Published in final edited form as:

Int J Geriatr Psychiatry. 2017 September ; 32(9): 991–999. doi:10.1002/gps.4558.

The SAGES Telephone Neuropsychological Battery: Correlation with In-Person Measures

Lydia Bunker, BA[1,*], **Tammy T. Hshieh, MD, MPH**[2,3,*], **Bonnie Wong, PhD**[4,5], **Eva M. Schmitt, PhD**[3], **Thomas Travison, PhD**[3], **Jacqueline Yee, BA**[3], **Kerry Palihnich, BA**[6], **Eran Metzger, MD**[3,5], **Tamara G. Fong, MD, PhD**[4,5,±], and **Sharon K. Inouye, MD, MPH**[3,6,±]

[1]New York Medical College, Valhalla, New York

[2]Division of Aging, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston

[3]Aging Brain Center, Institute for Aging Research, Hebrew SeniorLife, Boston

[4]Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA

[5]Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA

[6]Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston. MA

Abstract

Objective—Neuropsychological test batteries are administered in-person to assess cognitive function in both clinical and research settings. However, in-person administration holds a number of logistical challenges that makes it difficult to use in large or remote populations, or for multiple serial assessments over time. The purpose of this descriptive study was to determine whether a telephone-administered neuropsychological test battery correlated well with in-person testing.

Methods—50 English-speaking patients without dementia, over 70 years old and part of a cohort of patients in a prospective cohort study examining cognitive outcomes following elective surgery, were enrolled in this study. Five well-validated neuropsychological tests were administered by telephone to each participant by a trained interviewer within 2–4 weeks of the most recent in-person interview. Tests included the Hopkins Verbal Learning Test-Revised, Digit Span, Category Fluency, Phonemic Fluency and Boston Naming Test. A General Cognitive Performance (GCP) composite score was calculated from individual subtest scores, as a Z-score.

Results—Mean age was 74.9 years (SD = 4.1), 66% female and 4% non-white. Mean and interquartile distributions of telephone scores were similar to in-person scores. Correlation analysis of test scores revealed significant correlations between telephone and in-person results for

Corresponding Author/Request for Reprints: Tammy T. Hshieh, M.D., Division of Aging, Brigham and Women's Hospital; One Brigham Circle, 3rd floor; Boston, MA 02115; thshieh@partners.org; Phone: 401-536-6075.

[*]Lydia Bunker and Tammy Hshieh contributed equally to the paper as co-first authors.

[±]Tamara Fong and Sharon Inouye contributed equally to the paper as co-senior authors.

each individual subtest, as well as for the overall composite score. A Bland-Altman plot revealed no bias or trends in scoring for either test administration type.

Conclusions—In this descriptive study, the telephone version of a neuropsychological test battery correlated well with the in-person version, and may provide a feasible supplement in clinical and research applications.

Keywords

Neuropsychological testing; test battery; telephone; cognition; cognitive testing

Introduction

Neuropsychological testing is used in both clinical and research settings to assess for cognitive impairment. Many different neuropsychological tests are utilized to evaluate specific cognitive domains. Several individual tests may be combined to create a neuropsychological test battery, which clinicians and researchers can utilize to assess and to track cognitive status over time. Typically, formalized neuropsychological testing is administered in person by a trained clinician, such as a neuropsychologist or physician and may take over two hours to complete. However, face-to-face evaluation can be challenging, due to limitations such as staff time and costs, geographic distance, and patient constraints including time, morbidity, and disability.

While assessments for clinical dementia (major neurocognitive disorder) and neurological examinations must necessarily be administered in person, there is mounting evidence supporting the added value of remote assessment of some aspects of patient performance, such as neuropsychological testing by telephone. Telephone-based cognitive test batteries are convenient for both participants and interviewers and may hold particular advantages for remote evaluation of patients such as in rural areas or for large-scale epidemiological studies to maximize retention, decrease time burden, and increase cost-effectiveness of such studies (Castanho et al., 2014). The availability of comparable face-to-face and telephone versions of a neuropsychological test battery would enable clinicians and researchers to obtain reliable and consistent follow-up data on patients, enhancing the feasibility of tracking neuropsychological changes over time.

