Published in final edited form as:

Alzheimer Dis Assoc Disord. 2010 ; 24(3): 256–263. doi:10.1097/WAD.0b013e3181d7109f.

Pilot Study to Show the Feasibility of a Multicenter Trial of Home-based Assessment of People Over 75 Years Old

Mary Sano, PhD^{*,†}, Susan Egelko, PhD^{*,†}, Steven Ferris, PhD^{‡,§}, Jeffrey Kaye, MD^{||}, Tamara L. Hayes, PhD^{||}, James C. Mundt, PhD^{||}, Michael Donohue, PhD[#], Sarah Walter, MS[#], Shelly Sun, MS[#], and Luis Saucedo-Cerda[¶]

^{*}Mount Sinai School of Medicine, New York

[†]The James J Peters VAMC, Bronx

[‡]NYU School of Medicine, New York

[§]Nathan Kline Institute, Orangeburg, NY

^{||}Oregon Health & Science University, Portland, OR

[¶]Healthcare Technology Systems, Madison, WI

[#]UCSD, La Jolla, CA

Abstract

This report describes a pilot study to evaluate feasibility of new home-based assessment technologies applicable to clinical trials for prevention of cognitive loss and Alzheimer disease.

Methods—Community-dwelling nondemented individuals 75 years old were recruited and randomized to 1 of 3 assessment methodologies: (1) mail-in questionnaire/live telephone interviews (MIP); (2) automated telephone with interactive voice recognition (IVR); and (3) internet-based computer Kiosk (KIO). Brief versions of cognitive and noncognitive outcomes were adapted to the different methodologies and administered at baseline and 1-month. An Efficiency measure, consisting of direct staff-to-participant time required to complete assessments, was also compared across arms.

Results—Forty-eight out of 60 screened participants were randomized. The dropout rate across arms from randomization through 1-month was different: 33% for KIO, 25% for IVR, and 0% for MIP (Fisher Exact Test $P = 0.04$). Nearly all participants who completed baseline also completed 1-month assessment (38 out of 39). The 1-way ANOVA across arms for total staff-to-participant direct contact time (ie, training, baseline, and 1-month) was significant: $F(2,33) = 4.588$; $P = 0.017$, with lowest overall direct time in minutes for IVR (Mn = 44.4; SD = 21.5), followed by MIP (Mn = 74.9; SD = 29.9), followed by KIO (Mn = 129.4; SD = 117.0).

Conclusions—In this sample of older individuals, a higher dropout rate occurred in those assigned to the high-technology assessment techniques; however, once participants had completed baseline in all 3 arms, they continued participation through 1 month. High-technology home-based assessment methods, which do not require live testers, began to emerge as more time-efficient over the brief time of this pilot, despite initial time-intensive participant training.

Keywords

alzheimer disease; prevention trials; home-based assessment

Major challenges exist for conducting Alzheimer disease (AD) primary prevention trials in recruitment and retention of elderly cohorts and evaluating them over extended periods of time. First, elderly populations often present with a range of physical, social, and health issues, which limit study participation, particularly when a clinical site is geographically remote. Second, identifying a target population for a clinical trial within a general medical setting requires establishing operational study criteria and diagnostic categories at entry to classify homogeneous subgroups of participants with or without memory deficits or other forms of mild cognitive impairment (MCI). Third, assessment batteries need to be developed that are: sensitive to change over time; have sufficient breadth to tap key domains that mark transition from cognitive health to dementia; and suitable for large-scale administration.

This report describes a pilot study in which we incorporated the cognitive and noncognitive domains that Ferris et al¹ identified as important to characterizing the transition to dementia. Our goal was to adapt these assessment measures to a home-based longitudinal study of a population of community-dwelling elderly people, using technologies that reduce the need for face-to-face assessment and participant travel. If such home-based assessment is feasible, it would allow a potentially larger and broader sampling of individuals to participate in studies.²

This study included screening procedures to identify nondemented elders, categorizing them as having normal cognition or MCI by a battery of well-established in-person neuropsychologic measures.^{3,4} Participants were then randomly assigned to 1 of 3 experimental methods of home-based longitudinal assessment: (1) mail-in/phone (MIP): A validated cognitive assessment was conducted by phone with a live, trained tester, supplemented by written mail-in questionnaires assessing noncognitive domains; (2) assessment and data collection through telephone-based automated interactive voice response (IVR)^{5–7} for all cognitive and noncognitive measures (no live tester); and (3) computer-based, video-directed assessment, and data collection through kiosk (KIO) (no live tester). Abbreviated versions of cognitive and noncognitive measures from key domains associated with change from nondemented to MCI and dementia were administered in formats specific to the technology of experimental method of assessment.

In addition, medication adherence to a twice-daily vitamin regimen, a performance-based instrumental activity of daily living, was assessed by methods specific to each technology arm.

