



Published in final edited form as:

Alzheimer Dis Assoc Disord. 2006 ; 20(4): 232–241. doi:10.1097/01.wad.0000213862.20108.f5.

Spanish Instrument Protocol: New Treatment Efficacy Instruments for Spanish-speaking Patients in Alzheimer Disease Clinical Trials

Mary Sano, PhD^{*,†}, Susan Egelko, PhD[‡], Shelia Jin, MD, MS[§], Jeffrey Cummings, MD^{||}, Christopher M. Clark, MD[¶], Sonia Pawluczyk, MD[#], Ronald J. Thomas, PhD[§], Mario Schittini, MD, MPH^{*,†}, Leon J. Thal, MD[§], and the members of the Alzheimer's Disease Cooperative Study

*Mount Sinai School of Medicine

†The Bronx VA Medical Center, Bronx

‡Psychological Consulting Practice, New York, NY

§Alzheimer's Disease Cooperative Study, UCSD, La Jolla

||David Geffen School of Medicine at UCLA Los Angeles

#University of Southern California Keck School of Medicine, Los Angeles, CA

¶Department of Neurology, Alzheimer's Disease Center and the Institute on Aging, University of Pennsylvania, Philadelphia, PA

Abstract

Objective—To evaluate the feasibility of longitudinal assessment and the psychometric properties of both established and new outcome measures used in clinical trials of patients with dementia in a cohort of Spanish-speaking elders in the United States.

Methods—This is a prospectively collected multicenter study comparing patients with Alzheimer disease (AD) (N = 77) and elderly controls (N = 17) who are primary Spanish speakers. Spanish-speaking individuals with AD (SSI AD) were selected to represent predefined categories of impairment as determined by a Mini-Mental State Examination score. Controls were selected to approximately match by age and education (SSI C). Subjects were administered a series of Spanish translations of established outcome measures (Mini-Mental State Examination, Clinical Dementia Rating, Geriatric Dementia Scale), and Functional Assessment Staging (FAST)] and new outcome measures developed for United States in clinical trials to assess cognition, function, behavioral disturbance, and clinical global change. Half of the subjects were assessed at 1 and 2 months to evaluate reliability; all subjects were assessed at 6 and 12 months. Comparisons were made between patients and controls and between the Spanish-speaking cohort and a similar English-speaking cohort.

Results—The 12-month completion rate was 77%, with a trend toward greater impairment in those with full retention. Both established and new measures demonstrated good internal consistency and test-retest reliability in this cohort. All but one measure of cognition demonstrated excellent discriminability between AD subjects and controls. The SSI AD cohort declined significantly on measures of cognition, function, and clinical global change over the 12-month

assessment period. The SSI AD and English AD (ESI AD) cohorts declined equivalently on the most common outcomes in clinical trials of AD (delayed recall, clinical global change). Likewise, the most common behavioral changes were also similar in the ESI and SSI groups. However, the annual change was lower in SSI AD than in the ESI AD on several other measures of cognition and function.

Conclusions—These results support the recruitment of Spanish-speaking patients and the use of Spanish language translations for use in the clinical trials for AD.

Keywords

Alzheimer disease; Spanish language assessment; outcome measures; clinical trials

The search for new treatments for Alzheimer disease (AD) has identified the need for improved outcome measures specific to this disease^{1–10} and sensitive to pharmacologic intervention. Addressing this need, the Alzheimer’s Disease Cooperative Study (ADCS) initiated a multicenter trial to develop and refine novel assessment instruments targeted to key clinical domains of AD.^{1–10} The ADCS described results of this investigation, organized along the dimensions of cognition (including cognitive dysfunction in patients with severe impairment), functional activity, behavioral symptoms, and global clinical ratings.^{1–10}

As a part of an additional mandate to extend utility to as broad a population base as possible in the United States, these newly developed instruments were translated into Spanish; a cohort of Spanish-speaking individuals (SSI) was recruited in a multicenter trial, in a design which paralleled the English-speaking instrument study (ESI).^{2–10} Previously, we reported on the translation procedure, recruitment, and baseline characteristics of that SSI population.⁹ Comparisons between individuals enrolled in the SSI and ESI studies indicated that the former cohort had fewer years of formal education and were more likely to have female caregivers. Despite these differences, the intercorrelations at baseline among the established disease severity variables (EDSVs) were high and comparable in the 2 groups.