Previous studies have demonstrated that brief cognitive screening tests may be modified for ease of administration by telephone. Examples of screening tests modified for telephone include the widely used Telephone Interview of Cognitive Status (TICS) (Brandt, 1988) which was derived from the Mini Mental State Examination (MMSE) (Folstein et al., 1975) and the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005) which was recently adapted for telephone application in the t-MoCA (Pendlebury et al., 2013). A few groups have previously developed and validated neuropsychological batteries (at least 25 minutes in length) for use on the telephone (Castanho et al., 2014). A recent review by Castanho et al. found 19 studies in which neuropsychological testing was validated for telephone administration. Only five of these – the Structured Telephone Interview for Dementia Assessment (6 subscales) (Go et al., 1997), Telephone Cognitive Assessment Battery (6 neuropsychological tests) (Debanne et al., 1997), Minnesota Cognitive Acuity

Screen (9 subtests) (Knopman et al., 2000), Brief Test of Adult Cognition by Telephone (6 subtests) (Tun and Lachman, 2006) and Cognitive Telephone Screening Instrument (6 subtests) (Kliegel et al., 2007) – assessed multiple cognitive domains with detailed testing. The current study extends previous work by testing a telephone battery with strong assessment of executive functioning, as well as with an existing well-tested composite measure (Jones et al., 2010, Gross et al., 2014). Thus, the aim of the current descriptive study was to assess the correlation between telephone-based and in-person, face-to-face versions of the neuropsychological test battery used for the Successful Aging after Elective Surgery (SAGES) study. We hypothesize that there will be high correlation between the two neuropsychological test batteries, so that the telephone and in-person assessments can be used as comparable methods in future research studies.

Methods

Study Design

The SAGES study is an ongoing prospective cohort study of 566 older adults undergoing major elective surgery. The study design and methods have been described in detail previously (Schmitt et al., 2012). In brief, eligible participants were age 70 years and older, English speaking, and scheduled to undergo elective surgery at two Harvard-affiliated academic medical centers, with an anticipated length of stay of at least 3 days. Surgical procedures, which had moderate to high risk of incident delirium, included total hip or knee replacement, lumbar, cervical, or sacral laminectomy, lower extremity arterial bypass surgery, open abdominal aortic aneurysm repair, and open or laparoscopic colectomy. Exclusion criteria included evidence of dementia, active delirium or hospitalization within 3 months, terminal condition, legal blindness or severe deafness, history of schizophrenia or psychosis, and history of alcohol abuse or withdrawal.

Sub-study Population

From September 16, 2013 – February 10, 2014, consecutive participants (n=95) undergoing follow-up in the SAGES study were offered enrollment into the present substudy. Inability to hear on the telephone was the only additional exclusion criterion; ten persons were excluded due to being hard of hearing. Fifty patients volunteered to participate in this sub-study and provided verbal consent to proceed with the telephone interview. The remaining 35 patients could not be reached to complete the interview during the pre-specified study period.

All study procedures were approved by the institutional review boards of Beth Israel Deaconess Medical Center and Brigham and Women's Hospital, the two study hospitals, and Hebrew SeniorLife, the study coordinating center, all located in Boston, Massachusetts.

Approach

Every 6 months following their elective surgery, SAGES subjects underwent the full neuropsychological test battery in person, as part of a longer 75 minute follow-up interview that also collected demographic, functional, mood, and health-related data. For the present sub-study, a 30 minute telephone neuropsychological test battery (described below) was administered to the volunteer sub-group, within 2–4 weeks of the in-person interview. This time frame was chosen to optimize correlation, and minimize intervening factors in the interim that could impact cognitive performance. The neuropsychological tests chosen all have modest-to-low test-retest issues over one year (Mitsis et al., 2010, Rankin et al., 2005). There were specific, standardized instructions for the telephone test battery, asking participants to be in a quiet environment away from others and to refrain from writing items down. Experienced interviewers who had undergone training and standardization with inter-rater reliability assessment completed both in-person and telephone batteries, and were kept strictly blinded to each other's results.

The individual neuropsychological subtests were administered in the same order as the in-person and telephone-based administrations. Because of logistic constraints and the need to complete the SAGES in-person assessment on a precise timeline, we were not able to randomize the order of test administration approaches (i.e., in-person vs. telephone). For the present study, the in-person assessment always occurred first.

