



Published in final edited form as:

J Am Geriatr Soc. 2012 September ; 60(9): 1616–1623. doi:10.1111/j.1532-5415.2012.04111.x.

Validation of a cognitive assessment battery administered by telephone

Stephen R. Rapp, PhD^{a,b}, Claudine Legault, PhD^b, Mark A. Espeland, PhD^b, Susan M. Resnick, PhD^c, Patricia E. Hogan, MS^b, Laura H. Coker, PhD^b, Maggie Dailey, PhD^b, and Sally A. Shumaker, PhD^b for the CAT Study Group

^aDepartment of Psychiatry and Behavioral Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina, 27157

^bDivision of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, 27157

^cLaboratory of Behavioral Neuroscience, National Institute on Aging, NIH, Baltimore, MD

Abstract

Background—While the gold standard method of cognitive assessment is a face-to-face administration, telephone-based assessments offer several advantages if they demonstrate reliability and validity.

Design—Observational study; 110 participants randomly assigned to receive two administrations of the same cognitive test battery 6 months apart in one of four combinations (1st administration/2nd administration): telephone/telephone; telephone/face-to-face; face-to-face/telephone; or face-to-face/face-to-face.

Setting—Academic medical center

Participants—110 non-demented women between the ages of 65 and 90 years.

Measures—The battery included tests of attention, verbal learning and memory, verbal fluency, executive function, working memory and global cognitive functioning plus self-report measures of perceived memory problems, depressive symptoms, sleep disturbance and health-related quality of life. Test-retest reliability, concurrent validity, relative bias associated with telephone administration, and change scores were evaluated.

Results—There were no statistically significant differences in scores on any of the cognitive tests or questionnaires between randomly assigned modes of administration at baseline indicating equivalence across modes. There was no significant bias for tests or questionnaires administered by telephone ($p > 0.01$). Nor was there a difference in mean change scores between administration modes except for the Category Fluency ($p = 0.01$) and the California Verbal Learning Test long delay-free recall ($p < 0.01$). Mean test-retest coefficients for the battery were not significantly different across groups though individual test-retest correlation coefficients were generally higher within mode than across mode.

Conclusions—Telephone administration of cognitive tests and questionnaires to older women is both reliable and valid. Use of telephone batteries can substantially reduce the economic cost and

Address correspondence: Stephen R. Rapp, PhD, Department of Psychiatry and Behavioral Medicine, Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157; Phone: (336) 716-6995; Fax: (336) 716-0299; srapp@wakehealth.edu.

Trial Registration: Not applicable

burden of cognitive assessments and increase enrollment, retention and data completeness thereby improving study validity.

Keywords

cognition; assessment; telephone; validation; tests

Introduction

Performance on cognitive tests is critical for making diagnoses of cognitive disorders like dementia and mild cognitive impairment which require a documented decline in cognitive performance. Cognitive testing is also essential to define the magnitude and pattern of changes in cognition associated with normal aging, and in distinguishing clinically abnormal from normal cognitive changes. The 'gold standard' for assessing cognitive functioning is a face-to-face administration of a battery of standardized tests measuring key cognitive abilities such as attention, learning, memory, language, visuospatial abilities, and executive functions. Neuropsychological evaluations often include measures of mood and affect as well as information on well-being and quality of life.

The costs and feasibility of face-to-face cognitive testing are not trivial particularly in large samples from diverse geographic regions. A highly trained examiner is required to administer the tests, and patients or study participants must travel to the examiner, which can be particularly difficult for older and/or impaired individuals and sometimes requires in-home evaluations. Cost is magnified in the context of large, multi-site observational or clinical studies. Lowering the costs associated with face-to-face cognitive assessment could enable larger and more diverse samples to be tested as well as reduce burden substantially.

Telephone-administered cognitive assessments have emerged as a promising alternative to face-to-face assessments. The Telephone Interview for Cognitive Status (TICS)(1), for example, is a global measure of cognitive functioning that was modeled after the Mini Mental State Exam (2). With items assessing several cognitive abilities, the TICS yields a single score ranging from 0 to 41. Several studies have demonstrated the TICS' reliability and validity (3–16). Evans et al. reported a test-retest correlation of 0.70 for a modified version of the TICS (TICSm)(17) administered to 50 participants in the Nurse's Health Study twice with a 1 month interval (16). A significant limitation of the TICS, however, is that it provides only a coarse, global view of cognitive functioning and thus is not sufficient for making diagnostic determinations (14) or for measuring specific cognitive functions. Its performance may also be limited by study eligibility criteria (e.g., age, education, socio-economic status, health status) that contract its range (18).

Several attempts to develop a telephone-administered cognitive test battery that includes tests of different abilities have been made. Rankin et al. reported moderate to strong bivariate correlations ($\rho = 0.71-0.89$) between a battery of cognitive tests first administered face-to-face and then re-administered one year later over the telephone to 1,738 male and female participants 61 to 87 years old (mean = 75 years) in the Age-Related Eye Disease Study(19). The battery included the Logical Memory I & II and Digit Span-Backwards from the Wechsler Memory Scale- III (20;21) and a Category Fluency (animals) and Letter Fluency task (22).

Grodstein and colleagues administered a battery of 5 cognitive tests including the TICS, the East Boston Memory Test (23;24), the Category Fluency (Animals) task (25) and the Digit Span-Backwards test (20) by telephone to 19,319 women ages 70–79 years old who were participating in the Nurses' Health Study(26). The authors did not report the reliability or validity of the battery in this study, but they did examine its performance in a separate

sample of 61 participants over 70 years of age from the Religious Orders Study (16). Participants were administered the telephone cognitive battery and the same battery face-to-face. The authors reported a correlation of 0.81 between a single composite score for each battery but they did not report correlations for individual tests. They also found that scores below pre-determined cut-points on the TICS and on the EBMT were significantly related to independently determined diagnoses of dementia and to the apolipoprotein E (APOE) genotype, a risk factor for Alzheimer's dementia (27;28). Thus their battery demonstrated reliability and criterion-related validity.