The study was conducted by the Alzheimer Disease Cooperative Study (ADCS), a national consortium of sites with expertise in the conduct of clinical trials in cognitive loss and dementia.⁸ The purpose of this pilot study was to establish the feasibility of recruitment, randomization, and in-home evaluation of elderly participants. In addition to measures of participant performance, measures of efficiency (staff-to-participant contact needs) were developed to compare these different in-home evaluation methods. The pilot study offered an opportunity to standardize operational procedures, automate scoring, and establish transfer of data from individual sites to a centralized database, all of which provided a model for a subsequent large-scale study.

METHODS

Participants

Participants were recruited from 3 ADCS sites similar to those likely to be recruited into a primary prevention trial.

Inclusion criteria were: More than or equal to 75 years old; Mini-Mental State Examination (MMSE)⁹ score more than or equal to 26; willingness to take study multivitamins; minimal computer skills or willingness to learn; English fluency; ability to answer and dial a telephone; adequate speech, hearing, and vision to complete assessments; and independently living. A study partner, was encouraged, but not required.

Exclusion criteria were: dementia diagnosis; using prescriptive cognitive-enhancing drugs; intent to continue use of own nonprotocol multivitamins during the study; history or presence of major psychiatric, neurologic, or neurodegenerative conditions associated with significant cognitive impairment; unstable housing arrangements over the 3-month duration of the pilot study; or current participation in a clinical trial involving CNS medications or cognitive testing.

Recruitment

Each site targeted recruitment at communities with large numbers of age-appropriate elderly participants and located close enough to the research site to ensure ready access to in-person technical support. Each site used a variety of recruitment strategies, for example, lectures on memory enhancement, outreach through direct mail, and posting fliers. Informed consent was obtained and participants were offered a small financial incentive for study completion, in accordance with local institutional review boards (IRBs).

Experimental Methods of Home-based Assessment

Participants were randomized to 1 of 3 home-based assessment methods:

Mail-in/phone (MIP). The cognitive assessment was conducted by a trained evaluator through live telephone call with the participant. Noncognitive assessment and the experimental medication adherence were conducted by mail.

Interactive Voice Recognition (IVR) Assessment through automated telephone. A standard large-key telephone was installed in the home; cognitive, noncognitive, and medication adherence assessments were conducted through an automated telephone system. Responses were recorded and scored through voice recognition and keypad entry, requiring no live staff.

Kiosk (KIO) Computerized Assessment. A computer kiosk, consisting of a touch screen and attached telephone handset, and in most cases a new broadband internet connection, was installed in the home. Cognitive and noncognitive tests were presented orally and visually by an on-screen videotaped tester; responses were recorded by voice through a telephone handset or by touch screen. Experimental medication adherence was assessed by MedTracker,^{10,11} an instrumented 7-day reminder pill box, linked by Bluetooth to the in-home study computer to record precise times of pill-taking.

Staff Training

A day-long investigator meeting was conducted to train staff, including installation of equipment for IVR and KIO arms, and procedures for training participants in the use of in-home equipment. For the MIP arm that required a live phone tester for the cognitive assessment, coordinators received training in test administration and scoring. Tester competence was confirmed by a certification test assessing knowledge of administration and scoring of the cognitive assessment battery.

Timeline of Procedures (Fig. 1)

Preparatory steps at each site included identifying local broadband internet service providers and establishing relationships with targeted community sites. After participants signed consent forms, an in-person *screening visit* was conducted to determine eligibility. During screening the participant was administered an in-person neuropsychologic battery taken from the Uniform Data Set of the National Alzheimer Coordinating Center that is widely used in clinical research protocols (NACC-UDC, ADNI).^{3,12} The neuropsychologic battery assessed verbal episodic memory, attention, semantic memory/language, psychomotor speed, and executive functioning.³ The clinician categorized participants as normal or Mild Cognitive Impairment (MCI) based on impressions of memory impairment from interview and available neuropsychologic evaluation.

Training visit. The training visit took place in the participant's home for the IVR and KIO arms and either by phone or at the participant's home for the MIP arm. The training visit consisted of a mock demonstration of test taking by the participant. The baseline evaluation was scheduled within the next week.

Baseline Visit. For the MIP group the cognitive battery was conducted by phone with a live tester; the tester reminded the participant to mail in the paper-and-pencil noncognitive battery in a prestamped addressed envelope. For the 2 high technology arms, appointments were scheduled through their respective automated technologies. If the visit was not completed, staff contacted the participant and the effort by staff was captured on the Efficiency form. No staff time was required for either the cognitive or noncognitive evaluations for IVR and KIO arms.

One-Month Follow-up. An assessment was scheduled 1 month after the baseline visit and was to be initiated by the participant. If a visit was missed, staff contacted the participant and the effort was captured on the Efficiency form.

Domains of Assessments

The cognitive^{13–17} and noncognitive^{18–25} evaluations each included 8 domains, represented by brief instruments suitable for repeated assessment. Tests were adapted to the technological format of each of the study arms, preserving as much as possible the integrity of the original in-person test.