Measures

The neuropsychological test battery used for the SAGES study was selected by the study investigators in consultation with neuropsychological experts, with particular focus on with assessment of domains vulnerable to delirium including executive function (Schmitt et al., 2012). The 45 minute in-person battery included eight standardized and widely used neuropsychological tests evaluating executive function, visuospatial function, attention, semantic memory, verbal episodic memory, confrontation naming, and language ability (Schmitt et al., 2012). From this larger neuropsychological battery, five tests (Appendix Table 1) including the Hopkins Verbal Learning Test-Revised (Brandt, 2001), Digit Span Test Forwards and Backwards (Wechsler, 1989), Verbal Fluency (Benton, 1968), Semantic Fluency (Benton, 1969) and a modified version of the Boston Naming Test (BNT) Short Form (15-items) (Goodglass, 1983) were selected for the 30 minute telephone battery based on feasibility of telephone administration. Three tests –Trail-making Tests A and B, Visual Search and Attention Test and RBANS (Repeatable Battery for the Assessment of Neuropsychological Status) Digit Symbol Substitution – require pen and paper testing and were not included since comparable telephone versions were not available. Individual test details, administration times and scoring criteria are shown in Appendix Table 1. For most of the tests, the telephone administration was virtually identical to the in-person interview. HVLT-R measured verbal episodic memory and asked the participant to remember a list of words read aloud to them. Three recall trials were administered, followed by a delayed recall trial and a recognition trial 20–25 minutes later. The Digit Span Test assessed attention and short-term memory. Participants were asked to repeat increasingly long sequences of digits, followed by a second trial in which they were asked to repeat digits in reverse order. Phonemic fluency examined executive function and semantic memory by asking participants to generate as many words as possible starting with a given letter (F, A, or S) within 60 seconds. The Categorical Fluency test evaluated executive function, semantic memory, and language skills. The participant generated as many words as possible in a specific semantic category, such as grocery store items within 60 seconds.

Only the BNT test required extensive modification for telephone administration. The original 60-item in-person BNT was designed to assess confrontation naming and language. We utilized a validated short version with 15 items. The participant was asked to identify an image of an object or animal presented on a flashcard. Using previously validated methodology to develop an auditory naming test that assesses vocabulary and confrontation naming (Hamberger and Seidel, 2003), the SAGES study team created the modified telephone version of BNT in collaboration with an experienced neuropsychologist (BW). For this subtest, the interviewer read a short sentence describing the object, then the participant was asked to name it. The interviewer was allowed to give a phonemic cue (the first phoneme of the word) if the participant was unable to identify the object, although a correct answer following this cue was only awarded a half point. The list of objects to be identified in the telephone version was identical to the in-person version, as was the order in which objects were presented.

The General Cognitive Performance (GCP) composite score was used as a summary neuropsychological measure. The methods to create this composite have been described in detail previously (Jones et al., 2010, Gross et al., 2014). First, scores for each individual test were stratified into deciles to create more comparable score distributions. Subsequently, parallel analysis and item response theory were used to create the weighted composite score, which is presented as a score scaled from 0–100.

Other study variables presented to characterize the cohort at baseline included demographics, the Geriatric Depression Scale, Charlson comorbidity index, modified Mini – Mental State (3MS) score, Wechsler Test of Adult Reading, and functional impairment by Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs) (Charlson et al., 1987, Katz, 1983, Wechsler, 1981).

Statistical Analyses

Baseline characteristics of the study sample are presented with standard descriptive statistics, including means, standard deviations, and proportions. The neuropsychological test scores obtained in-person and over the phone were described using standard statistics, including means, standard deviations, medians, 25–75% inter-quartile ranges, and percentages at the floor (lowest possible score) and ceiling (highest possible score) of the distribution of each score. Differences in scores by assessment method were evaluated by calculated mean differences in scores, and compared using the paired t-test statistic and associated 95% confidence intervals. Agreement between in-person and telephone test scores was estimated by the Pearson correlation coefficient (r). To evaluate for any systematic bias in the telephone-based administration, a Bland-Altman plot of the GCP composite scores was examined (Bland and Altman, 1999). All analyses were conducted using Stata MP Version 13.0. Statistical significance was assessed using a two-tailed alpha level of 0.05.

Results

Characteristics of the 50 participants are summarized in Table 1. They closely resembled the overall SAGES cohort, with a slightly lower mean age (75 versus 77 years), slightly higher female predominance (66% versus 58%), lower rate of nonwhite (4% versus 8%), and lower rate of any IADL dependency (18% versus 28%). The average time elapsed between the in-person and the telephone assessment was 15.5 days (SD = 6.7, IQR 10–20). Overall, the sample was highly educated (mean education 15 years), cognitively intact (mean 3MS score of 95) and functionally independent with only 8% impaired in any Activities of Daily Living (ADL) and 18% impaired in any Instrumental Activities of Daily Living (IADL). None of the participants had delirium at the time of the neuropsychological testing.