In the current report, we describe the validity (including issues of sensitivity and discriminability) and reliability of the EDSVs and newly developed assessment instruments in their application to this Spanish cohort. Although our focus is primarily on quantitative analyses of the new measures, we provide supplementary qualitative analyses to address issues of cultural disparity between the SSI and ESI cohorts in the presentation of AD behavioral symptoms.

With respect to the EDSVs, convergent validity in the SSI cohort was assessed by examining the 12-month follow-up intercorrelations, providing comparison with the ESI cohort. As a measure of reliability, internal consistency was evaluated for each of the EDSVs in the SSI, again using the ESI as a frame of comparison. The sensitivity of EDSVs to detect decline in the SSI AD cohort was studied by examining 12-month change, controlling for baseline scores and education level; comparability with the ESI AD cohort was also evaluated. With respect to the new measures, reliability was assessed via test-retest (baseline to 1 mo) in the SSI cohort, with comparison with the ESI group. Discriminability of the new measures was determined by baseline comparisons between SSI AD and control subjects. Finally, sensitivity of the new measures to decline in the SSI AD cohort was studied in a manner comparable with the EDSVs: examining 12-month change, controlling for baseline scores and education level.

The new measures in the cognitive domain included in this report assess immediate and delayed memory, attention/concentration, and executive function; each of these measures

provides a high ceiling for test performance. These measures were chosen partly for their sensitivity to subtle changes at the milder end of the AD spectrum. In addition, for those patients with severe cognitive dysfunction, a Severe Impairment Battery¹¹ is administered in lieu of the new cognitive assessment battery. Within the functional activity domain, the newly developed instrument spans a full spectrum of activities of daily living (ADL), ranging from basic ADLs (eg, bathing, toileting, eating) to instrumental ADLs (eg, shopping, food preparation, personal finance). The behavioral disturbance domain includes noncognitive symptoms common in AD (eg, depression, agitation, hallucinations, and delusions) that represent important aspects of the disease, both for the patient and caregiver. Finally, a global rating of change is included in the new battery to provide an overall metric of clinical impact.

METHODS

Subjects

Participants were recruited from 10 ADCS sites with bilingual staff and Spanish-speaking patients in their clinical populations. Subjects were participants and patients at local Alzheimer centers. Spanish speakers were recruited from the satellite sites of the ADCS center. These satellite sites are often cosponsored by NIA Alzheimer's Center grants and subjects recruited at these locations are representative of subjects likely recruited into clinical studies. Although they may not be typical of US Spanish speakers, they were, like their English-speaking counterparts, typical of those who would participate in clinical research. All subjects had to be Spanish-speaking. Bilingual patients with equal fluency in English and Spanish were not encouraged to participate in this study because they were eligible for enrollment in English language clinical trials.

Informed consent was obtained in accordance with local Institutional Review Board standards. Patients were required to meet the diagnosis of probable AD, according to NINCDS-ADRDA criteria,¹² and were stratified by entry level Mini-Mental State Examination (MMSE) into 1 of 5 categories. All but the most severely impaired patients (MMSE \leq 9) were community-dwelling and not currently taking psychoactive medication. Severely impaired patients were eligible for the study regardless of housing status (ie, might be living in a nursing homes) and use of psychoactive medications, provided the dose had remained stable for 4 weeks. All AD subjects were required to have an informant who was willing to participate by providing information about that patient. In addition, a group of elderly control subjects was also recruited. Further details of subject selection have been previously reported.²⁻⁹

Procedure

EDSVs, consisting of translated versions of the Mini-MMSE¹³; the Clinical Dementia Rating (CDR: both global score and sum of boxes)¹⁴; the Global Deterioration Scale (GDS)¹⁵; and the Functional Assessment Staging (FAST),¹⁶ were collected at baseline and at 1 year to provide a measure of disease progression with instruments commonly used in English-speaking populations. The new measures in the cognitive, functional, and behavioral disturbance domains, and also a global clinical index, were administered at baseline, and repeated at 6-month and 12-month follow-up in all AD and control subjects. Half of the group, selected at random, was also tested at 1 and 2 months following the baseline to provide test-retest reliability. Because of potential floor effect, subjects in the 2 most severe AD groups (MMSE 0 to 4 and 5 to 9) were not administered the cognitive tests. Conversely, because of anticipated ceiling effects, subjects in the mildly impaired AD group (MMSE \geq 21) were not administered the severe impairment battery. For more details on the design, refer Ferris et al.²