Cognitive Performance Assessment

The cognitive performance battery was designed to require about 30 to 40 minutes in duration, presented in a set sequence (Table 1). The test order was designed to achieve an approximate 15 to 20 minute interval between the immediate and delayed recall of the East Boston Memory Story. A checklist of adverse events was collected midway through the cognitive battery as a noncognitive filler. Participants were requested to allot 40 minutes for the in-home cognitive test session and encouraged to not take any breaks. However, if a break occurred, testing was resumed at the beginning of the discontinued test.

Noncognitive Measures

The noncognitive portion of the home-based assessment (Table 1) was completed by mail for the MIP arm, by automated telephone for the IVR arm and by automated computer kiosk for the KIO arm. Participants were instructed to take the study multivitamin in the morning and evening and an experimental medication adherence measure was generated for each of the arms as described above. All participants also returned pills at the end of the pilot study, a gold standard medication adherence rate that will be used in the Main Study.

Efficiency Measures

Staff contact was measured by the frequency and length of time spent with participants, either in-person or by phone. These contacts included: nonscheduled contacts with staff (eg, nonevaluation phone calls), in-person staff time to train participants in the use of the assessment methods (ie, MIP, IVR, or KIO), and staff time to administer the cognitive battery for the MIP arm by phone. (The other 2 arms required no staff time for home-based experimental testing.) A total staff-to-participant length of direct contact variable (in minutes) was generated by summing: (a) time to train; (b) baseline additional staff-to-participant contact; (c) baseline cognitive testing (MIP only); (d) 1-month additional staff-to-participant contact; e) 1-month cognitive testing (MIP only).

Data Analyses

The disposition of recruited participants was reported for this pilot study according to the Consort recommendations (Fig. 2) and key descriptive data on participant demographic and health status were summarized. To address feasibility of enrollment, randomization, and repeated data collection over the brief, 1-month interval, the 3 arms were compared with respect to (a) the number of participants discontinued after randomization; (b) length of time from screening to baseline; and (c) retention rate through 1-month follow-up assessment. Statistical comparisons among methods were also made on the composite “efficiency” measure, described above. The number of incomplete tests in the cognitive battery (out of a total of 8) was compiled across arms.

Categorical data (ie, frequency counts) were analyzed by Fisher Exact Test of equal proportions and if $P < 0.05$, pairwise comparisons were conducted. For continuously measured variables, 1-way analyses of variance (ANOVA) were conducted across arms. If Bartlett Test of homogeneity of variances was nonsignificant ($P = 0.05$) for a given variable, a Tukey post hoc analysis was conducted. If Bartlett test was significant ($P < 0.05$), Welch t tests²⁶ (accommodating unequal variances) were conducted, with Hochberg²⁷ adjustment for multiple comparisons (ie, pairwise comparisons = MIP:IVR, MIP:KIO, IVR:KIO). All tests were 2-tailed with α set at 0.05.

The means and standard deviations for scores on the cognitive and noncognitive measures were summarized for each arm at baseline and 1 month. A test-retest Pearson r correlation was calculated between baseline and 1-month scores to estimate the reliability of these abbreviated tests. Statistical comparisons of cognitive and noncognitive scores across arms were deferred until a large sample size is collected in the main study.

RESULTS

Demographics

A total of 60 participants were enrolled in the study, of which 48 (80%) were randomized (Fig. 2). The reasons that 12 participants discontinued before randomization included: general unwillingness to be randomized ($n = 3$); only willing to be randomized to KIO arm ($n = 1$); health concerns ($n = 2$); failure on MMSE ($n = 1$); unwillingness to take study vitamin ($n = 1$); and unspecified reasons unrelated to any of the arms ($n = 4$). The 48 randomized participants were distributed as: MIP $n = 14$; IVR $n = 16$; and KIO $n = 18$. Nine of the 48 randomized participants discontinued before baseline: 6 from KIO (33%), 3 from IVR (19%). Five of the 6 participants who discontinued from KIO reported reasons specific to that arm’s technology: for example, “too much trouble getting broadband internet.” Two of the 3 IVR participants who discontinued were described as “unable to complete the baseline.” The others withdrew with no specific comment recorded. The dropout rate across arms from randomization through completion of the study at 1 month follow-up was

significantly different: 33% for KIO, 25% for IVR, and 0% for MIP (Fisher Exact Test $P = 0.04$; none of the pairwise comparisons were significant).

The mean intervals in days that elapsed between screening and baseline evaluation were: MIP = 30.1 (SD = 14.8); IVR = 24.5 (SD = 8.8); and KIO = 35.6 (SD = 15.2). These time intervals were not significantly different across arms [$F(2,36) = 2.17$; $P = 0.129$]. Nor were there significant differences across arms in retention rate from baseline to Month 1 evaluation (Fisher Exact Test, $P = 0.641$). Of the 39 participants who completed baseline, only 1 participant (in the IVR arm) failed to complete the assessments at Month 1, yielding a 97.4% retention.