Table 2 presents descriptive results for the average and range of values for the in-person and telephone individual test scores and composite score. The score distributions were comparable for both types of test administration. With both approaches, participants demonstrated the strongest performance on HVLT-R Delayed Recall and the BNT. Twenty percent of both in-person and telephone participants received a perfect score on the Delayed Recall HVLT-R. For the BNT, 68% of in-person participants and 60% of telephone participants achieved a perfect score. Both of these tests have relatively narrow ranges and have been reported previously to demonstrate ceiling effects (Castanho et al., 2014). Floor effects were minimal for all of the tests administered both in-person or by telephone.

All comparisons between in-person and telephone mean scores were not significant. The GCP mean value was consistently higher by telephone (mean = 64, SD = 8.3) compared with the in-person administration (mean = 62; SD = 8.1). In addition, HVLT-R Total Recall (mean = 28, SD = 5.6 vs. mean = 27, SD = 5.8), Delayed Recall (mean = 10, SD = 1.3 vs. mean= 9, SD =2.5), Discrimination Index (mean = 10, SD = 1.3 vs. mean = 10, SD = 1.4), and Digit Span (mean =19, SD =4 vs. mean = 17, SD = 3.7) were generally higher on the telephone as compared to in-person administration. The mean scores for Semantic Fluency (FAS) test (mean =44, SD = 14.5 vs. mean = 45, SD = 13.8) and the BNT (mean =14, SD = 1.4 vs. mean = 14, SD = 1.7) were slightly lower when administered by telephone.

Comparisons of results from the in-person and telephone interviews are presented in Table 3. While all of the correlations except HVLT-R Retention Percentage are statistically significant, the strongest correlations (r > 0.80) were observed for the GCP composite, HVLT-R Total Recall, Verbal Fluency, and BNT, indicating substantial agreement between in-person and telephone scores. The correlation between HVLT-R Retention Percentage scores by telephone and in-person were not significant correlated. All other measures of HVLT-R were and they are arguably more clinically relevant scores than HVLT-R Retention Percentage. The HVLT-R Retention Percentage, Verbal Fluency, and BNT had negative mean difference scores, indicating that participants scored higher when the test was administered in-person compared with telephone. For all other tests, and the GCP composite, participants had positive mean difference scores, indicating higher scores on the telephone compared with in-person administration. The largest mean differences were observed for the GCP composite (2.35, 95% confidence interval 0.96–3.74), HVLT-R Total Recall (1.64, 95% CI 0.82, 2.46) and Digit Span (1.52, 95% CI 0.41, 2.63), all of which showed higher scores on the telephone administration. The smallest mean differences, ranging from −0.26 to 0.30 were demonstrated for the BNT, HVLT-R Delayed Recall and HVLT-R Discrimination Index.

Figure 1 shows the statistically significant correlation between GCP scores by in-person vs. telephone administration, with correlation coefficient, r = 0.82 (p < 0.001). The Bland-Altman plot displays the mean of our paired measurements (telephone vs. in-person) on the x-axis, and the absolute difference between the same two measurements on the y-axis. All but three paired data points lie within a 10-point range of one another, on a total grading scale of 0–100 for the GCP. Additionally, there is no apparent trend in the difference between paired data as the averages increase, nor does the scatter increase or decrease overall. The distribution appeared homogenous, thus, there is no evidence suggesting a systematic bias in the GCP scores between in-person and telephone administrations.

Discussion

Our results demonstrate that a telephone battery comprised of five subtests assessing attention, executive functioning, memory, and language abilities, shows strong correlation with the in-person battery, and may be a useful additional tool for assessing cognitive status. The telephone battery, which takes about 30 minutes to administer, is feasible and well accepted by study participants. Ratings of acceptability were based on feedback from the telephone interviewers, and feasibility was demonstrated based on the 100% completion rates of all participants. In general, participants performed comparably on the two modes of administration, with highest overall scores on the telephone administration which was administered second, possibly reflecting a learning effect. The composite score based on the telephone battery was highly correlated (r > 0.80, p < 0.001) with the in-person score. Thus, the telephone battery may provide a useful addition to the toolbox of neuropsychological test batteries for both clinical and research purposes.