Wilson and Bennett reported a significant relationship between the APOE genotype and several composite scores from a telephone administered battery that included Immediate and Delayed Recall of Story A from the Wechsler Memory Scale-Revised (20), Digit Span Forwards and Backwards (20), Digit Ordering (29), and Category Fluency (Animals, Vegetables) among 996 participants of the Religious Order Study (30). Subsequently, Wilson and colleagues compared face-to-face administration to telephone administration of their battery in 1,584 men and women in the Late-Onset Alzheimer's Disease Family Study and found that mode of administration only accounted for 2% of the variance in composite scores (31).

Together these studies provide support for the use of telephone-administered cognitive test batteries to measure cognitive performance and for classifying some individuals as impaired or unimpaired. Additional research is needed, however, to evaluate further the psychometric properties of telephone batteries and the direction of possible biases (e.g., over- or underestimates in relation to face-to-face administrations). Moreover the use of composite summary scores rather than individual test scores can mask variability in the reliability and validity of individual tests assessing domain-specific cognitive function.

The Cognitive Assessment by Telephone (CAT) study was designed to compare cross-sectionally and longitudinally the performance of a telephone-administered cognitive test battery to the same battery administered face-to-face in women ages 65 years and older with a range of cognitive levels. Unlike prior studies, CAT compared individual test performance and estimates of bias across administration modes and over time.

Methods

Women were recruited from the Piedmont region of North Carolina through an institutional research volunteer data base. Only women were considered because this study was intended to evaluate the test battery to be used in the Women's Health Initiative Memory Study (32) Extension Study. Prospective participants were screened in a face-to-face interview for significant hearing deficits and administered the Modified Mini Mental State Exam (3MSE) (33) to estimate their overall cognitive functioning. Women were excluded from the study if any of the following criteria were met: significant hearing problems, a previous clinical diagnosis of dementia, and a stroke within the past 6 months, inability to complete the 3MSE screening instrument, or scoring <72 on the 3MSE. Participants with 3MSE scores from high normal (95–100), mid-normal (88–94)- and mildly impaired (72–92) ranges were recruited and randomly assigned, with equal probabilities, to receive two administrations of the same cognitive test battery spaced six months apart in one of four orders (1st administration/2nd administration): telephone/telephone; telephone/face-to-face; face-to-face/telephone; or face-to-face/face-to-face. Those participants who were consented and randomized to receive face-to-face first proceeded directly to administration of the assessment battery. Those participants randomized to receive telephone first were instructed about a future telephone administration and released. All tests were administered by a trained and certified cognitive examiner blind to the group assignment. Six months after the

initial assessment each participant was re-administered the battery by telephone or face-to-face according to assignment. All participants provided written consent to participate and CAT was approved by the Institutional Review Board.

Cognitive Test Battery

The cognitive test battery was designed to assess key cognitive abilities including attention, concentration, verbal learning and memory, verbal fluency, working memory and executive function as well as global cognitive functioning.

Verbal learning and verbal memory were assessed with a modified version of the California Verbal Learning Test (34) in which the participant was read a list of 16 common words (List A) and then asked immediately to recall the words. This was repeated three times and the sum of all three trials provided a measure of new learning and short-term memory (CVLT Immediate Free Recall; possible range: 0–48). A new list of 16 different words (List B) was then administered once (List B; range 0–16), and the participant was asked to recall the List B words. The participant was next asked to recall the words from List A first without a semantic cue (Short-Delay Free Recall; range 0–16) and then with a semantic cue (Short-Delay Cued Recall; range 0–16). After a delay of approximately 20 minutes the participant was asked to recall the List A words again, first without a cue (Long Delay Free Recall; range: 0–16) and then following a cue (Long-Delay Cued Recall; range: 0–16). Last, the subject was asked to identify the List A words from a list that also included novel distracter words read by the examiner (Recognition; range: 0–44).

Verbal fluency was assessed with the Letter Fluency task and Category Fluency (Animals) task. For Letter Fluency the participant named as many words that began with a specified letter in 60 seconds. Three trials were administered using different letters (F, A, S) and the score was the sum of the three trials (25). For the Category Fluency-Animals task, participants were asked to name as many animals as possible in 60 seconds (25).

The Digit Span-Forward test of the Wechsler Memory Scale-III (21) measured attention and concentration and required the participant to repeat numbers presented in progressively longer span lengths (range: 0–14). The Digit Span-Backward task (21) required participants to recite the digits in reverse order and measured executive function/working memory (range: 0–14).

The TICS (1) was used as a measure of global cognitive functioning. It consists of items assessing orientation (12 points), attention/concentration (2 points), mental calculation (5 points), naming (4 points), repetition (2 points), social knowledge (4 points), and praxis (2 pts), for a total of 31 points. We omitted the TICS word list learning sub-task (10 points) to avoid proactive interference with the CVLT.

In addition to assessments of cognitive performance, participants completed questionnaires assessing perceived memory problems, depression, sleep disturbance and health-related quality of life over the phone and in person. Perceived memory problems were measured with 7 items used in the Nurse's Health Study (e.g., "Do you have much more trouble than usual remembering recent events?" "Do you have any difficulty in understanding or following spoken directions?") with a range of possible scores from 0–7 (35). Depressive symptoms severity was assessed with the 15-item Geriatric Depression Scale-Short Form (range: 0–15) (36;37). Difficulty sleeping was assessed with the 5-item Women's Health Initiative Insomnia Rating Scale (WHIRS; range: 0–20) (38). Health-related quality of life was assessed using the Medical Outcomes Study (MOS) Short Form 12 (39). Weighted summary scores were generated for physical (Physical Component Summary score; range:

0–100) and mental (Mental Component Summary score; range: 0–100) dimensions of quality of life.