Demographic and health status variables were not significantly different across the 3 arms. Overall, the mean age of the 39 randomized participants was 82.1 (SD = 5.2) and the number of years of education was 15.8 (SD = 2.9; Range = 8 to 20). A total of 29 participants (74%) were female and 8 (21%) were married. The ethnicity of the pilot sample was homogeneous: 36 (92.3%) White; 2 (5.1%) African-American; and 1 (2.6%) Native American. The near-ceiling mean score of 29.0 (SD = 1.2) on the MMSE was reflective of the study cutoff score of Z26. A total of 27 participants (69%) reported cardiovascular medical condition, 18 (46%) self-reported a memory complaint, and 7 (15%) met the clinician rating for MCI.

Feasibility: Narrative Description of Technical Issues in Start-up of Home-based Assessment—Scorable data were produced for all cognitive and noncognitive experimental tests by the completion of the pilot study. Study coordinators from each of the 3 pilot sites (2 in New York City and 1 in Portland, OR) reported similar technical start-up issues. They reported no implementation problems with the low-technology MIP arm.

Site staff encountered the most frequent and time-consuming difficulties in the set up of the KIO arm. These difficulties included: delays in broadband installation; difficulty installing speakers; need to permit remote access and multiple computer access for the participant; insufficient power from the MedTracker battery pack; and need for real-time assistance for computer installation. This technical assistance was not captured in the Efficiency measures, which focused exclusively on contacts that staff made directly with participants. In addition, data were lost for the Digit Span Backwards because the length of pauses in utterance was not considered in the original scoring paradigm for KIO and administration was terminated prematurely until this problem was corrected. In addition, 1 KIO participant was missing the Abbreviated TICS at baseline and 1-month due to hearing difficulty. All other experimental tests, both cognitive and noncognitive, were complete for this arm.

Statistical Comparison of the Number of Incomplete Cognitive Tests—The 3 arms were compared at baseline and Month 1 with respect to the number of incomplete tests in the home-based experimental cognitive battery (max = 8 tests). At baseline, none of the participants in the MIP ($n = 14$) or IVR ($n = 13$) arms were missing any tests. In the KIO arm 8 participants were missing 1 test (Digit Span Backward), 1 participant was missing 2 tests (Digit Span Backward and the Abbreviated TICS); 3 participants had all cognitive tests completed. The Fisher Exact Test across groups was highly significant ($P < 0.001$). At Month 1, although all 12 KIO participants were retained for Month 1 testing, 9 participants were missing Digit Span Backward and 1 participant was missing both Digit Span Backward and Abbreviated TICS. As with baseline testing, none of the MIP ($n = 14$) or IVR ($n = 12$) participants had any incomplete tests at Month 1. The Month 1 Fisher Exact Test of number of incomplete tests across arms was highly significant ($P < 0.001$).

Description of Incomplete Items in Noncognitive Tests and Medication

Adherence—Noncognitive tests were largely completed in all 3 arms. All participants had complete data for BCFSI, QOL, Behavioral, and ADL at both baseline and Month 1; and the CGIC was completed for all groups at Month 1. On Participant Status 2 participants out of 12 were missing information about cognitive-enhancing medications in the IVR arm at Month 1. On the Resource Use Inventory, 1 out of 13 IVR participants was missing the number of hours/week with helpers at baseline.

For **Medication Adherence** all 12 KIO participants had scorable MedTracker data and according to the returned number of pills, all participants had taken at least some medication. In the MIP arm, all participants mailed in self-reported use of medication. However, 2 out of 14 MIP participants had a missing return date so the adherence was unscorable and 1 MIP participant returned all pills. In the IVR arm, 1 out of 12 IVR participants was missing medication adherence data collected by self-reported automated phone response. In addition, 1 IVR participant had a missing return date for pill count and 1 IVR participant returned all pills. The medication adherence data collection in this pilot was targeted to standardized implementation of the procedures only; the actual measurement and data analysis (how the experimental arms each compare with the gold standard of returned pill count) have been deferred to the Main Study.

Efficiency

Number of Participants Requiring Additional Staff Contact (outside of training or testing)—At baseline, 9 (75%) participants in KIO required study coordinator contacts outside of training and testing, as compared with 7 IVR (54%) and 3 (21%) MIP participants. The overall Fisher Exact Test was significant ($P = 0.024$), with a significant pairwise comparison, using the Hochberg adjustment, between KIO and MIP ($P = 0.048$). By 1-month, however, despite a borderline level of significance (Fisher Exact $P = 0.078$), the magnitude of the differences was less noticeable. The numbers of participants requiring outside staff contact at 1-month were: 8 KIO (67%), 11 IVR (85%), and 6 MIP (43%).