Measures

This report describes the study of newly developed outcome instruments, determined through research on the ESI protocol, applied to the SSI protocol. These new measures in the cognitive, functional, and behavioral disturbance domains, and also a global index, were gleaned from a larger pool of original tests and specific items and represent the current recommendations for the most useful additions to AD outcome measurement in the English-speaking population. Previously described Spanish translations for both task stimuli and verbatim instructions⁹ were used for all instruments in this study. These tests and instructions underwent extensive translation, and back translation for multiple Spanish language dialects found in US communities. The process is explained in detail elsewhere.⁹

Cognitive Domain

The *Word List Recall* consists of a 10-word list, which is presented 3 times with immediate verbal recall after each. The sum of these trials yields a Word List Recall score (range = 0 to 30). The *Delayed Recall* score (range = 0 to 10) is the number of words recalled after a brief delay. The *Cancellation Test* is a version of the well-established cancellation task that requires targeting of “either of 2 numbers” as a measure of visual attention and concentration. This particular cancellation task was identified as the most useful version out of 6 cancellation tasks in the ESI longitudinal study.³ The score, based on the number of accurate target “hits” minus the number of inaccurate hits and reminders, was coded on an ordinal scale.³ This scale, proceeding from best to worst performance, is: greater than 30 hits = 0; 24 to 30 hits = 1; 18 to 23 hits = 2; 12 to 17 hits = 3; 6 to 11 hits = 4; and 0 to 5 hits = 5.

The *Maze Test*, a measure of executive function, consists of a series of 7 mazes of increasing complexity, administered from easiest to hardest. The test is discontinued when 2 consecutive mazes are failed (cutoff of 2 consecutive failures). The task is scored as *Number of Mazes Completed* (range = 0 to 7). In addition, a *Maze Speed Score*, based on the second maze in the series, is calculated using a 6-point ordinal scale developed by Mohs (personal communication): 0 to 30 in = 0; 31 to 60 in = 1; 61 to 90 in = 2; 91 to 120 in = 3; 121 to 239 in = 4; 240 in, or failure = 5.

Severe Impairment Battery—This test, developed to assess patients with severe dementia, was adapted for multicenter clinical trial use. The instrument consists of 40 questions associated with 9 areas of cognitive function: social interaction, memory, orientation, language, attention, praxis, visuo-spatial ability, construction, and orientation to name. A total score is generated (range = 0 to 100, with higher scores reflecting better performance).

Functional Domain

Activities of Daily Living (ADCS-ADL)—This informant-based instrument consists of 23 items selected from an original pool of 38 items necessary for personal care, communicating and interacting with other people, maintaining a household, conducting hobbies and interests, and making judgments and decisions. The measure captures the ability to perform these basic activities, and the amount of assistance required. The total possible ADL score ranges from 0 to 78, with higher scores reflecting better performance.

Behavioral Disturbance Domain

Two informant-administered instruments were adapted for use in AD clinical trials and were included in the English Instrument protocol. The *Behavioral Rating Scale for Dementia (BRSD)* is a 48-item scale which encompasses a range of behavioral and psychiatric disturbance associated with dementia.^{17,6} The BRSD scale includes items relating to

depression, psychosis, agitation, and withdrawal, yielding a single summary score. The *Cohen-Mansfield Agitation Inventory (CMAI)*, a 37-item inventory, measures verbal and physical aggressive and nonaggressive agitated behaviors.^{18,7} Higher scores reflect greater behavioral symptomatology on both of these measures.

