



Published in final edited form as:

Int J Geriatr Psychiatry. 2010 May ; 25(5): 531–539. doi:10.1002/gps.2373.

Evaluating cognition in an elderly cohort via telephone assessment

Effie M. Mitsis^{1,2,†}, Diane Jacobs³, Xiaodong Luo^{1,2}, Howard Andrews⁴, Karen Andrews⁴, and Mary Sano^{1,2}

¹Alzheimer's Disease Research Center, Department of Psychiatry, Mount Sinai School of Medicine, New York, NY, USA

²Research & Development, James J. Peters VA Medical Center, Bronx, NY, USA

³Consulting Neuropsychologist, San Diego, CA, USA

⁴Mailman School of Public Health, Columbia University, New York, NY, USA

Abstract

Objective—Longitudinal neuropsychological assessment provides the opportunity to observe the earliest transition to cognitive impairment in healthy, elderly individuals. We examined the feasibility, and its comparability to in-person assessment, of a telephone administered battery of established neuropsychological measures of cognitive functioning in healthy, elderly women.

Methods—Fifty-four women (age = 79 ± 7.7 ; education = 15.4 ± 3.3) who were in self-reported good health were recruited from senior centers and other community sources. A two-way cross-over design was used in which participants were randomly assigned to receive either (1) in-person neuropsychological assessment followed by telephone assessment and (2) telephone assessment followed by in-person assessment, separated by approximately 4 weeks. Linear regression models were used to determine whether there were performance differences by method (in-person vs. telephone), and equivalence testing assessed comparability of the two methods.

Results—There were no statistically significant differences in performance between in-person and telephone assessments on most neuropsychological tests, with the exception of digit span backward, Oral Trail Making Test Part A, and delayed recall on the SRT, the latter likely related to non-comparable exposure (6-trials in-person vs. 3-trials telephone). Equivalence testing differences fell in the pre-specified clinical equivalence zones, providing evidence of comparability of the two methods.

Conclusions—These pilot data support telephone administration of a neuropsychological battery that yields comparable performance to in-person assessment with respect to most instruments. Significant differences in scores on some measures suggest care should be taken in selecting specific measures used in a neuropsychological battery administered by telephone.

Keywords

telephone assessment; aging; cognition

Copyright © 2009 John Wiley & Sons, Ltd.

Correspondence to: E. M. Mitsis, PhD, effie.mitsis@mssm.edu.

[†]Assistant Professor.

Disclosure

The authors report no conflicts of interest.

Introduction

Longitudinal neuropsychological assessment, typically administered in-person, provides the opportunity to observe the earliest transition to cognitive impairment in healthy, elderly individuals and is a mainstay in the routine monitoring of dementia progression in mild cognitive impairment (MCI) and Alzheimer's disease (AD). Generally, individuals who are followed longitudinally are enrolled in clinical research studies where they are administered a lengthy battery of neuropsychological tests. Comprehensive batteries of neuropsychological tests administered in-person have their advantages and disadvantages. The advantages include reliability and diagnostic accuracy and the disadvantages include the requirement of hours of testing and the effort associated with transportation to health centers for seniors who may have mobility and/or finance difficulties. Therefore, for elderly individuals, lengthy testing and the effort and expense to travel to sometimes distant centers is burdensome. Thus, retention of an aging cohort for in-person assessment is challenging, making telephone assessment an attractive alternative. Several studies have proposed using telephone assessment as an alternative to in-person assessment, though few have systematically compared the two methods.