Statistical Methods

Test-retest reliability was assessed with Pearson correlation coefficients for each test administered by the same mode over the 6-month interval. The concurrent validity of the telephone battery for detecting changes was assessed by fitting mixed effect general linear models to the data collected from the two time periods for both modes of administration. Estimates of the bias of telephone assessments relative to face-to-face assessments (i.e. telephone assessment minus face-to-face assessment) were expressed in standard deviation units, to allow comparisons among different tests. We examined cross-sectional means for each test and mean changes over time. In addition, we examined the correlation between test scores over time across the battery of cognitive tests, using analyses of variance to assess for differences in these longitudinal correlations depending on the sequence of administration. We also examined, using interaction terms, whether there was evidence that any relative biases might be related to characteristics of the women (i.e., age, baseline 3MS, education and race/ethnicity). Given the number of statistical tests and the exploratory nature of this study, a level of 0.01 was used for statistical significance.

Results

Participants

One-hundred and ten women were randomized into the study (Figure 1). Sixty-four participants scored > 95 (high normal range) on the 3MSE at screening, 31 participants scored between 88–95 (medium normal range) and 15 participants scored lower than 88, but not below 72 (low normal/mildly impaired range). Baseline data were collected on 105 participants and 91 completed follow-up (Table 1). Table 2 summarizes baseline demographic characteristics; randomization provided good balance among the assigned administration sequences for these variables. Overall, study participants were predominantly Non-Hispanic White, well-educated, and had a mean age of 72.4 years (SD = 5.7 yr.) with a range from 65 to 90 years. Women completing the study did not differ from non-completers in age but completers were more likely to be White ($p=0.03$), more educated ($p=0.03$), and score higher on the 3MSE ($p=0.06$) compared to non-completers.

Telephone versus Face-To-Face Administration

There were no statistically significant differences in scores on any of the cognitive tests or questionnaires between randomly assigned modes of administration at baseline (Table 3) indicating equivalence across modes. Table 4 presents estimates of the bias of telephone assessments relative to face-to-face assessments (i.e. telephone assessment minus face-to-face assessment). There was no statistically significant bias for any of the 12 cognitive tests or 5 questionnaires ($p>0.01$).

Table 4 also shows the mean change scores for each test and questionnaire administered 6-months apart using the same administration mode. There were no significant differences between modes except for Category Fluency and the CVLT Long Delay-Free Recall subtest where there was some evidence of differences associated with mode of administration over time. Category Fluency scores declined slightly over 6 months when administered face-to-face, but tended to increase when administered by telephone (nominal $p=0.01$) while CVLT-Long Delay-Free Recall scores changed little for the face-to-face group but improved for the group assessed by telephone. There were no significant differences in 6-month change scores between modes for the questionnaires. There were no significant interactions ($p>0.01$) between modes of administration and race/ethnicity, age, education or baseline

3MSE scores, except for the Recognition subtest of the CVLT. Non-Whites tended to show worse performance on the telephone versus face-to-face administration relative to Non-Hispanic Whites ($p=0.0002$): mean (SD) standardized relative bias is -1.24 (0.34) for Non-Whites and 0.16 (0.12) for Non-Hispanic Whites.

Lastly, Table 5 shows the bivariate correlations between Time 1 and Time 2 administrations for each test and questionnaire for each group. Significant moderate to high correlations were found for each cognitive test when the mode was the same across the two administrations but a slightly more variable pattern of individual test correlations emerged when different modes were used for the two assessments. Across the 12 cognitive tests, there was a significant difference in the mean correlation depending on the sequence of administration modes ($p=0.004$ based on two-way analyses of variance). However a relatively high mean (SD) correlation occurred when both administrations were by telephone, both administrations were by face-to-face, or when a telephone administration was followed by a face-to-face administration: 0.74 (0.09), 0.67 (0.12), and 0.66 (0.21), respectively. A slightly lower mean correlation occurred when the mode of administration changed from face-to-face to telephone: 0.50 (0.21) which trended toward but did not reach statistical significance difference ($p > 0.01$) from the repeated face-to-face correlation based on a Scheffe' multiple comparisons test.

Discussion

Telephone-administered cognitive test batteries, if valid and reliable, could make cognitive testing more useful in clinical and research applications by reducing the cost and participant burden associated with face-to-face administrations and allowing more diverse geographic representation. In this study, we compared both modes of administration cross-sectionally and longitudinally in post-menopausal women with cognitive functioning ranging from mildly impaired to high normal.

Our results show that the two modes of administration yield scores for tests and questionnaires that are equivalent. We found no significant differences in scores on any of the individual cognitive tests or questionnaires between face-to-face and telephone administration in this sample of older women at baseline. Nor did we detect evidence of significant bias in scores attributable to test telephone administration. Our results also show a pattern of comparable changes in test and questionnaire scores over 6 months for both modes. The patterns of test-retest correlations between tests administered 6 months apart suggest that maintaining the same mode of administration will yield more reliable scores than if modes vary between administrations. Moreover our four study groups were balanced with respect to age, gender, education and race/ethnicity, and they did not differ in the severity of perceived cognitive problems, sleep problems, depressive symptom severity or health-related quality of life.

Mode of administration was associated with a slight differential change in scores over time for only 2 of 12 cognitive tests and none of the 5 questionnaires. The slightly higher scores on the Category Fluency test and the CVLT Free Recall-Delay subtest when administered by telephone, while statistically significant, are not likely to be clinically meaningful. Moreover these significant comparisons may be spurious given the number of statistical tests we performed.

The pattern of correlations between test scores given 6 months apart (test-retest reliability) in each of our comparison groups suggests that maintaining the same mode of administration over time will yield more reliable scores. We found a lower mean correlation for face-to-face administration followed by telephone administration ($r = 0.50$) and there was greater

variability in individual correlations (range: $r = 0.09-0.82$). When a telephone administration was followed by a face-to-face administration the mean correlation of all tests in the battery between administrations was higher ($r = 0.66$) and variability less (range: $r=0.24-0.88$). The strongest and most consistent test-retest correlations were in the two single mode groups. This slightly poorer performance for the dual mode groups may signal some heterogeneity in intra-individual biases when administration modes are altered in this sequence, leading to lower correlations. It is unlikely that actual cognitive change occurring between the first and second administrations can explain the weaker correlations as the effect is not consistent across groups or tests. Thus, using the same assessment mode over repeated assessments yields more reliable scores which improves statistical power.