Global Measure—The *ADCS-CGIC* is an interview designed to facilitate a standardized assessment of global change in patients with AD.⁴ This instrument is completed independent of other outcome measures by a rater/clinician who is experienced in assessing patients with AD. The evaluation is based on the CGIC interview alone, with additional reference to clinical data collected at the baseline evaluation. The instrument consists of probes within cognitive, behavioral, and functional domains, which the interviewer uses at baseline and later follow-up as a basis for the global assessment of change. The interview was administered in this study by either a Spanish-speaking clinician alone or an English-speaking clinician paired with a Spanish translator unrelated to the patient or informant. At each follow-up appointment, the clinician rates overall global change, and also change in the cognitive, behavioral, and functional domains, using a 7-point Likert scale, with the following categories: marked improvement, moderate improvement, minimal improvement, no change, minimal worsening, moderate worsening, marked worsening.

Statistical Analysis

The comparisons of 12-month follow-up rates between the overall SSI and ESI cohorts, and also between SSI AD and ESI AD groups, were each assessed by χ^2 analysis (follow-up rates calculated as the proportion of the subjects assessed at baseline completing 12 mo). Specific comparisons between completing and noncompleting subjects were examined separately within the SSI AD and ESI AD samples on demographic variables and baseline EDSVs with *t* tests (χ^2 for sex comparisons).

Comparisons Between SSI and ESI Cohorts on EDSVs

The interrelationships among EDSVs at 12-month follow-up (ie, convergent validity) were computed by Pearson correlation coefficients for the SSI and ESI cohorts separately. As a measure of reliability, internal consistency coefficients were computed for all 5 EDSVs (ie, MMSE, CDR, CDR-SB, GDS, and FAST), and a total score (a summation score, MMSE score reverse coded) for SSI AD and ESI AD patient groups at baseline.

To evaluate sensitivity of the established measures to 12-month change within the Spanish-speaking cohort, SSI AD and control subjects were compared on 12-month EDSV scores via analysis of covariances (ANCOVAs), controlling for baseline scores and education level. To evaluate differences in sensitivity of the established measures to 12-month changes between SSI AD and ESI AD patients, a series of ANCOVAs were performed on 12-month EDSVs, controlling for baseline scores and education level.

Comparisons Between SSI and ESI Cohorts on New Measures

To assess reliability of the new instruments, we calculated Pearson test-retest reliability coefficients between baseline and 1-month scores (one half the SSI and ESI cohorts had been randomly assigned to this 1-month evaluation).

The evaluation of sensitivity of the new cognitive and functional measures to Spanish-speaking patients was addressed by: (a) *t* test comparisons between SSI AD patient and control performance at baseline, (b) paired *t* test comparisons between SSI AD patient performance at baseline and 1 year to detect decline, and (c) a series of ANCOVAs between SSI AD and ESI AD groups, using baseline score and level of education as covariates, on 12-month scores. Within the behavioral disturbance domain, we did not expect decline

within the SSI AD group. However, we evaluated differences between SSI AD and ESI AD groups with respect to change over time through ANCOVAs between SSI AD and ESI AD groups, using baseline score and level of education as covariates, on 12-month BRSD and CMAI scores. Also, we conducted a qualitative comparison between SSI and ESI AD symptoms of agitation by identifying CMAI items endorsed by at least 50% of caregivers in each of these two groups. A Wilcoxon signed-rank test was used to assess the short-term reliability of the ADCS-CGIC between 1 and 2 months for SSI AD patients. To examine differences between the SSI and ESI AD patient groups in ADCS-CGIC change over 12 months, a χ^2 test was conducted on 12-month ratings for the completer sample, categorized as: “improved or no change” versus “worsened.”

RESULTS

Comparisons Between SSI and ESI Cohorts on EDSVs: Issues of Retention, Validity, Reliability, and Sensitivity

A total of 72 out of 94 SSI group (77%) reported in the initial paper completed the 1-year study, as compared with 275 of 306 (90%) ESI group, a difference that was statistically significant ($\chi^2 = 9.90$; $P = 0.002$). A lower completion rate for Spanish-speaking participants was also evident when the comparison was restricted to subjects with AD: 74% (SSI) versus 88% (ESI) ($\chi^2 = 7.75$, $P = 0.005$). Table 1 presents the baseline demographic and EDSVs for SSI and ESI AD patients, categorized according to whether or not they completed the 1-year study. As previously reported, the SSI cohort has significantly fewer years of formal education and is more likely to have a female caregiver.⁹ Comparisons between completer and noncompleter groups were conducted within the SSI and ESI AD samples. Although none of these group comparisons were significant, several trends were noted. SSI AD completers had a higher rate of female informants and were more impaired at baseline than SSI AD noncompleters. Conversely, the ESI AD completers were marginally less likely to have female informants and were less impaired at baseline than ESI AD noncompleters.