The available cognitive assessments that have been used specifically for telephone administration, such as the Telephone Interview for Cognitive Status (TICS) (Brandt *et al.*, 1988) and the TICS-modified (Plassman *et al.*, 1994) are useful as screeners, especially for large-scale epidemiological studies (van Uffelen *et al.*, 2007), in identifying individuals with cognitive impairment (i.e., see Lines *et al.*, 2003; Crooks *et al.*, 2005; Debling *et al.*, 2005) who are referred for a more thorough subsequent in-person assessment. However, these screeners are primarily tests of mental status, which are limited in their ability to detect subtle or mild cognitive dysfunction (Tombaugh and McIntyre, 1992; Crooks *et al.*, 2005). The few studies to have addressed the issue of telephone assessment using established neuropsychological tests sensitive to brain dysfunction (Taichman *et al.*, 2005; Christie *et al.*, 2006; Unverzagt *et al.*, 2007) have shown that telephone assessment captures cognitive test scores and self-reported mood and memory ratings as reliably and accurately as the traditional in-person method (Christie *et al.*, 2006; Unverzagt *et al.*, 2007), although others have observed learning effects for verbal memory measures (Taichman *et al.*, 2005) when in-person assessment is followed by telephone assessment.

This pilot study used a two-way, cross-over design to examine comparability of a telephone versus an in-person administered battery of neuropsychological measures to assess cognitive functioning in healthy, elderly women. Participants were randomly assigned to receive either of the following two sequences of administration: (1) in-person assessment followed by telephone assessment and (2) telephone assessment followed by in-person assessment. Feasibility and reliability outcomes from the pilot study were used to guide the selection of tests for a large scale observational follow-up study of women who had completed their participation in the longitudinal, multi-site, primary prevention trial of AD using hormonal replacement therapy (HRT) in postmenopausal women (PREPARE) (Sano *et al.*, 2008) that was being transitioned from in-person assessment conducted at the various sites nationwide, to telephone follow up to be conducted from our main, centralized sites (MSSM and JJP VAMC). The PREPARE study began in 1998 as a clinical trial designed to determine if HRT delays AD or memory loss. Women who were in good general health and had intact memory functioning (mean age and education at baseline 72.8 and 14.2 years, respectively) were randomized to HRT or placebo. Primary outcomes were incident AD and memory decline on neuropsychological testing. In response to the Women's Health Initiative (WHI) May 2002 report of increased incidence of heart disease, stroke, pulmonary embolism, and breast cancer among women randomized to HRT, the PREPARE study underwent extensive modifications, including discontinuation of study medication. The study continued to follow

these participants, as an observational trial, for a total of 5 years blind to the original medication assignment (Sano *et al.*, 2008). The 5-year observational PREPARE study has now ended across all sites.

The primary hypothesis in this study was that telephone assessment and in-person assessment are comparable methods. If our hypothesis was not borne out, the findings would nevertheless provide effect size information that would allow us to adjust the types of tests to be used in our PREPARE (Sano *et al.*, 2008) telephone follow-up project and second, it would provide information for a larger, independent study to examine these two methods of administration.

Methods

Subjects

Women, 65 years of age or older were recruited from senior centers and other community sources within New York City. Inclusion criteria were designed to identify participants who were similar to the original PREPARE cohort (Sano *et al.*, 2008). Subjects were included if they had self-reported good health, without self-reported major neurological and psychiatric illnesses, were English-speaking, and were willing to be assessed twice: once in-person and once by telephone. Subjects were excluded if they were deemed to be medically unstable, if their responses to the medical questionnaire (see Clinical Ratings Section) raised the possibility of neurological or psychiatric illness, and if they had any self-reported hearing difficulty beyond that associated with normal aging (i.e., had been seen by a doctor for hearing loss), wore hearing aids, or had hearing aids on order from their physician. A total of 54 subjects were randomized to receive either telephone or in-person assessment first. Eighty-three percent were Caucasian. There were no differences in age or education among those who received in-person assessment first ($N = 28$) versus telephone assessment first ($N = 26$) (Table 1).

Procedures

The two assessments were separated by approximately 4 weeks. For the telephone assessment, subjects were instructed to turn off radio, TV, or cell phones and were requested not to write anything down to assist their recall. In-person assessments were conducted in a quiet room at the community center, in the participants' home, or in the offices of the Alzheimer's Disease Research Center (ADRC) at MSSM. All participants signed informed consent approved by the MSSM and JJP VAMC Institutional Review Boards. If a subject was randomized to receive telephone assessment at Time 1, a copy of the consent form, along with a self-addressed and stamped envelope for expediency and courtesy, was mailed to the participant to be signed prior to participation. An appointment was scheduled immediately upon receipt of the signed consent.

