 participants per group),[3,4] limiting the generalizability of the results. Second, the majority of studies compared in-person and remote assessments conducted within the same day or 1 to 2 weeks apart,[7,13,15] with few studies examining the 2 to 10 months range.[10,19] Most longitudinal studies that switched to remote assessment in response to COVID-19 repeated assessments annually, making it unclear whether scores would still be comparable in this longer timeframe. Third, remote assessments in previous studies were conducted primarily in controlled satellite clinics[7,15] rather than at home, where the environment can be influenced by additional factors (e.g., noise, technology problems). Only a few studies have examined cognitive assessments conducted in remote at-home settings,[9,20] including one that examined test–retest reliability across 4 to 6 months in 150 healthy adults [age mean (SD) = 33.98 (15.00)].[21] Consistent with previous research, this study showed good test–retest reliability for Phonemic Fluency (ICC = 0.76) and moderate reliability for Digit Span Forward (ICC = 0.61), Digit Span Backward (ICC = 0.66), and Category Fluency (Animals; ICC = 0.52).

The National Alzheimer's Coordinating Center (NACC) provides a large-scale opportunity to overcome these limitations and examine the impact of remote versus in-person administration on neuropsychological test scores. The NACC Uniform Data Set (UDS) neuropsychological battery[22] has been administered remotely from March 2020 in response to the COVID-19 pandemic. The core UDSv3 neuropsychological battery was largely preserved, but tasks with visual stimuli were excluded (i.e., visual items on the MoCA, Benson Complex Figure Copy and Recall, Trail Making Test, and Multilingual Naming Test).[23] Annual neuropsychological assessments were continued and as of December 2021, a total of 3838 participants have had at least one remote assessment at home about 1.479 years (SD = 0.902) after their last in-person assessment.

The goals of the present study were to (1) quantify the reliability between in-person and remotely collected neuropsychological scores, (2) determine if cognitive trajectories differed by modality, and (3) identify modality-specific patterns in cognitively unimpaired (CU) and impaired (CI) older adults.

RESEARCH IN CONTEXT

1. **Systematic review**: We reviewed the literature on the reliability of scores obtained from remote at-home assessments. No studies to date have examined remote at-home administration of the National Alzheimer's Coordinating Center (NACC) Uniform Data Set (UDS) neuropsychological battery, even though it is being used to monitor cognition in more than 45,000 participants over time.
2. **Interpretation**: In-person and remote neuropsychological scores demonstrated moderate to good reliability in both cognitively unimpaired and cognitively impaired older adults. When examining longitudinal data, we found that the rate of score changes was comparable between remote and in-person modalities for 7 of 11 cognitive measures. Our results support the combination of in-person and remote neuropsychological UDS scores for the majority of remote NACC UDS neuropsychological tests.
3. **Future directions**: Further research is needed to understand the impact of remote assessments on longitudinal scores obtained from more than two timepoints as well as the reliability of neuropsychological tests that require visual stimuli.

2 | METHODS

2.1 | Participants

We examined NACC UDS data collected between September 2005 and December 2021 from 43 ADRCs; 37/43 ADRCs are currently active and 26/43 ADRCs collected remote neuropsychological data

TABLE 1 Demographic information for subsets of participants included for analyses.

	(A) Association analysis (*n* = 3838)	(B) Between-subject comparison of annual score changes		(C) Within-subject comparison of annual score changes (*n* = 263)
		Remote (*n* = 311)	In-person (*n* = 622)	
Age, mean (SD)	74.8 (8.90)	74.7 (8.35)	74.2 (9.15)	75.1 (8.45)
Sex, n (%)	–	–	–	–
Male	1493 (38.9%)	107 (34.4%)	209 (33.6%)	89 (33.8%)
Female	2345 (61.1%)	204 (65.6%)	413 (66.4%)	174 (66.2%)
Education, mean (SD)	16.3 (2.71)	16.4 (2.65)	16.6 (2.88)	16.3 (2.69)
Race, n (%)	–	–	–	–
White	3101 (80.8%)	257 (82.6%)	524 (84.2%)	214 (81.4%)
Black or African American	560 (14.6%)	43 (13.8%)	75 (12.1%)	39 (14.8%)
Asian	121 (3.2%)	5 (1.6%)	16 (2.6%)	5 (1.9%)
Other	56 (1.5%)	6 (1.9%)	7 (1.1%)	5 (1.9%)
Ethnicity, n (%)	–	–	–	–
Non-Hispanic	3648 (95%)	300 (96.5%)	599 (96.3%)	253 (96.2%)
Hispanic	190 (5%)	11 (3.5%)	23 (3.7%)	10 (3.8%)
Cognitive status, n (%)	–	–	–	–
Cognitively unimpaired	2697 (70.3%)	264 (84.9%)	483 (77.7%)	226 (85.9%)
Impaired—Not MCI	197 (5.1%)	7 (2.3%)	33 (5.3%)	5 (1.9%)
MCI	547 (14.3%)	25 (8%)	61 (9.8%)	19 (7.2%)
Dementia	397 (10.3%)	15 (4.8%)	45 (7.2%)	13 (4.9%)
Visit number, mean (SD)	5.87 (3.76)	6.04 (3.37)	5.17 (1.30)	6.77 (3.16)

Remote visit type, n (%)	**–**	**1st visit**	**2nd visit**	**–**	**1st visit**	**2nd visit**
Telephone	2608 (67.9%)	265 (85.2%)	242 (77.8%)	–	234 (89%)	224 (85.2%)
Videoconferencing	1143 (29.8%)	46 (14.8%)	65 (20.9%)	–	29 (11%)	36 (13.7%)
Combination	87 (2.3%)	0	4 (1.3%)	–	0	3 (1.1%)

In subsets (A), (B) remote, and (C), diagnosis is based on initial remote visit; for subset (B) in-person, diagnosis based on first in-person visit included in the analysis.