Convergent Validity

The interrelationships among EDSVs at 12-month follow-up were examined via Pearson correlation coefficients and were found to be relatively high, ranging from 0.79 to 0.97 for the SSI cohort (the baseline correlation coefficients for both the SSI and ESI cohorts had been previously presented in an earlier manuscript).⁹ These 12-month intercorrelations are reported in Table 2, along with comparable values from the ESI cohort.

Internal Reliability

Internal consistency coefficients, as evaluated by Cronbach α , were also uniformly high for the SSI ($n = 77$) AD patients and equivalent to the ESI ($n = 242$) AD patients on standardized scores of EDSVs at baseline. On the MMSE, α coefficients were 0.936 and 0.933 in the SSI and ESI AD groups, respectively; likewise on the CDR, α coefficients were 0.939 and 0.928 in the SSI and ESI AD groups, respectively. A Total Score, on the basis of a composite of all 5 established measures (MMSE, CDR, CDR-SB, GDS, and FAST), was also found to have high consistency for both SSI (Cronbach $\alpha = 0.96$) and ESI AD patients (Cronbach $\alpha = 0.95$).

Sensitivity to 12-month Change

Twelve-month change scores were calculated for the EDSVs for AD patients and controls, as presented in Table 3. As expected, SSI patients declined to a greater extent than SSI controls, with the exception of the CDR. In a series of ANCOVAs, controlling for baseline

scores and education level, these significant differences in 12-month scores between SSI AD patients and controls were as follows: MMSE [$F = 4.61(1,60)$; $P = 0.036$], CDR-SB [$F = 4.01(1,62)$; $P = 0.050$], GDS [$F = 10.42(1,68)$; $P = 0.002$], and FAST [$F = 10.76(1,68)$; $P = 0.002$].

Despite similarly high internal consistency and intertest correlations at 1-year follow-up, the SSI AD patients demonstrated significantly less decline (ie, from 36% to 75% less decline) over the 1-year period than the ESI AD patients on the MMSE, CDR, and CDR-SB (Table 3). ANCOVAs, controlling for baseline scores and education level, identified significant differences in 12-month scores between SSI and ESI AD patients for MMSE ($F = 7.14[1,228]$; $P = 0.008$), CDR ($F = 10.68[1,253]$; $P = 0.001$), and CDR-SB ($F = 4.88[1,253]$; $P = 0.028$). Differences in 12-month scores between the SSI and ESI AD patient groups on the GDS approached but did not reach significance. Changes on the FAST from baseline to 12 months were small in both patient groups and not significantly different from each other.

Comparisons Between SSI and ESI Cohorts on New Measures: Issues of Reliability, Discrimination, and Sensitivity

Test-retest Reliability of New Measures—Table 4 presents test-retest reliability data on the new instruments generated between baseline and 1-month scores (one half of the original cohort evaluated at baseline was randomly assigned to this extra 1-month reliability assessment). Although both Pearson and Spearman correlations were calculated for all measures, only the Pearson correlations are presented in this report as in all cases the 2 computations yielded equivalent findings. Prior test-retest reliability findings based on the English cohort had also reported only the Pearson values, with the exception of BRSD in which case Spearman ρ had been reported. All test-retest correlations were highly significant and generally equivalent between the SSI and ESI cohorts. However, the test-retest correlation for the SIB was notably lower in the SSI, as compared with the ESI, cohort although both were significant.

Discrimination Between SSI AD and Controls on the New Measures—The ability to discriminate between SSI AD and control subjects at baseline was assessed through t test comparisons, the resulting P values are reported in Table 5. Overall, the new measures in the cognitive, functional, and behavioral disturbance domains succeeded in differentiating patient and control groups. The one exception was the Maze Speed Score, which failed to differentiate the 2 groups owing to slow performance by the SSI controls.