Two previously published telephone neuropsychological batteries, comprised of the TICS and other individually validated subtests, demonstrated significantly similar mean scores on subtests; however, these batteries were examined in exclusively female populations (Rapp et al., 2012, Mitsis et al., 2010). A group from the Age-Related Eye Disease Study (AREDS) conducted a thorough validation of telephone and face-to-face batteries administered to 1,738 participants to assess subtest correlations (Rankin et al., 2005). However, the time elapsed between in-person and telephone interview ranged from 4.7 to 12 months. This potentially presents a substantial limitation as the physical and cognitive health of patients can change dramatically over the course of 4–12 months.

Our study is unique in that it examines the consistency of a detailed neuropsychological test battery comprised of nearly identical subtests, administered by telephone and face-to-face interviews (within an average interval of 15 days) in a population of older men and women with no pre-existing diagnosis of dementia or major neurocognitive disorder and therefore avoids some of the limitations seen in prior correlation studies between in-person and telephone administration. There are a number of other noteworthy strengths to our study, including the expertise of the well-trained interviewers, careful blinding of results, well-tested composite measure, and the existing rich and complete data collection on the SAGES participants which were leveraged for the present study. Despite these strengths, some limitations of this study are worthy of mention. First, while the telephone battery assesses a broad range of cognitive domains, it is more limited than the in-person battery since it excluded 3 tests assessing visuoperceptual and executive functioning. Second, the repeated administration of tests is a significant issue, and practice effects often cannot be avoided. Although randomizing the order of test administration can sometimes help minimize such effects, due to logistic considerations of the parent SAGES study, it was not possible to do so. Thus, there may have been a learning effect influencing results of the telephone administration, which was always subsequent to the in-person administration; this in turn may have reduced the observed strength of association between in-person and telephone testing. However, all the tests for our neuropsychological test battery have been previously shown to have modest-to-low test-retest issues (Mitsis et al., 2010, Rankin et al., 2005), which may mitigate this learning effect problem. Lastly, the modified telephone version of the BNT may have been slightly more difficult for the participants than picture identification. The interviewers also noted that some patients had difficulty understanding words such as “toothed instrument” (describing the object “saw”) and were therefore unable to properly identify the object. Overall, however, there was strong correlation between in-person and telephone BNT scores (Pearson correlation co-efficient 0.85 with CI 0.75–0.91).

Additional caveats to the study include the fact that there was less control over testing conditions when the battery was administered by telephone. Interviewers were not able to verify a quiet and calm environment or to ensure compliance with instructions about not writing down words or numbers. Despite exclusion of hearing impaired participants, it is also possible that the telephone made it difficult for participants to understand the interviewer’s instructions. This may have contributed, for example, to a lower mean score on the BNT administered by telephone. It is possible that the lower than expected (but still statistically significant) correlations for Digit Span and Category Fluency are due to the telephone administration, where test administrators have less control of the testing environment and scores may have been affected by distractions.

Furthermore, the volunteer participants in the study may not be representative of the general population and their sample size was small (n = 50). Only 4% of the sub-study population was non-white, most had some college education (mean 14.9 years) and all were undergoing elective surgery at tertiary care centers in New England. While the internal validity of our results should not be impacted, these results should be replicated in larger studies that include more diverse populations to assure generalizability. In addition, since this was intended to be a correlational study of different administration approaches, we did not validate the test scores against an external reference standard. This would be an important future step to validate the instrument. Finally, it is important to note that some of the neuropsychological tests used are proprietary and require a fee for use.

Conclusion

While requiring external validation, this descriptive study holds promise to provide additional tools for both clinical and research settings. Our telephone battery may be complementary to in-person testing for epidemiologic studies where multiple data points need to be collected. Our telephone battery is very feasible and cost-effective, allowing for longitudinal follow-up of participants in studies requiring repeated neuropsychological assessment. It also allows for accessing a potentially home-bound or rural-dwelling population for clinical evaluation who would not be available otherwise. In conclusion, the telephone neuropsychological test battery presented in this study shows strong correlation with in-person administration. While future validation is needed in a larger study population, this battery holds promise to serve as a useful tool for the growing field of telemedicine or other applications where obtaining in-person neuropsychological data from patients may not be feasible.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Acknowledgements and Sources of Funding Support:

This study was supported in part by Grants No. P01AG031720 (SKI), K07AG041835 (SKI), and R01AG044518 (SKI) from the National Institute on Aging. Ms. Bunker was supported by Grant No. T35AG038027 from the National Institute on Aging, John A. Hartford Foundation, and MetLife Foundation. Dr. Hshieh was supported by Grant No. T32AG000158 from the National Institute on Aging. Dr. Inouye holds the Milton and Shirley F. Levy Family Chair. The funding sources had no role in the study design, collection/analysis/interpretation of data, writing of the report, or decision to submit for publication. All co-authors had no potential financial or personal conflicts of interest.