Neuropsychological test battery

All neuropsychological testing was conducted by a single, trained psychometrist under the supervision of a licensed PhD neuropsychologist (EMM), who double-scored all assessments to ensure accuracy. All tests were scored according to published guidelines and were chosen on the basis of their established sensitivity in detecting impairment in normal aging and in MCI and AD patients. The battery was designed to be brief (approximately 45 min) yet comprehensive. The selection of specific measures was influenced by both our desire to maintain continuity with measures currently used in our PREPARE study (Sano *et al.*, 2008) and our desire to incorporate additional measures used in national multi-center trials of AD and MCI (Morris *et al.*, 2006).

The following tests were included for both telephone and in-person assessments: Mini Mental Status Examination (MMSE) Orientation (10-items) (Folstein *et al.*, 1975), Buschke's Selective Reminding Test (SRT) (Buschke, 1975), Wechsler Adult Intelligence Scale, 3rd edition (WAIS-III) Digit Span subtest, Wechsler Memory Scale, 3rd edition (WMS-III) Logical Memory I & II (Anna Thompson story only), Letter Fluency (CFL), Animal Fluency, WORLD spelled backward, and Oral Trail Making Test, Parts A and B (Table 2). At the end of the in-person assessment session, the remaining items on the MMSE were administered so that when combined with Orientation, a global measure of mental status could be derived for each of the participants in this pilot study (Folstein *et al.*, 1975). For the additional MMSE items, WORLD spelled backward was substituted with counting backward by seven. Overall, the telephone assessment battery incorporated widely used measures from multi-center studies (Morris *et al.*, 2006) in addition to the well validated measures used in our PREPARE study (Sano *et al.*, 2008).

The SRT has extensive normative data from an ethnically diverse sample (Stricks *et al.*, 1998); has been used in large, community-based cohort studies (Stern *et al.*, 1992); and has been used with great success in the PREPARE study (Sano *et al.*, 2008; Wang *et al.*, 2008). Here, we cross-validated the 6-trial version of the SRT administered in-person with a 3-trial version administered over the phone. The decision to abbreviate the SRT was made due to concern regarding the feasibility of conducting the full 6-trial version over the telephone. While length of time was of some concern (approximately 6–8 min longer), the primary concern was that the lack of face-to-face contact with an examiner would make it easier for participants to terminate the testing session prematurely by hanging up the telephone, should they become frustrated by the multiple learning trials of the entire SRT. Furthermore, repeated trials may have increased the possibility of participants writing down the words. Moreover, one of the goals of conducting this study was to examine the comparability of the 3- and 6-trial versions of the SRT, consistent with our intent to administer a telephone battery that would be both comprehensive yet brief in the follow up of the PREPARE cohort (Sano *et al.*, 2008).

Clinical ratings

Clinical ratings administered during both assessments were the 5-item Geriatric Depression Scale (Sheikh and Yesavage, 1986), the 8-item Activities of Daily Living Questionnaire from the Alzheimer's Disease Cooperative Studies (ADCS) (Galasko *et al.*, 2006), and a brief Health questionnaire requiring a yes/no response to a variety of medical, neurological, and psychiatric disorders. Subjects were required to report any hospitalizations or significant psychosocial stressors (requiring a yes/no response) within the last 1 year (Table 1).

Study design

To compare in-person versus telephone administration, a two-way cross-over design was used. Cross-over designs use within-subject variation to provide a more powerful statistical test. Test administration was counterbalanced with half of the subjects randomly assigned to receive telephone assessment first followed by in-person and the other half to receive in-person first followed by telephone assessment.