On the CVLT Recognition score, we found that non-Whites did more poorly when the CVLT was administered by telephone compared to face-to-face while Non-Hispanic Whites did slightly better. As we found no race/ethnicity effect on any other test of cognitive function including the other six CVLT parameters, this finding may be spurious. Additional research into possible racial or ethnic bias associated with telephone administration of tests and questionnaires, however, will help clarify this finding.

A telephone-based cognitive assessment approach similar to the one we examined is being used currently in the Women's Health Initiative Memory Study (40). A geographically diverse sample of 2,858 postmenopausal women between the ages of 75 and 93 are being evaluated annually with a telephone administered battery consisting of the TICSm (17), the East Boston Memory Test(23), the Oral Trail Making Test (41), Category Fluency-Animals, Digit Span-Forward and Backward (20), the Geriatric Depression Scale-Short Form(37), and the Women's Health Initiative Insomnia Rating Scale (38). Administration of the battery occurs annually from the WHIMS Coordinating Center at Wake Forest School of Medicine (Winston-Salem, NC) by 12 trained and certified examiners. To date over 4,000 telephone administrations have been administered. Several other research groups also have used telephone-administered cognitive batteries in studies with older adults (30;31;35;42). However we believe this is the first study to examine the reliability, validity and potential bias of individual cognitive tests and questionnaires administered by telephone to older adults over time.

Despite these advantages, telephone batteries have limitations. The range of cognitive functions that can be assessed over the telephone is more limited than in the conventional face-to-face administration. For example, tests involving motor performance cannot be administered. Hearing difficulties and distracting properties of the testing (e.g., home) environment cannot be as easily controlled as with in-person assessments. Also, the use of disallowed 'aids' such as note pads, clocks or calendars or receiving help from another person, cannot be absolutely prevented. In WHIMS we specifically instruct participants about these issues before beginning a testing session. Finally, whether administered by telephone or face-to-face, cognitive test batteries are insufficient by themselves for making diagnoses such as dementia and mild cognitive impairment.

Limitations notwithstanding, our data suggest that the results of cognitive testing and questionnaire administration over the phone to older women with normal to mildly impaired global cognitive functioning are comparable to face-to-face examinations when administered by trained examiners. Telephone administration of cognitive tests and questionnaires can be used to reach prospective study participants who might not otherwise participate, who live a great distance from the study center, or who cannot easily travel. In this way, telephone administration significantly reduces the participant burden compared to conventional face-to-face administration, thereby increasing enrollment, retention and data completeness which improve the validity of the study results. Our study demonstrates that telephone-

administered test batteries and questionnaires assessing cognition and other important participant-reported outcomes can be measured efficiently, reliably and validly.

Acknowledgments

The Cognitive Assessment by Telephone study was funded by the Intramural Research Program, National Institute on Aging, NIH, and the General Clinical Research Center of Wake Forest University Baptist Medical Center (M01-RR07122).

Reference List

1. Brandt J, Spencer M, Folstein MF. The telephone interview for cognitive status. *Neuropsychiatry Neuropsychol Behav Neurol*. 1988; 1:111–117.
2. Folstein MF, Folstein SE, McHugh PR. 'Mini Mental State': a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry*. 1975; 12:189–198.
3. Barber M, Stott DJ. Validity of the Telephone Interview for Cognitive Status (TICS) in post-stroke subjects. *Int J Geriatr Psychiatry*. 2004 Jan; 19(1):75–79. [PubMed: 14716702]
4. Brandt J, Spencer M, Folstein M. The Telephone Interview for Cognitive Status. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology*. 1988 Jun; 1(2):111–117.
5. Crooks VC, Petitti DB, Robins SB, et al. Cognitive domains associated with performance on the telephone interview for cognitive status-modified. *Am J Alzheimers Dis Other Demen*. 2006 Jan; 21(1):45–53. [PubMed: 16526589]
6. Crooks VC, Clark L, Petitti DB, et al. Validation of multi-stage telephone-based identification of cognitive impairment and dementia. *BMC Neurol*. 2005; 5(1):8. [PubMed: 15829005]
7. De Jager CA, Budge MM, Clarke R. Utility of TICS-M for the assessment of cognitive function in older adults. *Int J Geriatr Psychiatry*. 2003 Apr; 18(4):318–324. [PubMed: 12673608]
8. Debling D, Amelang M, Hasselbach P, et al. Assessment of cognitive status in the elderly using telephone interviews. *Z Gerontol Geriatr*. 2005 Oct; 38(5):360–367. [PubMed: 16244822]
9. Desmond DW, Tatemichi TK, Hanzawa L. The Telephone Interview for Cognitive Status (TICS): Reliability and validity in a stroke sample. *International Journal of Geriatric Psychiatry*. 1994 Oct; 9(10):803–807.
10. Ferrucci L, Del LI, Guralnik JM, et al. Is the telephone interview for cognitive status a valid alternative in persons who cannot be evaluated by the Mini Mental State Examination? *Aging (Milano)*. 1998 Aug; 10(4):332–338. [PubMed: 9825025]
11. Graff-Radford NR, Ferman TJ, Lucas JA, et al. A cost effective method of identifying and recruiting persons over 80 free of dementia or mild cognitive impairment. *Alzheimer Dis Assoc Disord*. 2006 Apr; 20(2):101–104. [PubMed: 16772745]
12. Hogervorst E, Bandelow S, Hart J Jr, et al. Telephone word-list recall tested in the rural aging and memory study: two parallel versions for the TICS-M. *Int J Geriatr Psychiatry*. 2004 Sep; 19(9): 875–880. [PubMed: 15352146]
13. Jarvenpaa T, Rinne JO, Raiha I, et al. Characteristics of two telephone screens for cognitive impairment. *Dement Geriatr Cogn Disord*. 2002; 13(3):149–155. [PubMed: 11893836]
14. Knopman DS, Roberts RO, Geda YE, et al. Validation of the telephone interview for cognitive status-modified in subjects with normal cognition, mild cognitive impairment, or dementia. *Neuroepidemiology*. 2010; 34(1):34–42. [PubMed: 19893327]
15. Welsh KA, Breitner JC, Magruder-Habib KM. Detection of dementia in the elderly using telephone screening of cognitive status. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology*. 1993 Apr; 6(2):103–110.
16. Evans DA, Grodstein F, Loewenstein D, et al. Reducing case ascertainment costs in U.S. population studies of Alzheimer's disease, dementia, and cognitive impairment--Part 2. *Alzheimer's and Dementia*. 2011 Jan; 7(1):110–123.
17. Welsh KA, Breitner J, Magruder-Habib KM. Detection of dementia in the elderly using the telephone interview for cognitive status. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology*. 1993; 6:103–110.