Sensitivity: Detecting 12-month Changes in SSI AD; Comparability of 12-month Changes in SSI and ESI AD Patients

Cognitive Domain and SIB—The P values based on paired t tests between SSI AD patients at baseline and 12-month follow-up are presented in Table 5. Performance on the cognitive battery deteriorated to a significant degree over this 12-month period on Word List Learning, Delayed Recall, and the SIB. It is noteworthy that the SSI control group demonstrated deterioration on the Word List Recall and Delayed Recall at 12 months. However, the 12-month total scores on both tests remained significantly higher than the AD group. SSI AD, but not SSI C, scores also worsened on Cancellation, Number of Mazes Completed, and Maze Speed Score, although not to the level of significance (vs. the SSI controls, who showed modest improvement on all 3).

A series of ANCOVAs, using baseline score and level of education as covariates, were performed between SSI and ESI AD patients on 12-month cognitive scores. The P values corresponding to the F tests of these ANCOVAs are reported in the last column of Table 5. On the Word List Learning and SIB, the SSI AD patients declined to significantly less than

the ESI AD patients (ie, 39% and 45% less, respectively). This effect is similar in size to the difference between SSI and ESI AD patients seen on several of the EDSVs in comparable tests of 12-month change. However, there were no significant differences between SSI and ESI AD patients in 12-month change on Delayed Recall, Cancellation, Number of Mazes Completed, or Maze Speed Score.

Functional Domain: ADCS-ADL—The total ADL score significantly deteriorated over 12 months in the SSI AD patient group. An ANCOVA between the 12-month ADL scores of SSI and ESI AD patients, with baseline scores and education level as covariates, demonstrated significantly less functional deterioration (50% less) in the SSI AD patient group. Interestingly, the SSI C group also demonstrated lower scores over 12 months, but as with the Word List described above, the ADL 12-month total score remained well above those of the AD group.

Behavioral Disturbance Domain—The BRSD and CMAI total scores did not change over the 12-month period in the SSI AD patient group. In contrast to the cognitive and functional domains, there were no significant differences in the degree of BRSD and CMAI changes over 12 months between ESI and SSI AD patients, as assessed by ANCOVAs on 12-month scores, covarying baseline scores and education level. We also addressed the issue of possible qualitative differences in behavioral disturbance symptoms in AD patients drawn from 2 divergent cultures. Table 6 provides a list of the CMAI items identified by at least 50% of caregivers of the SSI and ESI patients. Notably, caregivers in the 2 cohorts identified the identical 5 items as the most frequent behavioral manifestations of agitation in the AD patients for which they were informants, with the same rank ordering of the first 4 symptoms.

ADCS-CGIC—Test-retest reliability was assessed for this nonparametric clinical rating scale using Wilcoxon signed-rank test between 1-month and 2-month scores for SSI AD patients. At 1 month, 100% of SSI AD patients were scored in 1 of the 3 central categories: “minimal worsening” (25%), no change (75%), or “minimal improvement” (0%). At 2 months, this figure had dropped to 83%: minimal worsening (24%), no change (59%), and minimal improvement (0%). The mean change in CGIC ratings for SSI AD patients from month 1 to month 2 was not significantly different ($P = 0.81$ for signed ranks test), showing good short-term reliability (the comparable mean change in CGIC ratings for ESI AD patients from months 1 to 2 was also insignificant, $P = 0.20$). At 12 months, SSI AD patients showed significantly more worsening compared with SSI C (78% vs. 7%) as expected. To examine differences between the SSI and ESI AD patient groups in change over 12 months, a χ^2 test was conducted on 12-month ratings for the completer sample, categorized as: improved or no change versus worsened. A total of 34 out of 59 (69%) of the SSI AD patients versus 149 out of 191 (78%) of the ESI AD patients were rated as worsened, a trend for greater decline in the SSI AD patients which was not significantly different ($\chi^2 = 1.60$, $df = 1$, $P = 0.21$) (Table 7).