Statistical analysis

Chi-square tests and *t*-tests were used to compare group differences in characteristics and clinical ratings. The primary analysis consisted of two steps. For step 1, two-sample *t*-tests were used to compare the data of the two groups (in-person first vs. telephone first) on the initial administration of the battery, a method comparable to a 2-group parallel design. For equivalence testing, the indifference zone for each neuropsychological test was defined as the interval $(-SD/2, SD/2)$, where SD was the standard deviation of the in-person scores

from the initial assessments. Differences that fall within the specified indifference zone are considered clinically trivial and of no practical consequences (Equivalence Testing: http://www.graphpad.com/library/BiostatsSpecial/article_182.htm). The significance level of the equivalence test was considered to be <0.05 if the 90% confidence interval (CI) of the mean difference between telephone scores and in-person scores fell completely within the defined indifference zone. The results of this robust and conservative analysis (summarized in Table 2), minimizes power by using only data from the first administration of the tests. To increase power, step 2 permitted the use of all the data, accounting for the two-way cross over design. Difference scores were calculated for each cognitive measure by subtracting the score obtained on the 2nd administration from the score obtained on the 1st administration. These difference scores were used as dependent variables in linear regression models examining the effect of method of administration (i.e., in-person vs. telephone) on each measure (adjusted and unadjusted for age and education). A statistically significant result reflected a significant difference between telephone and in-person scores. Equivalence testing was used to determine whether the 90% CI of the difference of a difference score falls into the indifference zone.

The above procedure was carried out for all of the neuropsychological tests, with the exception of SRT delayed recall, due to the differential learning trials of the test (6-trials for in-person vs. 3-trials for telephone). To address this study design related carry-over effect, the sum of the first three trials of SRT immediate recall was calculated from the in-person administration and was compared to the sum from the 3-trial telephone assessment from the initial visits for each method of administration. SRT delayed recall was a comparison of total delayed recall score (i.e., 6-trial recall vs. 3-trial recall).

Results

There was no difference in age ($t = -1.15$; $p = 0.256$) or education ($t = 0.05$; $p = 0.960$) between women who were randomized to receive in-person assessment first versus telephone assessment first, nor were there any differences on the MMSE, GDS, and ADL scores as a function of method of administration (Table 1). Subjects were oriented, not depressed, and had no difficulties with ADL. The most frequently cited medical issue was related to falls and head injury, followed by myocardial infarction, depression and chronic obstructive pulmonary disease (Table 1). Women who received in-person assessment first reported a greater number of falls and myocardial infarction as compared to women who were first assessed by telephone (Table 1), which did not affect outcome in any analyses (data not shown). There were no other significant differences by method of administration on self-report ratings.

Data based upon step 1 of the primary analysis of the first administration of the assessments (either telephone or in-person) are presented in Table 2. There were no statistically significant differences on all 10 non-SRT tests between telephone and in-person assessment. Of these 10 tests, only two (LM I and LM II) were statistically equivalent, which is not surprising since this type of analysis lacks power as described above. Table 3 depicts the results of step 2 of the primary analysis. There were no significant differences by method of administration on most non-SRT neuropsychological tests, with the exception of DSB and Oral TMT Part A. Subjects performed statistically better on DSB during the telephone assessment as compared to the in-person assessment ($t = -2.09$, $p = 0.046$) and had better performance on Oral TMT Part A during the in-person assessment as compared to their performance during the telephone assessment ($t = -2.42$, $p = 0.021$). There was a trend for better performance on Letter Fluency during the in-person assessment ($t = 2.02$, $p = 0.058$). In this more powerful analysis, 7 of the 10 non-SRT tests were statistically equivalent (Table 3). Non-significant findings in equivalence testing for Orientation and WORLD spelled

backward (Tables 2 and 3) was due to the narrow indifference zones for these two tests as a result of a restricted range of obtained scores (9–10 for Orientation and 4–5 for WORLD spelled backward). SRT analysis was not done because of differences in methodology.

Discussion

We examined the comparability of in-person versus telephone assessment of cognition in an elderly, female cohort. These pilot data suggest that most of the neuropsychological tests yielded comparable scores whether administered in-person or by telephone, but some tests may result in differentially maximized performance across methods (Table 3) suggesting that care should be taken in selecting the specific measures that are used in a neuropsychological battery administered by telephone.