(downloaded on 02/23/2022, "investigator_nacc56.csv").[24] UDS data from 44,713 unique participants totaling 164,265 visits were available. Three subsets of this full data set were created. The first subset consisted of 3838 participants who had both a remote and in-person assessment. The second subset consisted of 311 participants who had two remote assessments as well as a group of 622 participants with two in-person assessments matched for age, sex, education, cognitive status, race, ethnicity, primary language, visit number, years between two visits, and number of completed neuropsychological tests (Methods S1). The third subset consisted of 263 participants who had two remote assessments as well as two preceding in-person assessments. Across all subsets, most participants completed their remote assessments via telephone (Table 1). Missing education and race data were imputed (Methods S2). Cognitive status was selected based on the "NACCUDSD" variable, with individuals categorized as "Normal cognition," "Impaired not mild cognitive impairment (MCI)," "MCI," and "Dementia." We used cognitive status at the first relevant timepoint for each analysis (e.g., the cognitive status at the in-person assessment was used for analyses comparing consecutive in-person and remote scores). Given the small sample size, participants who had a diagnosis of MCI not due to Alzheimer's disease were excluded from analyses that examined differences across cognitive status ($n = 214$ in association analysis, $n = 40$ between-subject, and $n = 8$ within-subject analyses). Institutional review board approval and informed consent was obtained from all participants at each research center.

2.2 | Cognitive tests

The UDS neuropsychological measures that were administered at both in-person and remote visits were the blind version of the MoCA,[25] Craft Story,[26,27] Semantic Fluency (Animals and Vegetables),[28] Phonemic Fluency ("F" and "L" words), and Number Span Forward and Backward.[22] The MoCA is a global cognitive screening measure and the blind version excludes visually-based items (i.e., trail making, cube drawing, picture naming, and clock drawing); the MoCA-blind was administered remotely and a MoCA-blind score was calculated from item-level responses for in-person visits. The Craft Story is a verbal story memory test; we used immediate and delayed verbatim recall scores. Semantic and Phonemic fluency are measures of language and executive functioning; total semantic and total phonemic scores were calculated by summing the individual components. Number Span Forward and Backward are measures of attention and working memory.

2.3 | Statistical analyses

Statistical analyses were performed in R v.4.2.1. Three main sets of analyses were conducted.

First, we examined the reliability between in-person and remote scores for each test. In individuals who had at least one in-person and one immediately following remote visit (Table 1A; $n = 3838$), we calculated ICCs using a two-way mixed, agreement design, and the single-measure output. Within this same subset of participants, we also calculated ICCs between two in-person visits in those who had two previous in-person visits ($n = 3016$). We additionally used linear regression models with in-person score predicting remote score at the subsequent visit, and also modeled an interaction term of cognitive status by in-person score to determine if these associations differed between CU and CI (i.e., MCI or dementia) groups.

Second, we examined if score trajectories differed by modality. We first examined this question using a between-subjects approach because remote assessments always occurred after in-person assessments and there could be increased task familiarity during the remote assessment. We used the R MatchIt package with 1:2 matching to identify a visit-number and demographically matched group of 622 participants with at least two in-person visits (Methods S1) to compare to the 311 participants who completed two remote assessments 0.868 years (SD = 0.200 years) apart (Table 1B). Annual score changes were calculated for each neuropsychological test, each modality, and each subject by extracting the slope from linear regressions using time from the first assessment to predict score. Linear regressions were performed on *z*-scored values, which were calculated using the means and SDs from all CU individuals at their first visit. *T*-tests were used to compare annual *z*-score changes between groups (in-person vs remote). We ran two-way analyses of variance (ANOVAs) with between-subject factors of modality (in-person, remote) and cognitive status (unimpaired, impaired), as well as post hoc analyses with pairwise contrasts to examine differences between cognitive groups.

We also examined whether score trajectories differed by modality using a within-subject approach. In participants with at least two in-person visits and two remote visits ($n = 263$; Table 1C), we calculated the annual *z*-score change for each test for each subject during their in-person assessments and during their two remote assessments. Paired-sample *t*-tests were used to compare annual *z*-score changes between in-person and remote visits. We ran two-way ANOVAs with a within-subject factor of modality (in-person, remote) and a between-subject factor of cognitive status (unimpaired, impaired), as well as post hoc analyses with pairwise contrasts to understand differences between cognitive groups.

3 | RESULTS

3.1 | Demographics

Across all three subsets, the average age was around 75 years of age and the majority of the participants were female, non-Hispanic, White, and highly educated (Table 1). The majority of participants were CU. Most remote visits began in March 2020 at visit 5.87 (SD = 3.76), which was ≈1.48 years (SD = 0.90) after the last in-person visit. Remote neuropsychological assessments via telephone, videoconferencing, and a combination of telephone and videoconferencing were completed by 77.8%–89%, 11%–20.9%, and 1.1%–2.3% of participants, respectively (Table 1).

3.2 | Reliability between in-person and remote scores

ICCs for in-person and remote scores ranged from moderate to good, with MoCA-blind, Craft Story Delayed Recall, Semantic Fluency (Total), and Phonemic Fluency (Total) showing the highest reliabilities (Table 2A; Figure S1)[6]. Although slightly lower, these ICCs were highly consistent with those calculated from two in-person assessments (Table 2A). All ICCs were significant. Correlations between in-person and remote scores ranged from 0.595 for Number Span Backward to 0.798 for Semantic Fluency (Total) (Table S1A), and in-person and remote scores were significantly related for all tests (all p values < 0.001, Figure 1A).

Within CI individuals, ICCs ranged from 0.501 for Number Span Backward to 0.771 for Craft Story Delayed Recall and were similar to the whole group ICCs (Table 2B). Within CU individuals, the ICCs were lower and ranged from 0.565 for Number Span Backward to 0.749 for Phonemic Fluency (Total) (Table 2B). Correlation values within each cognitive group were very similar to ICC values (Table S1B) and significant Score x Cognitive Status interactions were observed for MoCA-blind ($\beta = 0.39$, SE = 0.03, $p < 0.001$), Craft Immediate Recall ($\beta = 0.13$, SE = 0.03, $p < 0.001$) and Delayed Recall ($\beta = 0.21$, SE = 0.03, $p < 0.001$), Semantic Fluency—Animals ($\beta = 0.06$, SE = 0.03, $p = 0.046$), Semantic Fluency—Vegetables ($\beta = 0.10$, SE = 0.03, $p = 0.002$), Semantic Fluency Total ($\beta = 0.08$, SE = 0.03, $p = 0.004$), and Number Span Forward ($\beta = -0.09$, SE = 0.03, $p = 0.008$) (Figure 1B); no significant interactions were observed for Phonemic Fluency (F, L, Total), and Number Span Backward (all p's > 0.385). Of the tests that showed significant Score x Cognitive Status interactions, remote raw scores were more consistent with in-person raw scores for all tests except for Number Span Forward in the CI group.