18. Espeland MA, Rapp SR, Katula JA, et al. Telephone interview for cognitive status (TICS) screening for clinical trials of physical activity and cognitive training: the seniors health and activity research program pilot (SHARP-P) study. *Int J Geriatr Psychiatry*. 2011 Feb; 26(2):135–143. [PubMed: 21229597]
19. Rankin MW, Clemons TE, McBee WL. Correlation analysis of the in-clinic and telephone batteries from the AREDS cognitive function ancillary study. AREDS Report No. 15. *Ophthalmic Epidemiol*. 2005 Aug; 12(4):271–277. [PubMed: 16033748]
20. Wechsler, D. Wechsler Memory Scale - Revised Manual. 1987.
21. Wechsler, D. The Wechsler Memory Scale-3rd Edition (WMS-III). Psychological Corporation, Harcourt, Inc; 1996. Ref Type: Generic
22. Rosen WG. Verbal fluency in aging and dementia. *J Clin Neuropsychology*. 1980; 2:135–146.
23. Gfeller JD, Horn GJ. The East Boston Memory Test: a clinical screening measure for memory impairment in the elderly. *J Clin Psychol*. 1996 Mar; 52(2):191–196. [PubMed: 8771447]
24. Albert M, Smith LA, Scherr PA, et al. Use of brief cognitive tests to identify individuals in the community with clinically diagnosed Alzheimer's disease. *Int J Neurosci*. 1991 Apr; 57(3–4):167–178. [PubMed: 1938160]
25. Lesak, MD. Neuropsychological assessment. 3. New York: Oxford University Press; 1997.
26. Lee S, Kawachi I, Berkman LF, et al. Education, other socioeconomic indicators, and cognitive function. *Am J Epidemiol*. 2003 Apr 15; 157(8):712–720. [PubMed: 12697575]
27. Evans DA, Grodstein F, Loewenstein D, et al. Reducing case-ascertainment costs in US population studies of Alzheimer's disease, dementia and cognitive impairment--Part 2. *Alzheimer's and Dementia*. 2011
28. Kang JH, Logroscino G, De VI, et al. Apolipoprotein E, cardiovascular disease and cognitive function in aging women. *Neurobiol Aging*. 2005 Apr; 26(4):475–484. [PubMed: 15653176]
29. Wilson RS, Beckett LA, Barnes LL, et al. Individual differences in rates of change in cognitive abilities of older persons. *Psychol Aging*. 2002 Jun; 17(2):179–193. [PubMed: 12061405]
30. Wilson RS, Bennett DA. Assessment of Cognitive Decline in Old Age with Brief Tests Amenable to Telephone Administration. *Neuroepidemiology*. 2005 Apr 25; 25(1):19–25. [PubMed: 15855801]
31. Wilson RS, Leurgans SE, Foroud TM, et al. Telephone assessment of cognitive function in the late-onset Alzheimer's disease family study. *Arch Neurol*. 2010 Jul; 67(7):855–861. [PubMed: 20625093]
32. Shumaker SA, Reboussin BA, Espeland MA, et al. The Women's Health Initiative Memory Study (WHIMS): a trial of the effect of estrogen therapy in preventing and slowing the progression of dementia. *Controlled Clin Trials*. 1998 Dec; 19(6):604–621. [PubMed: 9875839]
33. Teng EL, Chui HC. The Modified Mini-Mental State (3MS) examination. *J Clin Psychiatry*. 1987 Aug; 48(8):314–318. [PubMed: 3611032]
34. Delis, DC.; Kramer, J.; Kaplan, E. The California Verbal Learning Test. New York: The Psychological Corporation; 1987.
35. Grodstein F, Chen J, Pollen DA, et al. Postmenopausal hormone therapy and cognitive function in healthy older women. *J Am Geriatr Soc*. 2000 Jul; 48(7):746–752. [PubMed: 10894312]
36. Yesavage JA. Geriatric Depression Scale. *Psychopharm Bull*. 1988; 24:709–711.
37. Burke WJ, Roccaforte WH, Wengel SP. The short form of the Geriatric Depression Scale: a comparison with the 30-item form. *J Geriatr Psychiatry Neurol*. 1991 Jul; 4(3):173–178. [PubMed: 1953971]
38. Levine DW, Kripke DF, Kaplan RM, et al. Reliability and validity of the Women's Health Initiative Insomnia Rating Scale. *Psychol Assess*. 2003 Jun; 15(2):137–148. [PubMed: 12847774]
39. Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996 Mar; 34(3):220–233. [PubMed: 8628042]
40. Shumaker SA, Reboussin BA, Espeland MA, et al. The Women's Health Initiative Memory Study (WHIMS): a trial of the effect of estrogen therapy in preventing and slowing the progression of dementia. *Control Clin Trials*. 1998 Dec; 19(6):604–621. [PubMed: 9875839]