The SRT was abbreviated for telephone administration. Not surprisingly, subjects were able to recall more words following a delay when they had six learning trials as compared to three learning trials. There was no difference between the two methods when the sums of the first three trials were compared ($t = -0.23$, $p = 0.817$). It is noteworthy that delayed recall for prose (Logical Memory) did not differ by method of administration, suggesting that when acquisition conditions are held constant, in-person and telephone delayed verbal recall are comparable. The decision to administer six learning trials for the in-person SRT was made to maintain consistency with the battery used with participants in the PREPARE study. However, to reduce the burden placed on participants with regard to time and effort associated with follow-up telephone testing, the 3-trial version of the SRT will be used in the follow up by telephone of the nationwide PREPARE subjects (Sano *et al.*, 2008). The data from this pilot study will be used to develop an algorithm to estimate the 6-trial score using the 3-trial version of the SRT.

Our study extends previous work examining in-person versus telephone-based assessment using less sensitive screening measures such as the TICS (Brandt *et al.*, 1988; Plassman *et al.*, 1994). The neuropsychological measures used in this pilot study are comparable to those used in a recent study investigating in-person and telephone-based assessments in a younger cohort of women (Unverzagt *et al.*, 2007). Specifically, our findings are similar to those of Unverzagt *et al.* (2007), who adapted measures of memory, attention, information processing speed, verbal fluency, and self-report measures of mood and memory for telephone administration in breast cancer survivors and healthy controls (Unverzagt *et al.*, 2007). Their findings in well-educated women with an average age of 59 years demonstrated that telephone assessment captured cognitive test scores and self-reported mood and memory ratings as reliably and accurately as the traditional in-person method. Similarly, Taichman *et al.* (2005) administered a brief neurocognitive battery over the telephone in a sample of somewhat younger individuals (49.7 ± 13.9 years of age) with pulmonary arterial hypertension (Taichman *et al.*, 2005). Their findings resulted in intraclass coefficients between 0.5 and 0.8 for tests that assessed a variety of cognitive domains ($p < 0.05$ for each domain). Learning effects were notable only for verbal memory measures. However, the Taichman *et al.* study did not randomize the subjects to receive telephone or in-person assessment first. All subjects were tested in-person first, with subsequent telephone assessment 2.5 months later. While we found differences in delayed recall as a result of learning effects, that is, having a greater number of trials on the SRT during in-person assessment as compared to telephone assessment, we found no differences when we compared the scores from the first three trials of the SRT between both methods of administration and no differences in delayed recall when subjects receive comparable exposure to the task (LM I and II).

There were some notable limitations to this study. First, the subjects in this pilot project were recruited based upon their self-report of good physical, neurological, and mental health. Subjects who endorsed a significant medical problem, hearing loss, were depressed, or were taking medications deemed to impact upon cognitive performance were excluded. Furthermore, our findings should be interpreted in light of the fact that our subject sample included a highly educated group of relatively healthy, older women. Application of telephone assessment to men or the general population may produce slightly different results than those obtained here. The degree to which external factors (i.e., television, radio in the home, interruptions) and internal factors (i.e., hearing loss, medical illness) impact upon an individual's performance during telephone assessment cannot be determined based upon these findings. We attempted to reduce any type of interference during the telephone assessment by requesting at the start of the session that each subject turn off all media, ignore call-waiting, and not have any writing utensils nearby that could potentially be used to jot down words or numbers.

Another limitation is related to the study design. A two group, randomized, four condition design (see Unverzagt *et al.*, 2007), would have eliminated problems associated with practice effects as a consequence of randomization to a two-group design in which each subject either had telephone assessment first or in-person assessment first. However, the two-way, cross-over design used allows for participants to serve as their own control, thus increasing power by reducing between-subject error. In addition, all the tests were identical with the exception of the SRT. Comparison of the first three learning trials of the 6-trial SRT to the 3-trial telephone administered version of the SRT was not significant ($t = -0.23$, $p = 0.817$) suggesting that it is unlikely that the type of design used here biased our findings.