TABLE 2 Intraclass coefficients (ICCs) for remote and in-person visits (left column) and two preceding in-person visits (right column) in (A) all participants and (B) within cognitive status groups.

(A) All participants	ICC (95% CI)	
	Remote and last preceding in-person visit (n = 3838)	Two previous in-person visits (n = 3016)
MOCA blind	0.758 (0.743–0.772)	0.771 (0.755-0.785)
Craft story immediate recall	0.711 (0.694–0.727)	0.725 (0.707-0.742)
Craft story delayed recall	0.770 (0.756–0.783)	0.775 (0.759-0.789)
Semantic fluency-animals	0.720 (0.696–0.741)	0.741 (0.724-0.757)
Semantic fluency-vegetables	0.704 (0.678–0.728)	0.714 (0.696-0.732)
Semantic fluency-total	0.787 (0.758–0.812)	0.808 (0.794-0.821)
Phonemic fluency-F	0.686 (0.664–0.706)	0.703 (0.685-0.721)
Phonemic fluency-L	0.697 (0.680–0.714)	0.700 (0.681-0.718)
Phonemic fluency-total	0.771 (0.754–0.786)	0.787 (0.773-0.800)
Number span forward	0.653 (0.634–0.672)	0.710 (0.691-0.728)
Number span backward	0.590 (0.566–0.613)	0.693 (0.673-0.711)

(B) Within cognitive status groups	Cognitively unimpaired		Cognitively impaired	
	Remote and last preceding in-person visit (n = 2730)	Two previous in-person visits (n = 2236)	Remote and last preceding in-person visit (n = 894)	Two previous in-person visits (n = 581)
MOCA blind	0.567 (0.538–0.594)	0.643 (0.607–0.659)	0.694 (0.648–0.734)	0.734 (0.692–0.771)
Craft story immediate recall	0.595 (0.570–0.620)	0.658 (0.632–0.682)	0.671 (0.630–0.707)	0.689 (0.642–0.732)
Craft story delayed recall	0.623 (0.598–0.647)	0.686 (0.662–0.708)	0.771 (0.741–0.798)	0.760 (0.722–0.794)
Semantic fluency-animals	0.653 (0.627–0.677)	0.692 (0.669–0.713)	0.663 (0.592–0.719)	0.683 (0.635–0.726)
Semantic fluency-vegetables	0.616 (0.587–0.642)	0.650 (0.626–0.674)	0.679 (0.612–0.733)	0.698 (0.650–0.740)
Semantic fluency-total	0.720 (0.692–0.746)	0.758 (0.739–0.775)	0.731 (0.641–0.793)	0.767 (0.724–0.803)
Phonemic fluency-F	0.659 (0.636–0.681)	0.681 (0.658–0.703)	0.632 (0.571–0.685)	0.636 (0.585–0.683)
Phonemic fluency-L	0.666 (0.645–0.687)	0.677 (0.653–0.699)	0.639 (0.587–0.685)	0.662 (0.613–0.706)
Phonemic fluency-total	0.749 (0.731–0.765)	0.771 (0.754–0.788)	0.712 (0.650–0.762)	0.740 (0.700–0.775)
Number span forward	0.657 (0.635–0.678)	0.700 (0.678–0.721)	0.536 (0.485–0.583)	0.684 (0.634–0.728)
Number span backward	0.565 (0.533–0.594)	0.673 (0.649–0.696)	0.501 (0.448–0.551)	0.640 (0.587–0.687)

3.3 | Consistency of score trajectories across modalities

With a between-subjects approach, there were no statistically significant differences in annual z-score changes between modalities for any of the tests (Figure 2A; Table S2A). No significant Modality x Cognitive Status interactions were observed (all p's > 0.180) except for MoCA-blind ($F(1,720) = 5.51$, $p = 0.019$; Figure 3A), which was characterized by a significant increase in z-scores during remote assessments in the CU group ($t(720) = -1.975$, $p = 0.049$) that was not observed during in-person assessments or in the CU group. Raw score changes, rather than z-score changes, are shown in Figure S2A and Figure S3A.

To capitalize on the rich longitudinal data that were available, we next used a within-subject approach to determine whether annual z-score changes observed on two remote assessments differed from those observed on in-person assessments within the same individual. In other words, we examined if an individual's score trajectory changed when visits were switched from in-person to remote. With this within-subject approach, there were no significant differences between modalities in annual z-score change for MoCA-blind, Craft Delayed Recall, Phonemic Fluency (L, Total), Number Span Forward,