Note: Some tests in the neuropsychological battery are proprietary and require licensing fees. All tests used in this paper were licensed and used with permission.

References

BENTON AL. DeveloVerbal Fluence-Neurosensory Center Comprehensive Examination for Aphasia. Neuropsychologica. 1968

BENTON AL. Development of a multilingual aphasia battery. Progress and problems. J Neurol Sci. 1969; 9:39–48. [PubMed: 5820858]

BLAND JM, ALTMAN DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999; 8:135–60. [PubMed: 10501650]

BRANDT, J., BENEDICT, RHB. Psychological Assessment Resources. Odessa, FL: 2001. Hopkins Verbal Learning Test - Revised: Professional Manual.

BRANDT J, SPENCER M, FOLSTEIN M. The telephone interview for cognitive status. Neuropsychiatry, Neuropsychological Behavior and Neurology. 1988; 1:111–117.

CASTANHO TC, AMORIM L, ZIHL J, PALHA JA, SOUSA N, SANTOS NC. Telephone-based screening tools for mild cognitive impairment and dementia in aging studies: a review of validated instruments. Front Aging Neurosci. 2014; 6:16. [PubMed: 24611046]

CHARLSON ME, POMPEI P, ALES KL, MACKENZIE CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987; 40:373–83. [PubMed: 3558716]

DEBANNE SM, PATTERSON MB, DICK R, RIEDEL TM, SCHNELL A, ROWLAND DY. Validation of a Telephone Cognitive Assessment Battery. J Am Geriatr Soc. 1997; 45:1352–9. [PubMed: 9361661]

FOLSTEIN MF, FOLSTEIN SE, MCHUGH PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975; 12:189–98. [PubMed: 1202204]

GO RC, DUKE LW, HARRELL LE, CODY H, BASSETT SS, FOLSTEIN MF, ALBERT MS, FOSTER JL, SHARROW NA, BLACKER D. Development and validation of a Structured Telephone Interview for Dementia Assessment (STIDA): the NIMH Genetics Initiative. J Geriatr Psychiatry Neurol. 1997; 10:161–7. [PubMed: 9453683]

GOODGLASS, H., KAPLAN, E. The assessment of aphasia and related disorders. Malvern, PA: Lea & Febiger; 1983.

GROSS AL, JONES RN, FONG TG, TOMMET D, INOUYE SK. Calibration and validation of an innovative approach for estimating general cognitive performance. Neuroepidemiology. 2014; 42:144–53. [PubMed: 24481241]

HAMBERGER MJ, SEIDEL WT. Auditory and visual naming tests: normative and patient data for accuracy, response time, and tip-of-the-tongue. J Int Neuropsychol Soc. 2003; 9:479–89. [PubMed: 12666772]

JONES RN, RUDOLPH JL, INOUYE SK, YANG FM, FONG TG, MILBERG WP, TOMMET D, METZGER ED, CUPPLES LA, MARCANTONIO ER. Development of a unidimensional composite measure of neuropsychological functioning in older cardiac surgery patients with good measurement precision. J Clin Exp Neuropsychol. 2010; 32:1041–9. [PubMed: 20446144]

KATZ S. Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living. J Am Geriatr Soc. 1983; 31:721–7. [PubMed: 6418786]

KLIEGEL M, MARTIN M, JAGER T. Development and validation of the Cognitive Telephone Screening Instrument (COGTEL) for the assessment of cognitive function across adulthood. J Psychol. 2007; 141:147–70. [PubMed: 17479585]

KNOPMAN DS, KNUDSON D, YOES ME, WEISS DJ. Development and standardization of a new telephonic cognitive screening test: the Minnesota Cognitive Acuity Screen (MCAS). Neuropsychiatry Neuropsychol Behav Neurol. 2000; 13:286–96. [PubMed: 11186165]

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; 25:531–9. [PubMed: 19697298]

NASREDDINE ZS, PHILLIPS NA, BEDIRIAN V, CHARBONNEAU S, WHITEHEAD V, COLLIN I, CUMMINGS JL, CHERTKOW H. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005; 53:695–9. [PubMed: 15817019]