Telephone-based neuropsychological assessment has an advantage to in-person assessment by reducing the effort and cost associated with traveling to and from medical centers for elderly individuals who may be in poor health and have mobility and/or finance problems. In view of our findings that telephone assessment is generally comparable to in-person assessment, these results suggest telephone assessment may be a useful, cost-containing alternative for epidemiological and clinical research studies. There are some disadvantages to telephone assessment such as lack of visual stimuli and motor tasks, the elimination of the examiner's ability to directly observe behavior, and the lack of a controlled environment (i.e., quiet room, no interruptions) as found with in-person assessment. Nevertheless, the benefit of telephone assessment far outweighs the disadvantages for many elderly individuals who would otherwise not be evaluated nor followed regularly, in the monitoring of cognitive decline.

We conclude that telephone assessment may be a useful, cost-effective alternative to in-person assessments for large scale epidemiological studies and may be effective in reducing attrition that often results in longitudinal clinical research studies of cognitive decline and dementia in an aging cohort, by reducing the burden associated with frequent follow up, in-person testing.

Acknowledgments

The authors wish to thank the directors of the community centers in the New York metropolitan area who provided us with the opportunity to recruit within their midst. This study was supported by R01 AG15922 (Dr. Sano).

References

- Brandt J, Spencer M, Folstein M. The telephone interview for cognitive status. *Neuropsychiatry Neuropsychol Behav Neurol.* 1988; 1:111–117.

- Buschke H. Method for evaluating language competence in neurological patients. *Trans Am Neurol Assoc.* 1975; 100:169–171. [PubMed: 1226600]
- Christie JD, Biester RC, Taichman DB, et al. Formation and validation of a telephone battery to assess cognitive function in acute respiratory distress syndrome survivors. *J Crit Care.* 2006; 21(2):125–132. [PubMed: 16769455]
- Crooks VC, Clark L, Petitti DB, Chui H, Chiu V. Validation of multistage telephone-based identification of cognitive impairment and dementia. *BMC Neurol.* 2005; 5(1):8. [PubMed: 15829005]
- Debling D, Amelang M, Hasselbach P, Sturmer T. Assessment of cognitive status in the elderly using telephone interviews. *Z Gerontol Geriatr.* 2005; 38(5):360–367. [PubMed: 16244822]
- 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(3):189–198. [PubMed: 1202204]
- Galasko D, Bennett DA, Sano M, et al. ADCS Prevention Instrument Project: assessment of instrumental activities of daily living for community-dwelling elderly individuals in dementia prevention clinical trials. *Alzheimer Dis Assoc Disord.* 2006; 20(4 Suppl 3):S152–S169. [PubMed: 17135809]
- Lines CR, McCarroll KA, Lipton RB, Block GA. Telephone screening for amnesic mild cognitive impairment. *Neurology.* 2003; 60(2):261–266. [PubMed: 12552041]
- Morris JC, Weintraub S, Chui HC, et al. The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. *Alzheimer Dis Assoc Disord.* 2006; 20(4): 210–216. [PubMed: 17132964]
- Plassman B, Newman T, Welsh K, Helms M, Breitner J. Properties of the telephone interview for cognitive status. *Neuropsychiatry Neuropsychol Behav Neurol.* 1994; 7:235–241.
- Sano M, Jacobs D, Andrews H, et al. A multi-center, randomized, double blind placebo-controlled trial of estrogens to prevent Alzheimer’s disease and loss of memory in women: design and baseline characteristics. *Clin Trials.* 2008; 5(5):523–533. [PubMed: 18827045]
- Sheikh, JI.; Yesavage, JA. *Geriatric Depression Scale: Recent Evidence and Development of a Shorter Version.* The Haworth Press; New York: 1986.
- Stern Y, Andrews H, Pittman J, et al. Diagnosis of dementia in a heterogeneous population. Development of a neuropsychological paradigm-based diagnosis of dementia and quantified correction for the effects of education. *Arch Neurol.* 1992; 49(5):453–460. [PubMed: 1580806]
- Stricks L, Pittman J, Jacobs DM, Sano M, Stern Y. Normative data for a brief neuropsychological battery administered to English- and Spanish-speaking community-dwelling elders. *J Int Neuropsychol Soc.* 1998; 4(4):311–318. [PubMed: 9656604]
- Taichman DB, Christie J, Biester R, et al. Validation of a brief telephone battery for neurocognitive assessment of patients with pulmonary arterial hypertension. *Respir Res.* 2005; 6:39. [PubMed: 15850486]
- Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. *J Am Geriatr Soc.* 1992; 40(9):922–935. [PubMed: 1512391]
- Unverzagt FW, Monahan PO, Moser LR, et al. The Indiana University telephone-based assessment of neuropsychological status: a new method for large scale neuropsychological assessment. *J Int Neuropsychol Soc.* 2007; 13(5):799–806. [PubMed: 17697411]
- van Uffelen JG, Chin APMJ, Klein M, van Mechelen W, Hopman-Rock M. Detection of memory impairment in the general population: screening by questionnaire and telephone compared to subsequent face-to-face assessment. *Int J Geriatr Psychiatry.* 2007; 22(3):203–210. [PubMed: 17044136]
- Wang S, Jacobs D, Andrews H, et al. Cardiovascular risk and memory in non-demented elderly women. *Neurobiol Aging.* 2008 Epub ahead of print.