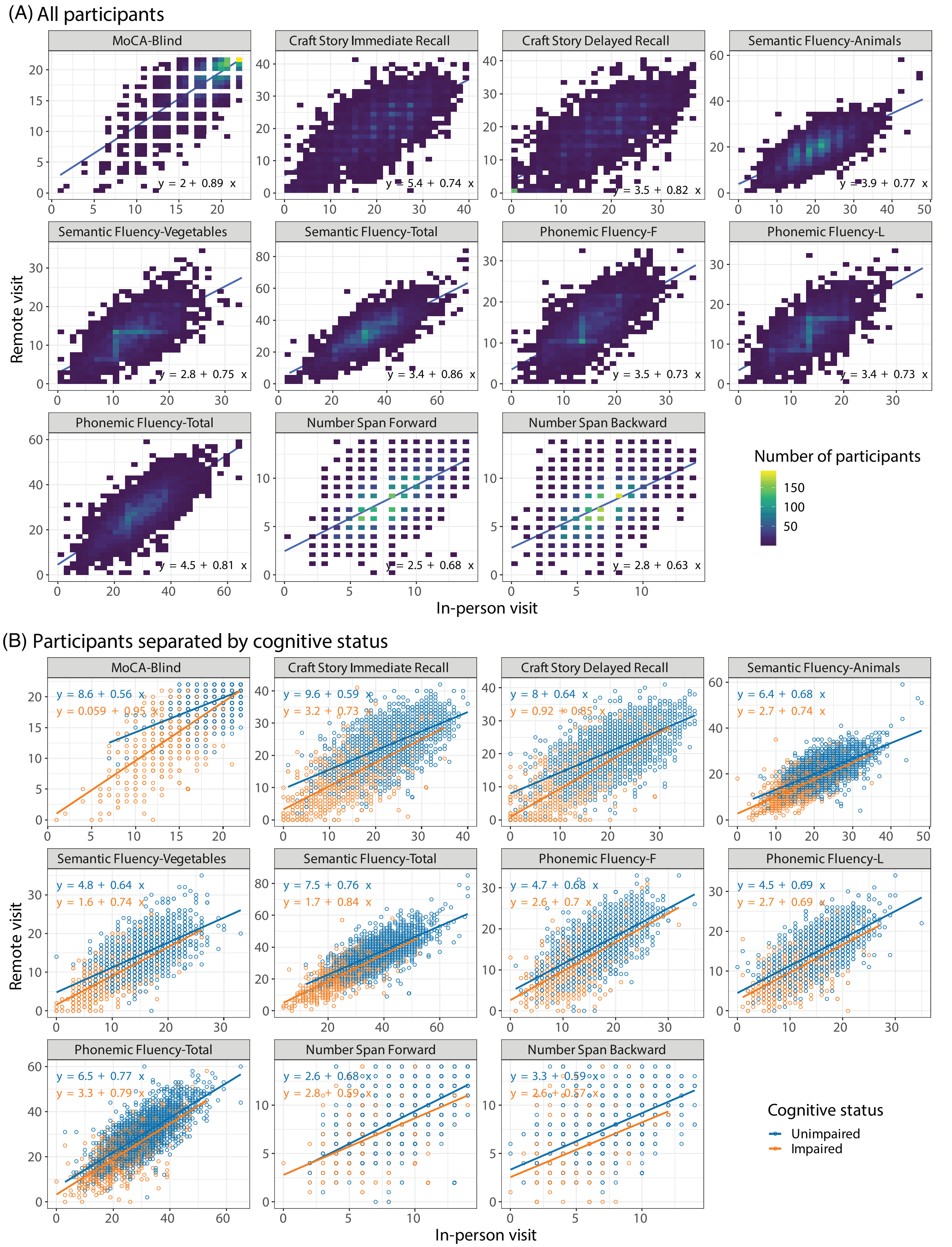

FIGURE 1 Scatterplots of in-person and remote scores. In-person and remote scores are highly correlated (A) in all participants with a remote visit and (B) within cognitively unimpaired and cognitively impaired groups.

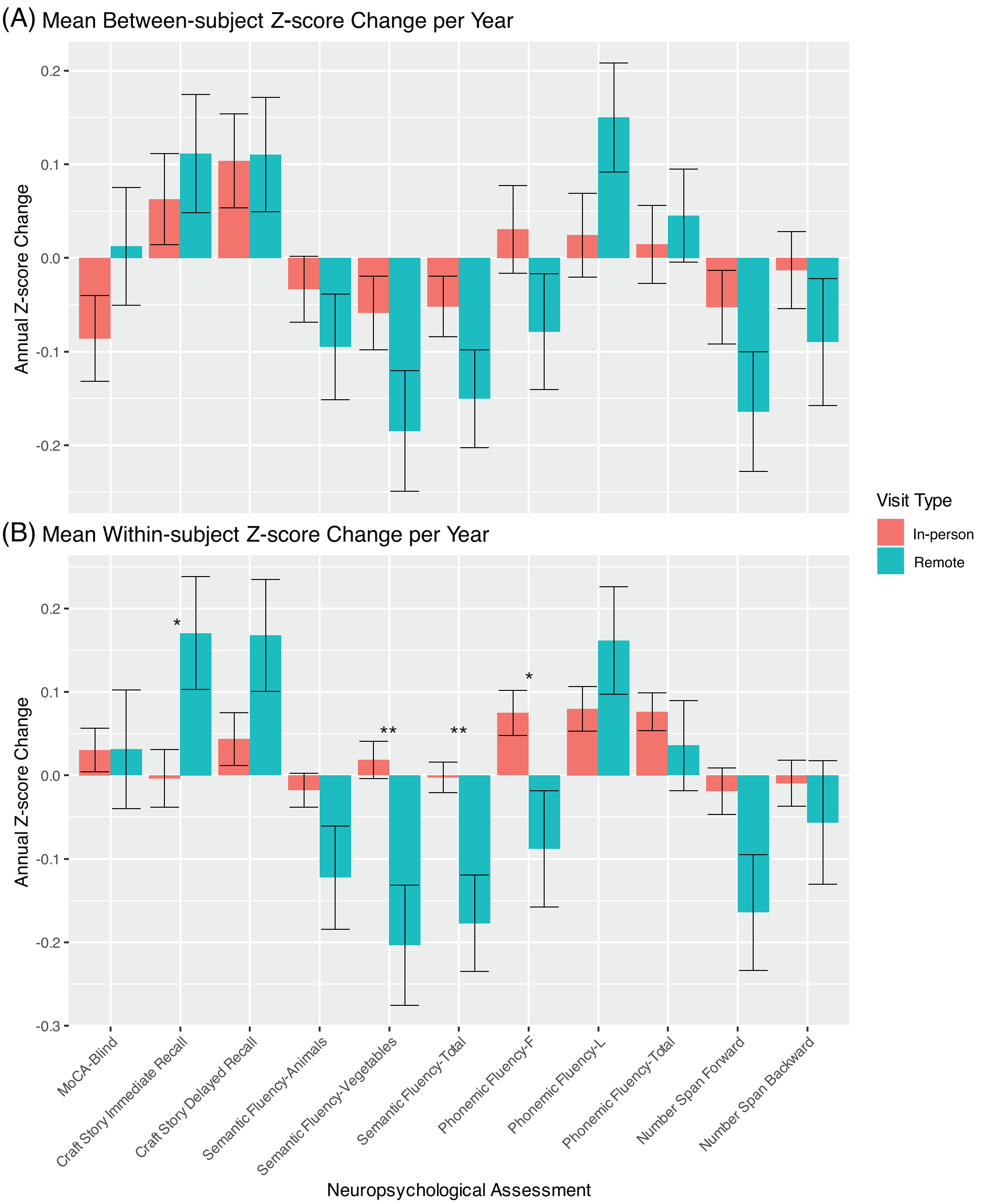

FIGURE 2 Comparisons of score trajectories calculated from multiple remote and multiple in-person assessments. (A) Between-subject comparisons of annual *z*-score changes from two remote visits and two in-person visits using an independent sample matched for visit number and demographic variables. (B) Within-subject comparisons of the annual *z*-score changes from two remote visits and all preceding in-person visits. *Note*: * $p < 0.05$, ** $p < 0.01$.