Total Time for Staff to Participant Contact: Training, Additional Contacts, and Testing—As noted in Table 2, the arms were significantly different in the amount of staff-to-participant training time required: $F(2,36) = 8.834$; $P = 0.001$. The mean time in minutes for training subjects in each arm's technology were: MIP Mn = 15.2 (SD = 6.2); IVR Mn = 31.6 (SD = 6.2) and KIO Mn = 104.2 (SD = 101.2). KIO training time was significantly longer than MIP ($P = 0.022$) and IVR ($P = 0.031$) and IVR training time was significantly longer than MIP ($P < 0.001$). At baseline, additional staff-to-participant contact time varied across groups [$F(2,36) = 3.988$; $P = 0.027$]. However, none of the pairwise comparisons were significant. By 1 month there were no significant differences across groups in additional staff-to-participant contact time [$F(2,36) = 0.144$; $P = 0.866$]. The 1-way ANOVA across arms for total staff-to-participant direct contact time (ie, training time + additional contact time + time live tester spent in experimental cognitive testing for MIP arm) was significant: $F(2,33) = 4.588$; $P = 0.017$. The group that required the lowest overall direct time in minutes was IVR (Mn = 44.4; SD = 21.5), followed by MIP (Mn = 74.9; SD = 29.9), followed by KIO (Mn = 129.4; SD = 117.0). The only significant pairwise difference was found between IVR and MIP, the former being the most time efficient ($P = 0.034$) with respect to total staff-to-participant direct contact time through the entire length of the study.

Test Results for Experimental Cognitive and Noncognitive Home-based Assessments

—The scores for cognitive and noncognitive tests by arm are presented in Table 3. Pearson test-retest reliability coefficients comparing baseline and 1 month scores were significant ($P < 0.001$) for all cognitive and noncognitive experimental measures,

ranging from 0.58 to 0.84. As the sample size was small in this pilot, statistical comparisons of scores were deferred until data are analyzed in the main study.

DISCUSSION

This pilot study showed the challenges of recruitment of elderly individuals to a home-based study of cognitive and noncognitive change using traditional and novel methods for evaluation. The cohort was quite old, with a mean age above 80, and nearly 70% reported cardiovascular conditions, thereby showing that an “at risk” population can be recruited. Randomization into the novel technology arms (KIO and IVR) was associated with an increased likelihood of discontinuation and dropout before baseline; participants assigned to KIO in particular cited the inconvenience of the technology. This finding points to the need to vet new technologies fully with real-life experience before deploying them to home use. It is noteworthy that once participants were familiarized with study procedures (ie, completing baseline) they remained engaged, as supported by the over 90% retention from baseline to 1-month follow-up across all arms.

Objective findings corroborate subjective accounts of greater difficulty with the initiation of the novel, automated methods, and particular difficulty with the KIO installation. The staff time was greater for the KIO arm, particularly during training and baseline data collection, and did not include the additional time for installation or the “real time” technical assistance for computer trouble-shooting. Nevertheless, the staff-time at startup for the new technologies resulted in more efficiency over time, even over this short study. For example, when total staff time was tallied for training, baseline, and 1-month follow-up periods, the IVR was already more time efficient than the MIP arm. Given the estimate from the current pilot study that each live cognitive assessment in the MIP arm requires 22 minutes for completion, after 3 additional follow-up assessments the KIO arm would also be projected to require less staff time, assuming no additional technological problems occurred. Certainly, longer follow-up with more assessments might result in significant savings with these automated methodologies.

This protocol represents the first use of this cognitive battery without in-person administration. It is also the first use of the abbreviated noncognitive assessments. Data collection for all measures was nearly complete across all arms. The test-retest reliability of all experimental measures was good, supporting their stability when collected by these novel formats. There were no significant differences across the 3 arms in any of the experimental cognitive scores at 1 month, with scores from each of the 3 experimental arms comparable with in-person “gold standard” tests.

Medication adherence served as a performance-based measure of “activities of daily living” and the data collection techniques were unique to each of the 3 technologies. Scorable data were obtained from all but 3 participants, indicating that these methods are reasonable for data collection in an elderly population.

This pilot study informed the currently ongoing, much larger study designed to evaluate the sensitivity of these methods over a 4-year interval. Some of the limitations of the pilot study, such as limited diversity of race/ethnicity and levels of education, have been addressed. The main study targets participants with a diversity of demographic factors that play a role in generalizability of findings (eg, 20% minority enrollment is required at each participating site).

The pilot study identified the gap between the relatively high initial enthusiasm for higher technologies (such as computers) and the reluctance to fully participate once randomized. The pilot study also identified the fear of the size and inconvenience of internet-based

technologies and prompted us to develop strategies to address these concerns. For example, computer Kiosks delivered in shipping cartons intimidated potential participants with small living quarters, and delivery of uncrated computers helped alleviate this concern. Providing companionship while service people installed cable connections overcame anxiety about having a “stranger” in the home. “Real time” help lines for local site staff for installation problems allowed the participant to see the technical issues as manageable.