DISCUSSION

This report illustrates the feasibility of assessment and psychometric properties of both established and new outcomes used in clinical trials of patients with dementia in a cohort of Spanish-speaking elders in the United States. More than three fourths of the cohort completed the 1-year trial. Although lower than the completion rate of the ESI cohort, this was far superior to the completion rates in most commercial treatment trials.² Although increasing dementia severity was associated with noncompletion among the ESI cohort, this was not the case among the SSI cohort.

In general, both established and new outcomes demonstrate excellent convergent validity, internal consistency, and test-retest reliability, with comparability to indices in the English-speaking cohort. Established and new measures of cognition and function differentiated between SSI control and AD subjects in all but a single measure. The anomaly, Mazes, a cognitively complex, timed test, was characterized by poor performance in the control group, perhaps related to low level of formal education. It is noteworthy that others have reported that nondemented Spanish-speaking elders are particularly likely to perform poorly on nonverbal tests of cognition.¹⁹

SSI AD patients showed cognitive and functional deterioration over 12 months on all but the one measure (ie, Mazes) which had poor discriminability as described above. The decline in this cohort provides confidence in our diagnosis of AD, a progressively degenerative disease. The one-year change in cognitive and functional measures was smaller in the SSI AD group than in the ESI AD group for several measures, both established and new. However, the deterioration in delayed recall, which is considered a hallmark cognitive impairment in AD, was comparable in the SSI and ESI groups. Others have described the importance of verbal memory in diagnostic specificity among Spanish speakers.²⁰ To our knowledge, this is the first systematic assessment of change in scores over time in an AD cohort. Because memory scores reflect the highest portion of assessment items in cognitive scales used in clinical trials, it is encouraging to see that this measure was equivalently sensitive to change as an English-speaking cohort.

Two behavioral measures, BRSD and CMAI, were examined and both showed greater symptomatology in the SSI AD group compared with the SSI control group, as would be expected. There was no systematic increase in behavioral disturbance over time, which is not unexpected since the entry criteria required that subjects have no significant use of medications for behavior except in the most impaired strata. This trend matches the findings in the ESI cohort. Since assessments were performed at relatively long intervals, subjects with symptoms were likely to be treated and continue in the trial. To further evaluate the cohorts on behavioral disturbance, the rank order of the most commonly endorsed behaviors on the CMAI was compared between the ESI and SSI AD groups and found to be nearly identical for both the ESI and SSI, strengthening the impression of comparability of this measure and utility of this instrument in these cohorts. It is very encouraging that the measurements of behavioral disturbance were sufficiently robust that differences between Spanish and English populations in several dimensions of caregiving did not alter the report. Of note, measures of restlessness and irritability were the most commonly endorsed items, and it is well known that they reflect disturbances that occur across dementia severity.

The rationale for test selection and item inclusion were based on work in English-speaking cohorts, with the goal to identify tools to assess efficacy in clinical trials and not for diagnostic purposes.² Specifically, it was our intent that single tools with an identical scoring metric be developed in the most change-sensitive and reliably administered domains and translated to allow clinical trials to enroll across a broad class of Spanish-speaking patients living in the United States. It is important to note that tests were not developed for all cognitive domains. For example, visual-spatial and constructional tasks were not included in this battery. Such tests are not often included in multicenter clinical trials, perhaps because scoring is difficult to standardize.

Each domain evaluated in this study is unique and our expectation is that not all domains will be used in any given trial, but will be selected on the basis of the specific type of efficacy related to that intervention. With this specificity in mind, we have presented raw scores for each measure in each domain in this manuscript rather than collapse across domains, thereby providing the reader with maximal information about these tools. Effect

size estimates, presented in Tables 3 and 5, were calculated by dividing the 12-month change score by baseline SD, and ranged between 0.2 and 0.5 in the Spanish AD group over 1-year change (in all areas except the behavioral disturbance domain). The effect sizes for memory measures were moderate and are large enough to see a change with therapeutic intervention with sample sizes typical of those used in AD trials.