or Number Span Backward (all *p*'s > 0.080; Figure 2B; Table S2B). In contrast, *z*-scores increased significantly for Craft Immediate Recall and decreased significantly for Semantic Fluency (Vegetables, Total) during repeat remote assessments, and these score changes were not observed during in-person assessments; for Phonemic Fluency (F), *z*-scores increased over time for in-person assessments and decreased over time for remote assessments (Figure 2B; Table S2B). When we examined Modality x Cognitive Status interactions, there were no significant interactions on any tests (all *p*'s > 0.230) except MoCA-blind ($F(1,227) = 9.43$, $p = 0.002$), Phonemic Fluency—F ($F(1,235) = 5.71$, $p = 0.018$), and Phonemic Fluency–Total ($F(1,234) = 5.98$, $p = 0.015$; Figure 3B). On these three tests, CI individuals showed significant declines in *z*-scores during remote assessments (MoCA-blind, $t(227) = 2.85$, $p = 0.005$; Phonemic Fluency—F, $t(235) = 2.89$, $p = 0.004$; Phonemic Fluency–Total, $t(234) = 2.50$, $p = 0.013$), which was not observed during in-person assessments or in the CI group. Raw score changes, rather than *z*-score changes, are shown in Figure S2B and Figure S3B.

All results were similar when calculated separately for participants who completed their remote assessments over the telephone, videoconference, or a combination of telephone and videoconference (Tables S3–S5; Figures S4–S6).

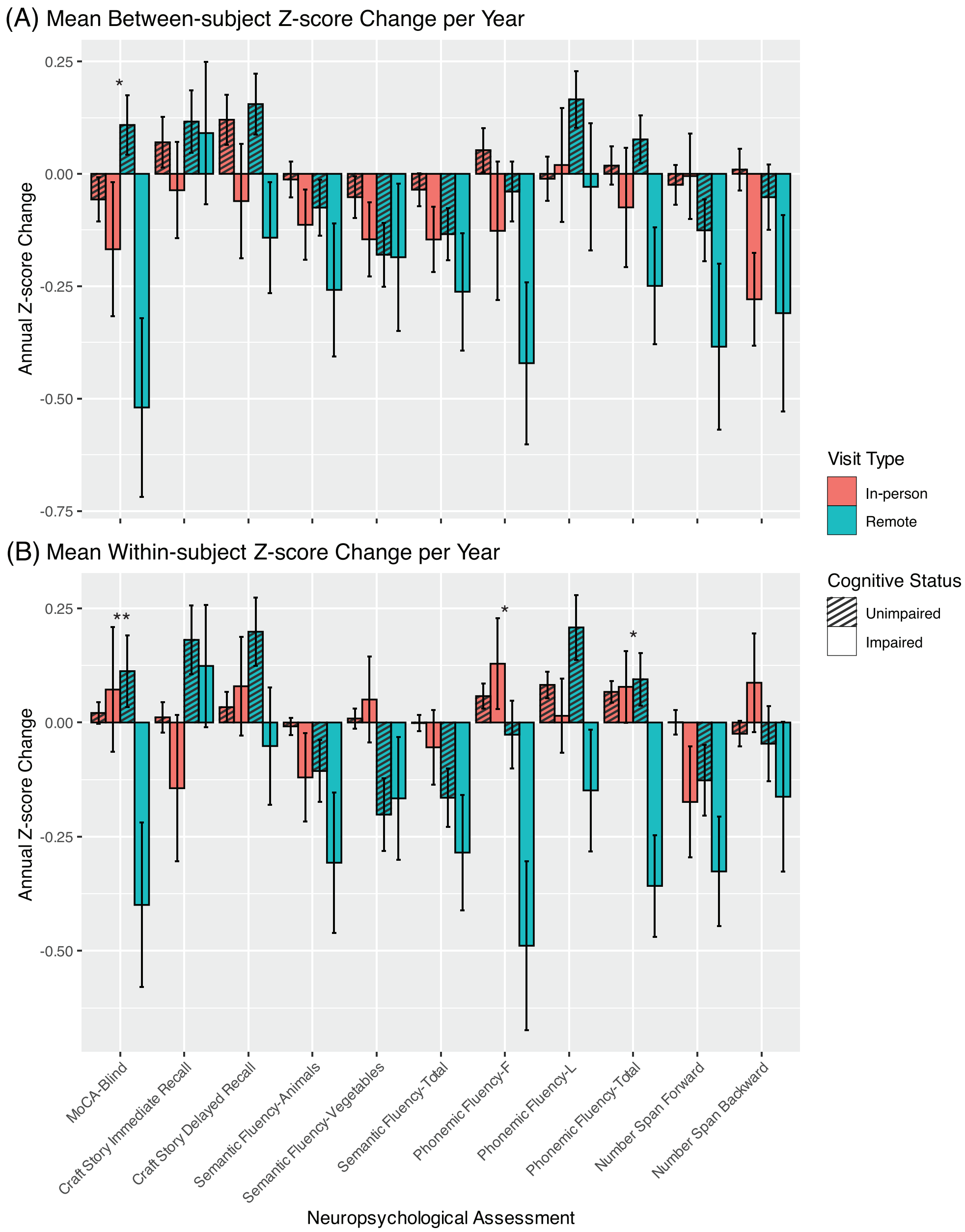

FIGURE 3 Comparison of score trajectories calculated from multiple remote and multiple in-person assessments in cognitively unimpaired and cognitively impaired groups. (A) Between-subject comparisons of annual *z*-score changes from two remote visits and two in-person visits using an independent sample matched for visit number and demographic variables. (B) Within-subject comparisons of the annual *z*-score changes from two remote visits and all preceding in-person visits. *Note*: * $p < 0.05$, ** $p < 0.01$.