PENDLEBURY ST, WELCH SJ, CUTHBERTSON FC, MARIZ J, MEHTA Z, ROTHWELL PM. Telephone assessment of cognition after transient ischemic attack and stroke: modified telephone interview of cognitive status and telephone Montreal Cognitive Assessment versus face-to-face Montreal Cognitive Assessment and neuropsychological battery. Stroke. 2013; 44:227–9. [PubMed: 23138443]

RANKIN MW, CLEMONS TE, MCBEE WL. Correlation analysis of the inclinic and telephone batteries from the AREDS cognitive function ancillary study. AREDS Report No. 15. Ophthalmic Epidemiol. 2005; 12:271–7. [PubMed: 16033748]

RAPP SR, LEGAULT C, ESPELAND MA, RESNICK SM, HOGAN PE, COKER LH, DAILEY M, SHUMAKER SA, GROUP CATS. Validation of a cognitive assessment battery administered over the telephone. J Am Geriatr Soc. 2012; 60:1616–23. [PubMed: 22985137]

SCHMITT EM, MARCANTONIO ER, ALSOP DC, JONES RN, ROGERS SO JR, FONG TG, METZGER E, INOUYE SK, GROUP SS. Novel risk markers and long-term outcomes of delirium: the successful aging after elective surgery (SAGES) study design and methods. J Am Med Dir Assoc. 2012; 13:818, e1–10.

TUN PA, LACHMAN ME. Telephone assessment of cognitive function in adulthood: the Brief Test of Adult Cognition by Telephone. Age Ageing. 2006; 35:629–32. [PubMed: 16943264]

WECHSLER, D. Manual for the Wechsler Adult Intelligence Scale - Revised. New York: Psychological Corporation; 1981.

WECHSLER, D. Wechsler Adult Intelligence Scale-Revised Manual. New York: Psychological Corporation, A Harcourt Assessment Company; 1989.

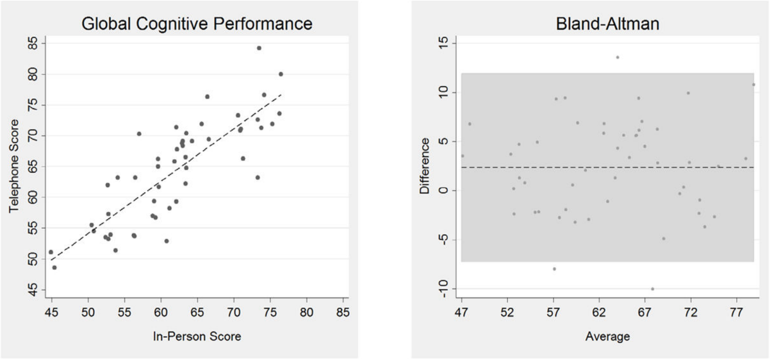

Figure 1. Correlation and Agreement of Telephone and In-Person GCP Composite Scores

The linear correlation between telephone and in-person General Cognitive Performance (GCP) composite scores is shown, r = 0.82 ($p \leq 0.001$). Bland-Altman plot shows the average of the telephone and in-person GCP value, in comparison to the difference between telephone and in-person mean scores. The data show no apparent trends as the GCP score increases, nor an overall increase or decrease in scatter, suggesting that there is no systematic bias in the correlation of the two modes of administration.

Table 1

Baseline Characteristics

Characteristic	Telephone (N = 50)	Overall Cohort (N= 566)
Age, mean(SD)	74.9 (4.1)	76.7 (5.2)
Female, *n (%)*	33 (66)	330 (58)
Hispanic or Non-White, *n (%)*	2 (4)	43 (8)
Years of Education, *mean (SD)*	14.9 (2.5)	14.9 (2.9)
Married, *n (%)*	29 (58)	335 (59)
Geriatric Depression Scale (GDS) Score, *mean (SD)*[†]	2.2 (2.5)	2.5 (2.5)
GDS > 5, *n (%)*	4 (8)	69 (12)
Charlson Comorbidity Index (CCI), *mean (SD)*[‡]	0.9 (1.3)	1.0 (1.3)
CCI ≥ 2, *n (%)*	12 (26)	167 (30)
Any Activity of Daily Living (ADL) Dependency, *n (%)*[β]	4 (8)	42 (7)
Any Instrumental ADL (IADL) Dependency, *n (%)*[δ]	9 (18)	157 (28)
Modified Mini Mental State (3MS) Score, *mean (SD)*[φ]	95.3 (3.6)	93.4 (5.4)
3MS < 85, *n (%)*	1 (2)	39 (6.9)
Wechsler Test of Adult Reading (WTAR), *mean (SD)*[χ]	37.7 (10.3)	37.7 (9.9)

[†] 15-point scale. Score > 5 suggestive of clinical depression. Score ≥ 10 highly predictive of clinical depression.