Key Points

- Telephone administration captured cognitive test scores and self-reported mood and memory ratings as reliably and accurately as the traditional in-person method.
- Total test times were slightly shorter for the telephone approach, an unexpected yet favorable outcome.
- Telephone assessment may decrease socially desirable responding via perception of anonymity by the respondent.
- Advantage of telephone assessment is savings in time and effort associated with physical travel to and from the clinic and it allows for greater scheduling flexibility.

Table 1

Subject characteristics

	In-person first		Telephone first	t/Chi-square	p	In-person range		Telephone range
	N = 28	N = 26	N = 26			Chi-square	p	
Demographics								
Age	77.82	80.23	80.23	-1.15	0.256	65-93	66-97	66-97
Education	15.39	15.35	15.35	0.05	0.960	6-20	10-20	10-20
Ethnicity (number Caucasian)	21	24	24	3.18	0.075			
Clinical ratings								
MMSE	28	28	28	NA	NA	22-30	21-30	21-30
GDS	0.75	0.69	0.69	0.19	0.852	0-4	0-4	0-4
ADL	8.32	8.63	8.63	-0.77	0.443	6-13	6-12	6-12
Psychosocial stressors (% yes) ^a	75%	60%	60%	1.36	0.243			
Hospitalization (% yes) ^a	50%	54%	54%	0.08	0.778			
Co-morbidities								
Falls ^b	18	64	64	9	35	4.41	0.036	0.036
Head injury	4	14	14	4	15	0.00	0.970	0.970
Depression	3	11	11	3	11	0.86	0.353	0.353
Myocardial infarction	4	14	14	2	8	0.51	0.477	0.477
COPD	2	7	7	2	8	0.00	0.978	0.978

Note: MMSE, Mini Mental Status Examination; GDS, Geriatric Depression Scale; ADL, Activities of Daily Living; COPD, Chronic Obstructive Pulmonary Disease.

^aWithin the past 1 year.

^b $p < .05$.

Table 2
Mean neuropsychological test scores and effect sizes by method of first administration