4 | DISCUSSION

In a large sample of older adults who underwent in-person and remote cognitive assessments with the UDS neuropsychological battery, we demonstrated moderate to good reliability between in-person and remote scores for CU and CI individuals. Leveraging longitudinal data, we showed that score trajectories were comparable for most UDS tests between modalities. Overall, our results suggest that UDS data collected from in-person and remote sessions can be combined and, more broadly, our results converged with previous conclusions that remote administration of cognitive tests is feasible and valid even in individuals with cognitive impairment.[4]

Our finding of moderate to good ICCs (ICC = 0.590–0.787) between in-person and remote scores for all neuropsychological tests were consistent with limited previous literature.[3,4] It is important to note that the ICCs between remote and in-person assessments across all tests were comparable to ICCs obtained from two in-person visits, suggesting that the reliability of combining in-person and remote scores is similar to what is normally done when combining in-person scores. ICCs remained moderate to good within CU (ranging from 0.565 to

0.749) and CI (ranging from 0.501 to 0.771) groups, providing support for remote assessments in both groups. Of interest, despite high correlations, there was a systematic increase in raw scores at the remote assessment across all tests as denoted by the positive intercepts in the linear regression models (Figure 1A). The increase was larger in CU individuals (Figure 1B) and can be partially explained by practice effects, or expected improved performance on repeated testing due to learning,[29,30] as the remote assessment always occurred after the in-person assessment. Previous studies have shown that CU individuals without biomarker evidence of Alzheimer's disease show stronger practice effects than CU individuals with biomarker evidence of Alzheimer's disease, individuals with MCI, and individuals with dementia.[31–35] For MoCA-blind, the systematic increase in raw remote score for CU older adults likely also reflects the restricted range in score, as most of the participants were at or near ceiling (Figure 1A).[36] In addition, CU individuals with a relatively low MoCA-blind score at the in-person visit must have had adequate scores on other neuropsychological tests to still be considered "unimpaired," and a low MoCA-blind score at a single timepoint may reflect random noise rather than true impairment.[37] This interpretation is consistent with our finding that the MoCA-blind ICC was lower in the CU group in comparison to the CI group, and had the second lowest ICC within the CU group (Table 2B). Experimental designs that vary the order of remote and in-person assessments will be helpful to determine if the score increases that are observed here during remote testing are reflective of test properties (e.g., practice effects, ceiling effects) or are reflective of something inherent to the remote modality.

Although there were no statistically significant differences in score trajectories between modalities when examined with between-subject matched groups, within-subject analyses showed differences on 4 of 11 neuropsychological tests. These discrepant results are likely due to the limited longitudinal data available for remote assessments. For our within-subject analyses, we calculated annual *z*-score change from up to 14 in-person assessments, providing a stable and accurate measure of score change over time for in-person assessments. In contrast, annual *z*-score changes for remote assessments were calculated from only two timepoints, which may capture normal score fluctuations that are not reflective of an individual's overall score trajectory. Indeed, the standard errors for in-person annual *z*-score changes were substantially smaller than those for remote annual *z*-score changes and the standard errors from the visit-number and demographically matched group with two in-person visits, highlighting the increased precision of the within-subject in-person measure (Figure 2).

Inherent features of a test may also make it more susceptible to increased variability during remote administration. Tasks that require sustained attention for longer periods may be impacted by remote administration, particularly if there are numerous salient distractors present in the environment. For example, Craft Story requires sustained attention for several sentences and attention ability is known to affect performance on immediate recall on story and word list memory tests.[38,39] It is interesting that three of the four tests that showed significant differences in score trajectories when examined in a within-subject manner were language fluency tests. Although previous studies focused on Semantic and Phonemic Fluency have generally shown good agreement across modalities,[4,6] some studies have reported significant differences[16,40] and poor to moderate reliability.[12] It has been suggested that using only a single trial of Semantic Fluency (e.g., Animals) may yield a score that lacks sufficient reliability.[4] Successful performance on fluency tasks requires intact semantic knowledge, ability to quickly generate and apply a search strategy, ability to associate and switch between subcategories, and monitoring of responses to avoid repetitions and out-of-category items.[41–43] It is possible that remote environments contain more distractions that can impact some of these processes (e.g., monitoring of responses, speed of word production). Inclusion of additional fluency trials may help provide a more stable estimate, although this needs to be tested explicitly in future research. Taken together, both the limited number of timepoints of the remote assessment data and potential factors such as fluctuating attention and varying cognitive strategies may be contributing to the discrepant annual score trajectory results observed for Craft Story Immediate Recall, Semantic Fluency, and Phonemic Fluency.

Our study has several methodological considerations. First, there may be a biased selection of who completed remote assessments due to a range of factors including technological readiness, access to good phone or internet connection, clinical severity, emotional well-being, and self-efficacy.[11,44–46] Second, annual *z*-score changes for remote assessments were calculated based on two timepoints, which is suboptimal for measuring change over time. Examination of reliable change indices and base rates of score changes can also improve the measurement and interpretation of score changes.[47] Third, our study included mixed telecommunication methods and most participants completed their remote assessments via telephone. Although our results demonstrate that the reliability was similar between telephone, videoconference, and combination of telephone and videoconference (Table S3A; Figure S4), previous studies have suggested that videoconferencing may be superior to telephone administration (see literature review[4]). Fourth, we did not correct for multiple comparisons when examining interaction effects among CU and CI individuals. Fifth, despite the data compromising multiple research centers across the United States, the participants included are predominantly Caucasian, non-Hispanic and highly educated, limiting the applicability of results with more inclusive and representative research required. Finally, because the remote NACC UDS battery did not include visuospatial assessments or measures of executive functioning that involved visual stimuli, we are unable to comment on the reliability of such tests.

Despite these limitations, this study has important implications for groups working with NACC UDS data and for those interested in remote neuropsychological testing more broadly. Our study provides evidence for the combination of remote and in-person UDS scores with the caveat that further examination of Craft Story Immediate Recall, Semantic Fluency, and Phonemic Fluency are needed. Overall, this study examining data from the NACC UDS is the largest comparison of remote and in-person neuropsychological assessment scores to date, and our results generally provide support for the feasibility and validity of remote neuropsychological testing across cognitive impairment levels.

ACKNOWLEDGMENTS

The National Alzheimer's Coordinating Center (NACC) database is funded by National Institute on Aging/National Institutes of Health (NIA/NIH) Grant U24 AG072122. NACC data are contributed by the NIA-funded Alzheimer's Disease Research Centers (ADRCs): P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI John Morris, MD), P30 AG066518 (PI Jeffrey Kaye, MD), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI David Bennett, MD), P30 AG072978 (PI Neil Kowall, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Eric Reiman, MD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Todd Golde, MD, PhD), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P20 AG068024 (PI Erik Roberson, MD, PhD), P20 AG068053 (PI Justin Miller, PhD), P20 AG068077 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James Leverenz, MD).