[‡] Comorbidities assigned a point value of 1–6, based on severity. Higher scores, particularly ≥ 2, suggest lower 1- or 2- year survival rate.

[β] Defined as requiring assistance to complete ≥ 1 personal care activity (bathing, dressing, toilet use, transferring, urine/bowel continence, eating).

[δ] Defined as requiring assistance to complete ≥ 1 day-to-day activity required to live independently (telephone use, shopping, meal preparation, housekeeping, laundry, transportation, medication management, finance management).

[φ] Expanded Mini-Mental Status Exam, testing 4 additional cognitive domains: long-term memory, abstract thinking, categorical fluency, and delayed recall. Maximum score 100, higher score indicates better performance. Score < 85 suggestive of cognitive impairment. Score <77 highly predictive of cognitive impairment.

[χ] Assesses peak lifetime intelligence as a function of correctly pronounced words. Maximum score 50, higher score indicates better performance.

Table 2

In-Person and Telephone Neuropsychological Battery Test Scores

TEST	**In-Person**				**Telephone**			
	Mean (SD)	***Median [IQI]***	***# at Floor n (%)***	***# at Ceiling n (%)***	***Mean (SD)***	***Median [IQI]***	***# at Floor n (%)***	***# at Ceiling n (%)***
GCP	62 (8.1)	62 [56–67]	1 (2)	1 (2)	64 (8.3)	65 [57–71]	1 (2)	1 (2)
HVLT-R Total Recall	27 (5.8)	28 [24–31]	1 (2)	1 (2)	28 (5.6)	30 [24–32]	2 (4)	4 (8)
Delayed Recall	9 (2.5)	10 [8–11]	2 (4)	10 (20)	10 (2.2)	10 [8–11]	2 (4)	10 (20)
Discrimination Index	10 (1.4)	10 [9–11]	2 (4)	6 (18)	10 (1.3)	10 [9–11]	2 (4)	9 (18)
Digit Span Test	17 (3.7)	16 [14–20]	2 (4)	3 (6)	19 (4.0)	16 [16–21]	1 (2)	1 (2)
Phonemic Fluency	45 (13.8)	47 [37–55]	1 (2)	1 (2)	44 (14.5)	46 [34–54]	2 (4)	1 (2)
Category Fluency	24 (5.9)	23 [20–29]	1 (2)	1 (2)	25 (6.3)	25 [20–29]	1 (2)	1 (2)
Boston Naming Test	14 (1.7)	15 [14–15]	1 (2)	34 (68)	14 (1.6)	15 [14–15]	1 (2)	30 (60)

Note: The mean and median scores for the General Cognitive Performance (GCP)and each individual test are shown, with interquartile intervals [IQI]. The number of participants (n) who achieved the highest test score is displayed as # at Floor, n (%) and # at Ceiling, n (%).

Table 3

In-Person vs. Telephone Correlations and Paired Tests

Test	Pearson Correlation Coefficient (95% CI)	Mean Difference (95% CI)
GCP	0.82 (0.71, 0.90) *	2.35 (0.96, 3.74)
HVLT-R Total Recall	0.87 (0.79, 0.93) *	1.64 (0.82, 2.46)
HVLT-R Delayed Recall	0.75 (0.60, 0.85) *	0.28 (−0.20, 0.76)
HVLT-R Discrimination Index	0.62 (0.41, 0.77) *	0.30 (−0.04, 0.64)
HVLT-R Retention Percentage	0.27 (−0.01, 0.51)	−1.37 (−6.15, 3.40)
Digit Span	0.50 (0.25, 0.68) *	1.52 (0.41, 2.63)
Verbal Fluency	0.92 (0.86, 0.95) *	−1.40 (−3.05, 0.25)
Category Fluency	0.63 (0.43, 0.77) *	1.12 (−0.36, 2.60)
Boston Naming Test	0.85 (0.75, 0.91) *	−0.26 (−0.52, −0.01)

CI= confidence interval; GCP= General Cognitive Performance; HVLT= Hopkins Verbal Learning Test

* p < 0.01