41. Ricker JH, Axelrod BN. Analysis of an oral paradigm for the Trail Making Test. *Assessment*. 1994 Mar; 1(1):47–51. [PubMed: 9463499]
42. Grodstein F, Chen J, Pollen DA, et al. Postmenopausal hormone therapy and cognitive function in healthy older women. *J Am Geriatr Soc*. 2000 Jul; 48(7):746–752. [PubMed: 10894312]

The CAT Study Group includes

Wake Forest University: Laura Coker, PhD; Maggie Dailey, PhD; Deborah M. Felton, BS; Mark A. Espeland, PhD; Darin Harris, BS; Patricia Hogan, MS; Ashley Lentz, BS; Claudine Legault, PhD; Carol Massa-Fanale, BS; Cynthia McQuellon, MA; Pamela D. Nance, BA; Debbie Pleasants, MEd; Cheryl Summerville, BS and Debbie M Booth, BA; Stephen Rapp, PhD; Sally A. Shumaker, PhD;

Conflict of Interest Checklist:

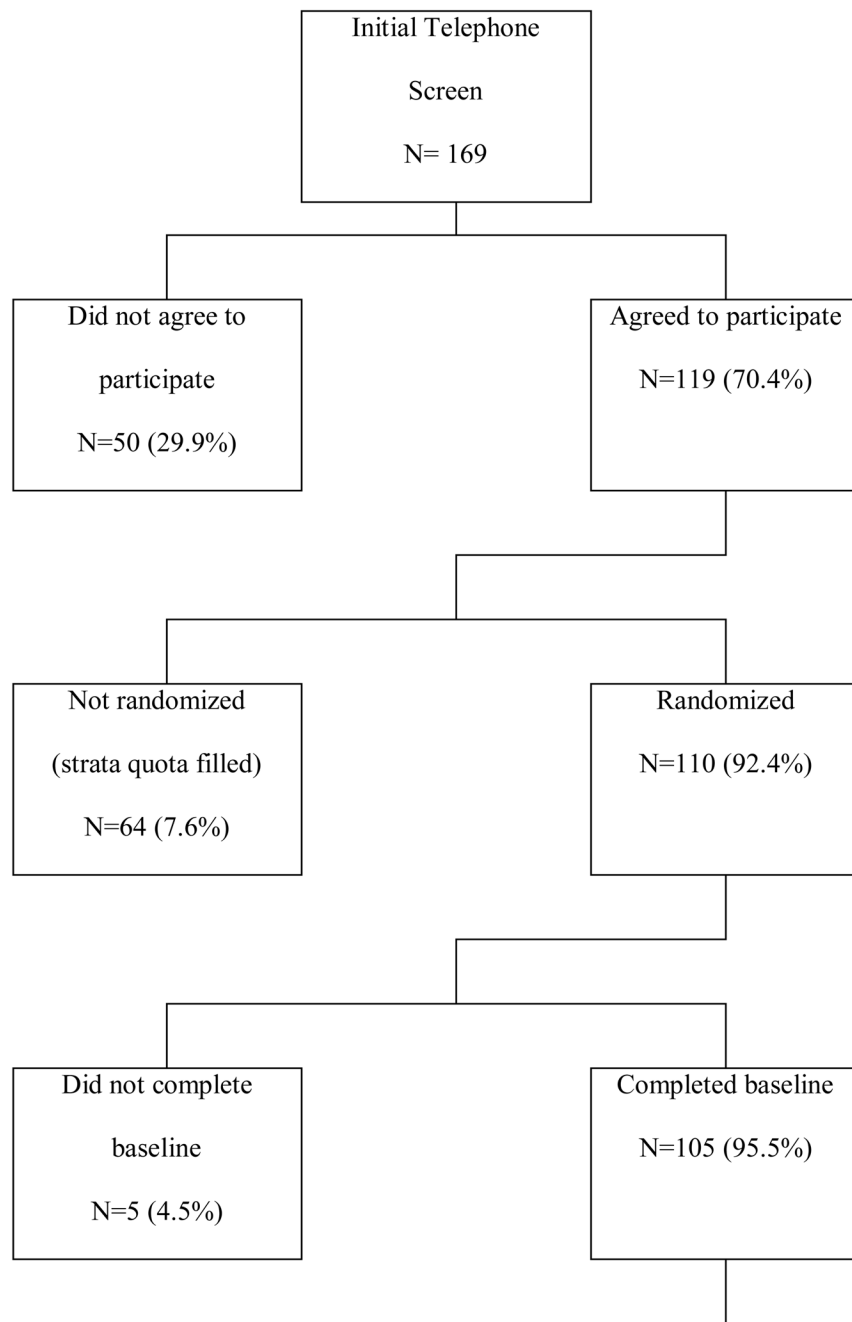
Elements of Financial/ Personal Conflicts	SR		CL		ME		SMR		PH		LC		MD		SS	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Employment or Affiliation		X		X		X		X		X		X		X		X
Grants/Funds		X		X		X		X		X		X		X		X
Honoraria		X		X		X		X		X		X		X		X
Speaker Forum		X		X		X		X		X		X		X		X
Consultant		X		X		X		X		X		X		X		X
Stocks		X		X		X		X		X		X		X		X
Royalties		X		X		X		X		X		X		X		X
Expert Testimony		X		X		X		X		X		X		X		X
Board Member		X		X		X		X		X		X		X		X
Patents		X		X		X		X		X		X		X		X
Personal Relationship		X		X		X		X		X		X		X		X

*Authors can be listed by abbreviations of their names

For “yes”, provide a brief explanation: _____

Author Contributions: SR, CL, ME, SMR, LC, MD, SS: concept and design, analysis and interpretation of data, and preparation of manuscript. PH: analysis and interpretation of data, and preparation of manuscript

Sponsor’s Role: Dr. Resnick (NIA Intramural Program) provided guidance in the design, methods and preparation of paper.