CONFLICT OF INTEREST STATEMENT

E. C. Mormino is a paid consultant for NeuroTrack, Roche, Genentech, and Eli Lilly. K. L. Poston is a paid consulstant for CuraSen Therapeutics Inc. All other authors have no disclosures relevant to the manuscript. Author disclosures are available in the Supporting Information.

CONSENT STATEMENT

Institutional review board approval and informed consent was obtained from all participants at each of the individuals research centers.

ORCID

Viktorija Smith MS https://orcid.org/0009-0008-7202-7532
Christina B. Young PhD https://orcid.org/0000-0001-9535-2137

REFERENCES

1. Parsons MW, Gardner MM, Sherman JC, et al. Feasibility and acceptance of direct-to-home tele-neuropsychology services during the COVID-19 pandemic. *J Int Neuropsychol Soc*. 2022;28:210-215. doi:10.1017/S1355617721000436
2. Hewitt KC, Block C, Bellone JA, et al. Diverse experiences and approaches to tele neuropsychology: commentary and reflections over the past year of COVID-19. *Clin Neuropsychol*. 2022;36:790-805. doi:10.1080/13854046.2022.2027022
3. Marra DE, Hamlet KM, Bauer RM, Bowers D. Validity of teleneuropsychology for older adults in response to COVID-19: a systematic and critical review. *Clin Neuropsychol*. 2020;34:1411-1452. doi:10.1080/13854046.2020.1769192
4. Hunter MB, Jenkins N, Dolan C, Pullen H, Ritchie C, Muniz-Terrera G. Reliability of telephone and videoconference methods of cognitive assessment in older adults with and without dementia. *J Alzheimer's Disease*. 2021;81:1625-1647. doi:10.3233/JAD-210088
5. Carlew AR, Fatima H, Livingstone JR, et al. Cognitive assessment via telephone: a scoping review of instruments. *Arch Clin Neuropsychol*. 2020;35:1215-1233. doi:10.1093/arclin/acaa096
6. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155-163. doi:10.1016/j.jcm.2016.02.012
7. Chapman JE, Cadilhac DA, Gardner B, Ponsford J, Bhalla R, Stolwyk RJ. Comparing face-to-face and videoconference completion of the Montreal Cognitive Assessment (MoCA) in community-based survivors of stroke. *J Telemed Telecare*. 2021;27:484-492. doi:10.1177/1357633X19890788
8. Katz MJ, Wang C, Nester CO, et al. T-MoCA: a valid phone screen for cognitive impairment in diverse community samples. *Alzheimer Dement*. 2021;13:e12144. doi:10.1002/dad2.12144
9. Stillerova T, Liddle J, Gustafsson L, Lamont R, Silburn P. could everyday technology improve access to assessments? a pilot study on the feasibility of screening cognition in people with Parkinson's disease using the Montreal cognitive assessment via internet videoconferencing. *Aus Occup Ther J*. 2016;63:373-380. doi:10.1111/1440-1630.12288
10. Abdolahi A, Bull MT, Darwin KC, et al. A feasibility study of conducting the montreal cognitive assessment remotely in individuals with movement disorders. *Health Informatics J*. 2016;22:304-311. doi:10.1177/1460458214556373
11. Lindauer A, Seelye A, Lyons B, et al. Dementia care comes home: patient and caregiver assessment via telemedicine. *Gerontologist*. 2017;57:e85-e93. doi:10.1093/geront/gnw206
12. Gierzynski TF, Gregoire A, Reader JM, et al. Evaluation of the Uniform data set version 3 teleneuropsychological measures. *J Int Neuropsychol Soc*. 2023:1-11. doi:10.1017/S1355617723000383
13. Wadsworth HE, Galusha-Glasscock JM, Womack KB, et al. Remote neuropsychological assessment in rural American Indians with and without cognitive impairment. *Arch Clin Neuropsychol*. 2016;31:420-425. doi:10.1093/arclin/acw030
14. Cullum CM, Weiner MF, Gehrmann HR, Hynan LS. Feasibility of telecognitive assessment in dementia. *Assessment*. 2006;13:385-390. doi:10.1177/1073191106289065
15. Cullum CM, Hynan LS, Grosch M, Parikh M, Weiner MF. Teleneuropsychology: evidence for video teleconference-based neuropsychological assessment. *J Int Neuropsychol Soc*. 2014;20:1028-1033. doi:10.1017/S1355617714000873
16. Wadsworth HE, Dhima K, Womack KB, et al. Validity of teleneuropsychological assessment in older patients with cognitive disorders. *Arch Clin Neuropsychol*. 2018;33:1040-1045. doi:10.1093/arclin/acx140
17. Vahia IV, Ng B, Camacho A, et al. Telepsychiatry for neurocognitive testing in older rural latino adults. *Am J Geriatr Psychiatry*. 2015;23:666-670. doi:10.1016/j.jagp.2014.08.006
18. Vestal L, Smith-Olinde L, Hicks G, Hutton T, Hart J. Efficacy of language assessment in Alzheimer's disease: comparing in-person examination and telemedicine. *Clin Interv Aging*. 2006;1:467-471.
19. Thompson NR, Prince MJ, Macdonald A, Sham PC. Reliability of a telephone-administered cognitive test battery (TACT) between telephone and face-to-face administration. *Int J Methods in Psychiatr Res*. 2001;10:22-28. doi:10.1002/mpr.97
20. Mitsis EM, Jacobs D, Luo X, Andrews H, Andrews K, Sano M. Evaluating cognition in an elderly cohort via telephone assessment. *Int J Geriatr Psychiatry*. 2010;25:531-539. doi:10.1002/gps.2373