	In-person		Telephone		Inp-Tel	SE	Effect size	p	90% Lower for diff.	90% Upper for diff.	In zone ^d	
	N	Mean	SD	N								Mean
MMSE Orientation	27	9.85	0.36	25	9.92	0.28	-0.07	0.09	-0.218	0.448	0.08	N
WORLD backwards	27	4.96	0.19	26	4.73	0.96	0.23	0.19	0.332	0.238	0.54	N
LM Immediate recall	28	13.43	3.70	26	13.54	2.60	-0.11	0.88	-0.034	0.899	1.33	Y
LM Delayed recall	28	11.79	4.28	25	11.64	3.64	0.15	1.10	0.038	0.894	1.95	Y
DSF	28	6.46	1.23	26	6.62	1.13	-0.15	0.32	-0.135	0.641	0.38	N
DSB	28	4.61	1.13	26	4.85	1.38	-0.24	0.34	-0.190	0.491	0.32	N
O-TMT Part A ^b	28	8.11	2.47	26	8.77	2.70	-0.66	0.70	-0.255	0.353	0.50	N
O-TMT Part B ^b	27	37.26	22.92	25	35.52	24.29	1.74	6.55	0.074	0.792	12.51	N
Letter Fluency	28	45.29	14.06	26	40.73	11.83	4.55	3.55	0.351	0.202	10.39	N
Animal naming	28	18.11	5.31	26	16.39	6.18	1.72	1.56	0.298	0.279	4.30	N
SRT												
Total recall (first 3 trials)	28	20.04	5.51	26	19.85	7.51	0.19	1.78	0.029	0.9168	3.12	N
Delayed Recall (6-Trials vs. 3-Trials)	28	6.75	2.91	25	4.4	3.27	2.35	0.85	0.759	0.0083	3.75	N

Note: LM I, Logical Memory Immediate Recall; LM II, Logical Memory Delayed Recall; DSF, Digit Span Forward; DSB, Digit Span Backward; O-TMT, Oral Trail Making Test; SRT, Selective Reminding Test.

^a Does the 90% confidence interval fall within the equivalence zone? Yes/No

^b Lower scores indicate better performance.

Table 3

Comparison of neuropsychological measures by method of administration

Test	SD of in-person at baseline	Unadjusted (inp-tel)					Adjusted (inp-tel) ^a						
		Diff.	SE	p	90% Lower for diff.	90% Upper for diff.	90% CI in equivalence zone?	Diff.	SE	p	90% Lower for diff.	90% Upper for diff.	90% CI in equivalence zone?
MMSE Orientation total	0.36	-0.20	0.16	0.227	-0.47	0.07	N	-0.23	0.17	0.172	-0.51	0.04	N
WORLD backwards	0.19	0.11	0.10	0.333	-0.06	0.28	N	0.08	0.10	0.428	-0.09	0.25	N
LM Immediate recall	3.70	-0.21	0.47	0.653	-0.98	0.56	Y	-0.14	0.47	0.763	-0.92	0.64	Y
LM Delayed recall	4.28	-0.71	0.48	0.128	-1.49	0.08	Y	-0.67	0.48	0.169	-1.46	0.12	Y
DSF	1.23	-0.15	0.16	0.370	-0.42	0.11	Y	-0.18	0.16	0.270	-0.44	0.09	Y
DSB *	1.13	-0.45	0.23	0.055	-0.83	-0.08	N	-0.48	0.23	0.046	-0.86	-0.10	N
O-TMT Part A *	2.47	-0.64	0.26	0.014	-1.06	-0.22	Y	-0.63	0.26	0.021	-1.06	-0.20	Y
O-TMT Part B	22.92	-1.02	2.68	0.717	-5.42	3.39	Y	-1.20	2.67	0.656	-5.59	3.20	Y
Letter Fluency	14.06	2.22	1.10	0.047	0.41	4.03	Y	2.20	1.13	0.058	0.34	4.06	Y
Animal naming	5.31	0.17	0.48	0.730	-0.63	0.96	Y	0.19	0.48	0.693	-0.60	0.98	Y
SRT													
Total recall in first 3 trials	5.51	-2.60	0.97	0.014	-4.20	-1.01	N	-2.74	0.97	0.007	-4.34	-1.14	N
Delayed Recall (6-Trials vs. 3-Trials)	2.91	0.22	0.41	0.597	-0.45	0.89	Y	0.23	0.42	0.582	-0.45	0.92	Y

Note: LM I, Logical Memory Immediate Recall; LM II, Logical Memory Delayed Recall; DSF, Digit Span Forward; DSB, Digit Span Backward; O-TMT, Oral Trail Making Test; SRT, Selective Reminding Test.

^aAdjusted for age and education.

* $p < 0.05$;

DSB = better performance during telephone first; O-TMT-A = better performance during in-person first.