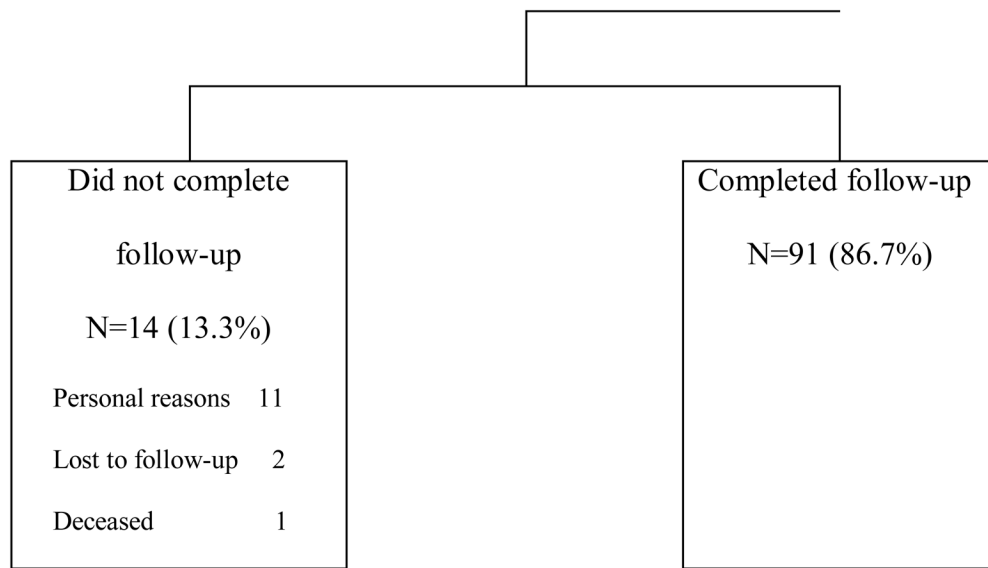


Figure 1.
CONSORT Diagram for CAT study.

Table 1

Enrollment and follow-up: number (%) of participants by visit and intervention assignment.

	Face/Face N=26	Face/Telephone N=26	Telephone/Face N=28	Telephone/Telephone N=25	Total
Baseline	26 (100)	26 (100)	28 (100)	25 (100)	105 (100)
Six-months	22 (85)	24 (92)	23 (82)	22 (88)	91 (87)

Table 2
 Characteristics at the time of enrollment grouped by intervention assignment: mean (SD) or percent.

Characteristic	Face/Face N=26	Face/Telephone N=26	Telephone/Face N=28	Telephone/Telephone N=25	P-value ¹
Race/ethnicity, n (%)					
Native American	0 (0)	1 (4)	0 (0)	0 (0)	
African American	2 (8)	5 (19)	2 (7)	2 (8)	0.282 ²
White	24 (92)	20 (77)	26 (93)	23 (92)	
Age, years, n (%)					
65–69	12 (46)	9 (35)	11 (39)	12 (48)	0.84
70–74	8 (31)	12 (46)	9 (32)	9 (36)	
75+	6 (23)	5 (19)	8 (29)	4 (16)	
Age, years, mean (SD)	72.0 (6.1)	72.9 (5.5)	72.9 (6.1)	71.5 (5.6)	0.79
Educational level, n (%)					
High school grad or less	5 (19)	2 (8)	4 (14)	3 (12)	0.69
Beyond high school	21 (81)	24 (92)	24 (86)	22 (88)	
3MSE, mean (SD)	93.9 (6.5)	93.5 (5.8)	94.6 (4.8)	93.6 (5.9)	0.90

¹ Analyses of variance for continuous measures; Pearson's chi-square or Fisher's exact tests for categorical variables

² Collapsed to white and non-white categories

Table 3

Mean (standard deviation) scores for cognitive tests and questionnaires at baseline, by mode of administration.

Measure	Possible Range of Scores	Face-to-Face N=52	Telephone N=53	P-value
Global Cognitive Functioning:				
Telephone Interview for Cognitive Status – Modified (TICS)	0–31	29.0 (1.91)	28.8 (2.60)	0.71
Verbal Fluency:				
Letter Fluency (F,A,S)	0 to infinity	34.6 (10.88)	32.9(11.28)	0.43
Category Fluency (animals)		18.9 (5.02)	17.4 (5.3)	0.14
Verbal Learning and Memory:				
CVLT				
Recall List A	0–16	27.1 (5.24)	26.6 (8.18)	0.72
Recall List B	0–16	6.4 (2.20)	6.8 (3.55)	0.51
Short Delay Free Recall	0–16	7.1 (3.16)	7.6 (3.72)	0.50
Long Delay Free-Recall	0–16	7.7 (3.20)	7.5 (4.01)	0.81
Short Delay Cued Recall	0–16	9.6 (2.48)	9.4 (3.33)	0.82
Long Delay Cued Recall	0–16	8.8 (3.15)	9.4 (3.70)	0.43
Recognition	0–44	12.6 (2.15)	12.9 (2.48)	0.53
Attention/Concentration:				
Digit Span Forward	0–14	7.9 (2.39)	8.0 (2.46)	0.88
Executive Function/Working Memory				
Digit Span Backward	0–14	6.5 (2.07)	6.8 (2.35)	0.45
Perceived Memory:				
Nurses' Health Study scale	0–7	2.31 (1.60)	2.30 (1.84)	0.98
Sleep				
Women's Health Initiative Insomnia Rating Scale (WHIIRS)	0–20	8.3 (5.20)	6.8 (4.22)	0.10
Health-Related Quality of Life				
MOS SF-12 Physical Component Summary	0–100	47.8 (8.68)	45.4 (11.92)	0.25
MOS SF-12 Mental Component Summary	0–100	53.9 (8.19)	53.2 (9.05)	0.70
Depression				
GDS-SF	0–15	1.29 (1.40)	1.62 (2.37)	0.39

Notes: TICS=Telephone Interview for Cognitive Status, CVLT=California Verbal Learning Test, HRQOL= Health-Related Quality of Life, Medical Outcomes Study (MOS) Short Form 12, GDS-SF=Geriatric Depression Scale Short Form

Table 4

Comparison of changes (6 months minus baseline) when both administrations use the same mode in standard deviation units and relative bias of telephone administration compared to face-to-face administration in standard deviation units. Covariates are age, baseline 3MSE, and race.