The most encouraging finding was the evidence that efforts early in the trial had long-lasting benefits in terms of participation and follow-up. The main study will provide larger sample sizes in which we will be able to evaluate and compare methods for efficiency and accuracy in early detection of cognitive deterioration and dementia.

Acknowledgments

The authors are grateful to Dr Paul Aisen for review of the manuscript and editorial advice. The authors would also like to acknowledge the contributions of Tracy Reyes and Ben Barth at Healthcare Technology Systems; Jacques H. de Villiers, Rachel Coulston, Esther Kladders, and John Paul Hosom at the OHSU Center for Spoken Language and Understanding; Jessica Payne-Murphy at the OHSU Oregon Center for Aging & Technology; and Study Coordinators Melissa Rushing at Mount Sinai School of Medicine; and Erica Maya at NYU School of Medicine.

Supported by these NIA grants: U01AG10483, P50AG005138, P30AG008051, and P30AG024978. Development of the Kiosk and MedTracker was supported in part by grants from NIA (P30-AG024978; P30-AG08017) and Intel Corporation.

REFERENCES

1. Ferris SH, Aisen PS, Cummings J, et al. Alzheimer's Disease Cooperative Study Group. ADCS Prevention Instrument Project: overview and initial results. *Alzheimer Dis Assoc Disord.* 2006; 20(suppl 3,4):S109–S123. [PubMed: 17135805]
2. Sano M, Zhu CW, Whitehouse PJ, et al. Alzheimer Disease Cooperative Study Group. ADCS Prevention Instrument Project: pharmacoeconomics: assessing health-related resource use among healthy elderly. *Alzheimer Dis Assoc Disord.* 2006; 20(suppl 3,4):S191–S202. [PubMed: 17135812]
3. 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; 20:210–216. [PubMed: 17132964]
4. Grundman M, Petersen RC, Ferris SH, et al. Alzheimer's Disease Cooperative Study. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Arch Neurol.* 2004; 61:59–66. [PubMed: 14732621]
5. Mundt JC, Kinoshita LM, Hsu S, et al. Telephonic remote evaluation of neuropsychological deficits (TREND): longitudinal monitoring of elderly community-dwelling volunteers using touch-tone telephones. *Alzheimer Dis Assoc Disord.* 2007; 21:218–224. (2007). [PubMed: 17804954]
6. Piette JD. Interactive voice response systems in the diagnosis and management of chronic disease. *Am J Manag Care.* 2000; 6:817–827. [PubMed: 11067378]
7. Mundt JC, Geralts DS, Moore HK. Dial “T” for Testing: technological flexibility in neuropsychological assessment. *Telemed J E Health.* 2006; 12:317–323. [PubMed: 16796499]
8. <http://www/ADCS.com>
9. 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; 12:189–198. [PubMed: 1202204]
10. Hayes TL, Hunt JM, Adami A, et al. An electronic pillbox for continuous monitoring of medication adherence. *Conference Proceedings: Engineering in Medicine and Biology Society.* 2006; 1:6400–6403.
11. Hayes TL, Larimer N, Adami A, et al. Medication adherence in healthy elders: small cognitive changes make a big difference. *J Aging Health.* 2009; 21:567–580. [PubMed: 19339680]

12. Mueller SG, Weiner MW, Thal LJ, et al. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement*. 2005; 1:55–66. [PubMed: 17476317]
13. Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*. 1989; 39:1159–1165. [PubMed: 2771064]
14. Salmon DP, Cummings JL, Jin S, et al. ADCS Prevention Instrument Project: development of a brief verbal memory test for primary prevention clinical trials. *Alzheimer Dis Assoc Disord*. 2006; 20(suppl 3,4):S139–S146. [PubMed: 17135807]
15. Petersen RC, Thomas RG, Grundman M. Vitamin E and donepezil for the treatment of mild cognitive impairment. *N Engl J Med*. 2005; 352:2379–2388. [PubMed: 15829527]
16. Reitan RM. Validity of the Trail Making Test as an indication of organic brain damage. *Percept Mot Skills*. 1958; 8:271–276.
17. Ricker H, Axelrod BN. Analysis of an oral paradigm for the Trail Making Test. *Assessment*. 1994; 1:47–51. [PubMed: 9463499]
18. Walsh SP, Raman R, Jones KB, et al. ADCS Prevention Instrument Project: the Mail-In Cognitive Function Screening Instrument (MCFSI). *Alzheimer Dis Assoc Disord*. 2006; 20(suppl 3,4):S170–S178. [PubMed: 17135810]
19. Patterson MB, Whitehouse PJ, Edland SD, et al. ADCS Prevention Instrument Project: quality of life assessment (QOL). *Alzheimer Dis Assoc Disord*. 2006; 20(suppl 3,4):S179–S190. [PubMed: 17135811]
20. Hoyt MT, Alessi CA, Harker JP, et al. Development and testing of a five-item version of the Geriatric Depression Scale. *J Am Geriatr Soc*. 1999; 47:873–878. [PubMed: 10404935]
21. Rinaldi P, Mecocci P, Benedetti C, et al. Validation of the five-item Geriatric Depression Scale in elderly subjects in three different settings. *J Am Geriatr Soc*. 2003; 51:694–698.
22. Avsarala JR, Cross AH, Trinkaus K. Comparative assessment of the Yale Single Question and the Beck Depression Scale in screening for depression in multiple sclerosis. *Mult Scler*. 2003; 9:307–310. [PubMed: 12814181]
23. Cummings JL, Mega M, Gray K, et al. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. *Neurology*. 1994; 44:2308–2314. [PubMed: 7991117]
24. 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(suppl 3,4):S152–S169. [PubMed: 17135809]
25. Schneider LS, Clark CM, Doody R, et al. ADCS Prevention Instrument Project: ADCS-clinicians' global impression of change scales (ADCS-CGIC), self-rated and study partner-rated versions. *Alzheimer Dis Assoc Disord*. 2006; 20(suppl 3,4):S124–S138. [PubMed: 17135806]
26. Welch BL. The generalization of 'student's' problem when several different population variances are involved. *Biometrika*. 1947; 34:28–35. [PubMed: 20287819]
27. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*. 1988; 75:800–803.