21. Fox-Fuller JT, Rizer S, Andersen SL, Sunderaraman P. Survey findings about the experiences, challenges, and practical advice/solutions regarding teleneuropsychological assessment in adults. *Arch Clin Neuropsychol*. 2022;37:274-291. doi:10.1093/arclin/acab076
22. Weintraub S, Besser L, Dodge HH, et al. Version 3 of the Alzheimer disease centers' neuropsychological test battery in the uniform data set (UDS). *Alzheimer Dis Assoc Disord*. 2018;32:10-17. doi:10.1097/WAD.0000000000000223
23. ADC Clinical Task Force (John C. Morris, MD, Chair) and the National Alzheimer's Coordinating Center (Walter A. Kukull, PhD, Director). T-cog Neuropsychological Battery Form (Form C2T) 2020.
24. National Alzheimer's Coordinating Center 2022. https://naccdata.org/
25. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatrics Soc*. 2005;53:695-699. doi:10.1111/j.1532-5415.2005.53221.x
26. Craft S, Newcomer J, Kanne S, et al. Memory improvement following induced hyperinsulinemia in Alzheimer's disease. *Neurobiol Aging*. 1996;17:123-130. doi:10.1016/0197-4580(95)02002-0
27. Craft S, Asthana S, Schellenberg G, et al. Insulin effects on glucose metabolism, memory, and plasma amyloid precursor protein in Alzheimer's disease differ according to Apolipoprotein-E genotype. *Ann N Y Acad Sci*. 2000;903:222-228. doi:10.1111/j.1749-6632.2000.tb06371.x
28. Morris JC, Heyman A, Mohs RC, et al. The consortium to establish a registry for Alzheimer's disease (CERAD). Part I. clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*. 1989;39:1159-1165. doi:10.1212/wnl.39.9.1159
29. Goldberg TE, Harvey PD, Wesnes KA, Snyder PJ, Schneider LS. Practice effects due to serial cognitive assessment: implications for preclinical Alzheimer's disease randomized controlled trials. *Alzheimers Dement*. 2015;1:103-111. doi:10.1016/j.dadm.2014.11.003
30. Hausknecht JP, Halpert JA, Di Paolo NT, Moriarty Gerrard MO. Retesting in selection: a meta-analysis of coaching and practice effects for tests of cognitive ability. *J Appl Psychol*. 2007;92:373-385. doi:10.1037/0021-9010.92.2.373
31. Samaroo A, Amariglio RE, Burnham S, et al. Diminished Learning Over Repeated Exposures (LORE) in preclinical Alzheimer's disease. *Alzheimer Dement*. 2020;12:e12132. doi:10.1002/dad2.12132
32. Jutten RJ, Rentz DM, Fu JF, et al. Monthly at-home computerized cognitive testing to detect diminished practice effects in preclinical Alzheimer's disease. *Front Aging Neurosci*. 2022;13.
33. Calamia M, Markon K, Tranel D. Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *Clin Neuropsychol*. 2012;26:543-570. doi:10.1080/13854046.2012.680913
34. Young CB, Mormino EC, Poston KL, et al. Computerized cognitive practice effects in relation to amyloid and tau in preclinical Alzheimer's disease: results from a multi-site cohort. *Alzheimer Dement*. 2023;15:e12414. doi:10.1002/dad2.12414
35. Bender AR, Ganguli A, Meiring M, Hampstead BM, Driver CC. Dynamic modeling of practice effects across the healthy aging-Alzheimer's disease continuum. *Front Aging Neurosci*. 2022;14.
36. Brooks BL, Sherman EMS, Iverson GL, Slick DJ, Strauss E. Psychometric foundations for the interpretation of neuropsychological test results. In: Schoenberg MR, Scott JG, eds. *The Little Black Book of Neuropsychology: A Syndrome-Based Approach*. Springer US; 2011:893-922. doi:10.1007/978-0-387-76978-3_31
37. Axelrod BN, Wall JR. Expectancy of impaired neuropsychological test scores in a non-clinical sample. *Int J Neurosci*. 2007;117:1591-1602. doi:10.1080/00207450600941189
38. Miller AL, Unsworth N. Attending to encode: the role of consistency and intensity of attention in learning ability. *J Mem Lang*. 2021;121:104276. doi:10.1016/j.jml.2021.104276
39. John SE, Ritter A, Wong C, Parks CM. The roles of executive functioning, simple attention, and medial temporal lobes in early learning, late learning, and delayed recall. *Aging Neuropsychol Cogn*. 2022;29:400-417. doi:10.1080/13825585.2021.2016583
40. Berns S, Davis-Conway S, Jaeger J. Telephone administration of neuropsychological tests can facilitate studies in schizophrenia. *Schizophr Res*. 2004;71:505-506. doi:10.1016/j.schres.2004.03.023
41. Reverberi C, Cherubini P, Baldinelli S, Luzzi S. Semantic fluency: cognitive basis and diagnostic performance in focal dementias and Alzheimer's disease. *Cortex*. 2014;54:150-164. doi:10.1016/j.cortex.2014.02.006
42. Price SE, Kinsella GJ, Ong B, et al. Semantic verbal fluency strategies in amnestic mild cognitive impairment. *Neuropsychology*. 2012;26:490-497. doi:10.1037/a0028567
43. Unsworth N, Spillers GJ, Brewer GA. Variation in verbal fluency: a latent variable analysis of clustering, switching, and overall performance. *Q J Exp Psychol*. 2011;64:447-466. doi:10.1080/17470218.2010.505292
44. Nygård L, Starkhammar S. Telephone use among noninstitutionalized persons with dementia living alone: mapping out difficulties and response strategies. *Scand J Caring Sci*. 2003;17:239-249. doi:10.1046/j.1471-6712.2003.00177.x
45. Choi NG, DiNitto DM. Internet use among older adults: association with health needs, psychological capital, and social capital. *J Med Internet Res*. 2013;15:e2333. doi:10.2196/jmir.2333
46. Udeh-Momoh CT, de Jager-Loots CA, Price G, Middleton LT. Transition from physical to virtual visit format for a longitudinal brain aging study, in response to the Covid-19 pandemic. Operationalizing adaptive methods and challenges. *Alzheimers Dement*. 2020;6:e12055. doi:10.1002/trc2.12055
47. Gavett BE, Ashendorf L, Gurnani AS. Reliable change on neuropsychological tests in the uniform data set. *J Int Neuropsychol Soc*. 2015;21:558-567. doi:10.1017/S1355617715000582

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Smith V, Younes K, Poston KL, Mormino EC, Young CB. Reliability of remote National Alzheimer's Coordinating Center Uniform Data Set data. *Alzheimer's Dement*. 2023;15:e12498. https://doi.org/10.1002/dad2.12498