Measure	Mean (SE) 6-Month Changes		p-value	Mean Relative Bias of Telephone Versus Face-to-Face		p-value
	Face-to-Face	Telephone		Mean	SE	
Global Cognitive Functioning						
TICS	0.20(0.13)	0.01(0.13)	0.33	-0.16	0.10	0.09
Verbal Fluency:						
Letter Fluency (F,A,S)	-0.01 (0.11)	0.13 (0.11)	0.43	-0.09	0.08	0.26
Category Fluency-Animals)	-0.22 (0.12)	0.28 (0.12)	0.01	-0.08	0.10	0.39
Verbal Learning and Memory:						
CVLT						
Recall List A	0.09 (0.14)	0.33(0.14)	0.28	-0.02	0.11	0.86
Recall List B	-0.14 (0.15)	0.03 (0.15)	0.44	0.24	0.11	0.03
Short Delay Free Recall	0.35 (0.13)	0.37 (0.13)	0.91	0.20	0.10	0.04
Long Delay Free-Recall	0.06 (0.12)	0.60 (0.11)	<0.01	0.17	0.09	0.07
Short Delay Cued Recall	0.11 (0.14)	0.39 (0.14)	0.21	0.07	0.11	0.52
Long Delay Cued Recall	0.43 (0.12)	0.23 (0.12)	0.28	0.12	0.09	0.20
Recognition	0.47 (0.16)	0.23 (0.16)	0.32	0.00	0.12	0.70
Attention/Concentration:						
Digit Span – Forward	0.09 (0.14)	0.04 (0.14)	0.82	-0.01	0.11	0.94
Executive Function/Working/Memory						
Digit Span -Backward	0.00 (0.16)	0.24 (0.16)	0.33	0.28	0.12	0.02
Perceived memory						
Nurses' Health Study scale	-0.52 (0.12)	-0.40 (0.12)	0.54	0.14	0.09	0.14
Sleep						
WHIRS	-0.12 (0.13)	-0.18 (0.13)	0.76	-0.22	0.10	0.03
Health-related quality of life						

Measure	Mean (SE) 6-Month Changes		p-value	Mean Relative Bias of Telephone Versus Face-to-Face		p-value
	Face-to-Face	Telephone		Mean	SE	
MOS SF-12 PCS	-0.03 (0.11)	0.04 (0.11)	0.69	-0.02	0.09	0.85
MOS SF-12 MCS	0.09 (0.13)	0.15 (0.13)	0.75	0.14	0.10	0.17
Depression						
GDS-SF	0.10 (0.16)	-0.14 (0.15)	0.32	0.03	0.12	0.78

Notes: TICS=Telephone Interview for Cognitive Status, CVLT=California Verbal Learning Test, WHIRS=Women's Health Study Insomnia Rating Scale; MOS SF-12=Medical Outcomes Study Short Form 12; PCS=Physical Component Summary; MCS=Mental Component Summary GDS-SF=Geriatric Depression Scale Short Form

Table 5

Longitudinal correlations (p-value) over six months across the battery of cognitive tests.

Test	Face/Face	Face/Telephone	Telephone/Face	Telephone/Telephone
Mean (SD) correlation across 12 cognitive tests	0.67 (0.12)	0.50 (0.21)	0.66 (0.21)	0.74* (0.09)
Telephone Interview for Cognitive Status – Modified	0.57 (0.006)	0.09 (0.69)	0.88 (<0.0001)	0.70 (0.0003)
Letter Fluency	0.88 (<0.0001)	0.33 (0.12)	0.79 (<0.0001)	0.71 (0.0002)
Category Fluency-Animals	0.85 (<0.0001)	0.71 (0.0001)	0.85 (<0.0001)	0.88 (<0.0001)
CVLT List A	0.63 (0.002)	0.75 (<0.0001)	0.45 (0.03)	0.83 (<0.0001)
CVLT List B	0.51 (0.02)	0.27 (0.19)	0.24 (0.27)	0.77 (<0.0001)
CVLT Short Delay – Free recall	0.71 (0.0002)	0.54 (0.006)	0.80 (<0.0001)	0.71 (0.0002)
CVLT long delay – Free recall	0.74 (<0.0001)	0.52 (0.01)	0.82 (<0.0001)	0.82 (<0.0001)
CVLT Short Delay – Cued recall	0.69 (0.0004)	0.47 (0.02)	0.74 (<0.0001)	0.63 (0.002)
CVLT long delay – Cued recall	0.53 (0.01)	0.62 (0.001)	0.83 (<0.0001)	0.82 (<0.0001)
CVLT Recognition	0.48 (0.02)	0.29 (0.17)	0.52 (0.01)	0.69 (0.0004)
Digits Span Forward	0.73 (0.0001)	0.61 (0.002)	0.36 (0.10)	0.76 (<0.0001)
Digits Span Backward	0.68 (0.0006)	0.51 (0.01)	0.51 (0.01)	0.57 (0.007)
Perceived memory function	0.66 (0.0008)	0.82 (<0.0001)	0.77 (<0.0001)	0.68 (0.0005)
Women’s Health Initiative Insomnia Rating Scale	0.64 (0.001)	0.87 (<0.0001)	0.83 (<0.0001)	0.65 (0.001)
MOS Short Form 12				
Physical Component	0.80 (<0.0001)	0.62 (0.002)	0.82 (<0.0001)	0.86 (<0.0001)
Mental Component	0.77 (<0.0001)	0.59 (0.003)	0.48 (0.02)	0.60 (0.004)
Geriatric Depression Scale-Short Form	0.55 (<0.01)	0.27 (0.20)	0.65 (<0.001)	0.43 (0.05)

Note: CVLT = California Verbal Learning Test; MOS = Medical Outcome Study; * Statistically distinct (p<0.05) from the mean correlation for the telephone to face-to-face sequence via a Scheffe multiple comparisons test