FIGURE 1.
Visit Timeline.

FIGURE 2.
Enrollment Summary.

TABLE 1

Experimental Home-based Assessment Measures: Cognitive and Noncognitive Measures

	Description	Score
Experimental Cognitive Measures		
1. Immediate word list recall (CERAD) ¹⁴	10-item word list with recall after each of 3 trials.	Mean # correct of 3 trials: range = 0–10
2. East Boston memory test-immediate recall ¹⁵	Brief, 1-paragraph story recall with 6 key elements.	Range = 0–6
3. Abbreviated TICS ¹⁶	8 items from the telephone-based adaptation of the Mini-Mental State Examination (MMSE).	Range = 0–8
4. Backward digit span (WAIS-R)	Strings of numbers range between 2 and 7 digits. Discontinuation after 2 consecutive errors on 2 consecutive sequence lengths.	2 Scores: total # correct: range = 0–12; string length: range = 0–7
5. Adapted trail making test: A and B ¹⁷	This test is preserved in its visual-spatial format for the KIO arm and adapted to an oral format for MIP and IVR arms. Good comparability between visual and oral versions has been demonstrated. ¹⁸	4 Scores: time to first error part A/B; time to complete part A/B
Adverse event check list	Filler test to ensure proper time delay for the 2 memory tests.	—
6. Delayed word recall	Recall of the initial 10-item word list is tested without cueing.	Range = 0–10
7. Category fluency (animals) (CERAD) ¹⁴	The participant is asked to say as many different animal names as possible in a 1-minute period; no prompting provided.	# Unique names that qualify as animals
8. East Boston memory test-delayed recall	Recall of the initial paragraph was tested after a 15–20 min delay; no cueing provided.	Range = 0–6
Experimental Noncognitive Measures		
1. Brief cognitive function screening instrument (BCFSI) ¹⁹	8 items of self-reported declines in memory and thinking during the past year. Each item is rated on 3-point scale: 0 (no change), 0.5 (maybe), 1 (yes).	Range = 0–8; higher scores indicate greater impairment.
2. Quality of life (QOL) ²⁰	8 items include: current general health, energy, memory ability, whole self, ability to do chores, enjoyment, and financial status. Each item is rated on a 4-point scale that ranges from 1 (poor) to 4 (excellent).	Range = 0–32; only noncognitive test in which higher scores indicate <i>lower level</i> of impairment
3. Behavioral scale ^{21–24}	8 Items include key psychiatric symptoms, i.e., depression, irritability, apathy, and anxiety. The time frame is the past week and symptoms are rated as either 1 (present) or 0 (absent).	Range = 0–8, higher scores indicating greater impairment
4. Activities of daily living (ADL) ²⁵	8 Instrumental ADL items sensitive to subtle changes in cognition include: money management, remembering appointments, finding belongings, etc. The time frame is the past 3 months and ratings are made on a 4-point scale that ranges between 0 (no difficulty) to 3 (not done at all).	Range = 0–24, higher scores indicating greater impairment
5. Clinical global inventory of change (CGIC) ²⁶	The baseline instrument consists of 4 open-ended questions about specific health-related events, cognition, behavior/psychiatric, vegetative signs, ADL, and 1 overall status question. At follow-up, the participant reviews what he had said at baseline and rates how he has changed. The participant then rates perceived overall change in 1 of 7 categories, ranging between “very much better” to “very much worse.”	7-Item categoric scale
6. Resource use inventory ³	A tally of the # of hospitalizations in the past 1 month formal and informal help and hours of paid or voluntary work in the past 3 months.	
7. Participant status ³	A brief 4-question self-report assesses changes in the past 1 month in medication and/or recent healthcare encounters.	
8. Medication adherence	Written medication diary in MIP; self-reported key pad response in IVR; and MedTracker computer register for precise time of pill-taking in KIO.	

TABLE 2

Efficiency: Length of Components and Total Time of Staff/Subject Contacts for Experimental Cognitive Testing Only

Components of Staff/Subject Contact (min)	Mail-In/Phone: MIP (N = 14)	IVR (N = 13)	KIO (N = 12)	<i>P</i> [*]
Time to train	15.2 (6.2)	31.6 (12.4)	104.2 (101.2)	0.001 [†]
Baseline additional contact	1.2 (4.0)	4.2 (8.7)	13.3 (17.5)	0.027 [‡]
Baseline cognitive testing	21.5 (3.0) [§]	N/A	N/A	N/A
1-month additional contact	10.9 (21.5)	8.5 (6.0)	12.0 (17.7)	0.866
1-month cognitive testing	21.8 (5.7) [§]	N/A	N/A	N/A
Total time	74.9 (29.9)	44.4 (21.5)	129.4 (117.0)	0.017 [¶]

^{*} Staff/Subject Contact was compared across the 3 groups using 1-way ANOVA. When the overall *F* test was significant for a component, we used Welch Pairwise *t*-tests, with Hochberg adjustment, to evaluate dyad differences due to nonhomogeneous variance across groups.

[†] Significant dyads: KIO:IVR; KIO:MIP; IVR:MIP.

[‡] No significant pairwise comparisons.

[§] N = 12.

^{||} N = 11.

[¶] Significant dyad: IVR:MIP.

TABLE 3

Select Cognitive and Noncognitive Experimental Test Scores by Arm (Baseline and 1-mo Data) and Test-retest Reliability for Full Cohort (N = 38)

Experimental Cognitive Test Scores	Mail-In/Phone: MIP (N = 14)*	IVR (N = 13) [†]	KIO (N = 12)	Pearson Correlation B/L:1-Mo [‡]
<i>B/L</i> Immediate word list recall	6.90 (1.57)	5.59 (1.94)	5.94 (1.33)	0.635
<i>1-Mo.</i> Immediate word list recall	7.69 (1.14)	6.47 (1.81)	6.56 (1.32)	
<i>B/L</i> Delayed word list recall	5.14 (3.08)	5.46 (2.76)	3.83 (2.17)	0.598
<i>1-Mo.</i> Delayed word list recall	6.57 (2.28)	6.33 (2.10)	5.83 (2.82)	
<i>B/L</i> Immediate East Boston memory test recall	3.57 (1.22)	4.85 (1.46)	4.33 (1.87)	0.585
<i>1-Mo.</i> Immediate East Boston memory test	4.57 (1.34)	5.00 (1.04)	3.92 (1.56)	
<i>B/L</i> Delayed East Boston memory test	3.36 (1.55)	4.92 (1.61)	3.33 (2.02)	0.635
<i>1-Mo.</i> Delayed east Boston memory test recall	4.64 (1.39)	4.67 (1.37)	3.92 (1.68)	
<i>B/L</i> Immediate Abbreviated TICS	7.43 (0.65)	6.77 (1.30)	6.45 (0.93) [§]	0.583
<i>1-Mo.</i> Immediate Abbreviated TICS	7.50 (0.65)	6.92 (1.31)	6.73 (0.90) [§]	
<i>B/L</i> Category fluency (animals)	19.36 (5.09)	18.36 (7.19)	17.42 (4.01)	0.781
<i>1-Mo.</i> Category fluency (animals)	19.50 (5.52)	18.92 (6.69)	18.33 (5.60)	
<i>B/L</i> BCFSI	1.11 (1.29)	2.62 (1.85)	2.67 (1.45)	0.790
<i>1-Mo.</i> BCFSI	1.11 (1.29)	3.00 (2.55)	2.38 (1.33)	
<i>B/L</i> Quality of Life (QOL)	24.43 (2.90)	22.15 (3.58)	23.25 (4.03)	0.842
<i>1-Mo.</i> Quality of Life (QOL)	24.71 (3.15)	22.75 (3.84)	22.17 (4.02)	
<i>B/L</i> Behavioral Scale	1.29 (1.27)	1.85 (2.08)	1.58 (1.08)	0.669
<i>1-Mo.</i> Behavioral Scale	1.07 (1.00)	2.08 (2.61)	2.00 (1.91)	
<i>B/L</i> ADL	2.07 (2.27)	6.62 (4.48)	5.17 (7.21)	0.834
<i>1-Mo.</i> ADL	1.29 (1.94)	6.67 (3.73)	4.08 (4.64)	

* Mail-In/Phone (MIP) was the only arm to use live phone tester.

[†] IVR N = 12 at 1-Mo. for all variables.

[‡] All correlations were significant between baseline and 1-mo scores (N = 38) at *P* values <0.001.

[§] KIO N = 11 at baseline and 1-Mo. for TICS.