The ADCS-CGIC was sensitive to change in the SSI AD group and discriminated between patients and controls. The change in this cohort is comparable to what has been observed in the ESI, suggesting that the clinical sensitivity may be less affected by cultural factors. Other reports have also indicated that the clinical judgment of change in cognitively impaired populations may be accurate even in the presence of complicating factors such as medical comorbidities and preexisting mental disabilities.²¹ This is an encouraging result since clinical global measures are often primary outcomes in clinical trials. Several limitations are apparent when considering these results. First, SSI in the United States are a heterogeneous population and this work does not explore the details of this heterogeneity. Additionally, the mean education level in the SSI is significantly lower than in ESI and while education was examined as a covariate in many analyses, it undoubtedly reflects a lifetime of experience and opportunity which, although not clearly understood, is likely to impact performance in late life. It is noteworthy that change in the SSI AD group was somewhat less than in the ESI AD group. Also, unfortunately, the control group was too small to provide normative data in this particular study. However, the data did allow a comparison to a group with similar demographic features and identified certain tasks, which were not sensitive to disease because of poor performance in the non-AD Spanish-speaking sample.

Future studies with this and other cohorts may shed light on the interactions important variables that play in cognitive and functional performance in old age and this study is limited to United States-dwelling Spanish-speaking subjects, largely recent immigrant populations living in the United States. Our results do not address the use of these instruments in Spanish-speaking countries with native populations; further research will be needed to evaluate the utility of these measures in such settings and provide some index of acculturation (eg, length of time in the United States.). Further, these instruments, while targeted to monitoring temporal changes in AD populations, have not been designed to replace diagnostic instruments for AD or other dementias.

Overall, the results suggest that the SSI can participate in AD clinical trials and can be followed with excellent rates of retention. Although this cohort demonstrated lesser level of decline on some cognitive and functional measures, performance on memory measures and clinical global outcomes were quite comparable with that seen in an equivalent English-speaking cohort. These results support the recommendation of (1) active recruitment of non-English-speaking individuals into clinical trials using well-translated instruments; and (2) the use of measures of memory, function, behavior, and global clinical ratings in trials with these populations.

Acknowledgments

Supported by U01 AG10483, AG05138.

References

1. Ferris, SH.; Mackell, JA., editors. Alzheimer Dis Assoc Disord. New York: Lippencott-Raven; 1997. A Multicenter Evaluation of New Treatment Efficacy Measures for Alzheimer Disease. Results From the Instrument Development Project of the Alzheimer Disease Cooperative Study.

2. Ferris SH, Mackell JA, Mohs R, et al. A multicenter evaluation of new treatment efficacy instruments for Alzheimer's disease clinical trials: overview and general results. *Alzheimer Dis Assoc Disord.* 1997; 11(suppl 2):S1–S12. [PubMed: 9236947]
3. Mohs RC, Knopman D, Peterson RC, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's disease assessment scale that broaden its scope. *Alzheimer Dis Assoc Disord.* 1997; 11(suppl 2):S13–S21. [PubMed: 9236948]
4. Schneider LS, Olin JT, Doody RS, et al. Validity and reliability of the Alzheimer's Disease Cooperative Study—clinical global impression of change. *Alzheimer Dis Assoc Disord.* 1997; 11(suppl 2):S22–S32. [PubMed: 9236949]
5. Galasko D, Bennett D, Sano M, et al. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. *Alzheimer Dis Assoc Disord.* 1997; 11(suppl 2):S33–S39.
6. Patterson MB, Mack JL, Mackell JA, et al. A longitudinal study of behavioral pathology across five levels of dementia severity in Alzheimer's disease: the CERAD behavior rating scale for dementia. *Alzheimer Dis Assoc Disord.* 1997; 11(suppl 2):S40–S44. [PubMed: 9236951]
7. Koss E, Weiner M, Ernesto C, et al. Assessing patterns of agitation in Alzheimer's disease patients with the Cohen-Mansfield agitation inventory. *Alzheimer Dis Assoc Disord.* 1997; 11(suppl 2):S45–S50. [PubMed: 9236952]
8. Schmitt FA, Ashford W, Ernesto C, et al. The Severe Impairment Battery: concurrent validity and the assessment of longitudinal change in Alzheimer's disease. *Alzheimer Dis Assoc Disord.* 1997; 11(suppl 2):S51–S56. [PubMed: 9236953]
9. Sano M, Mackell JA, Ponton M, et al. The Spanish instrument protocol: design and implementation of a study to evaluate treatment efficacy Instruments for Spanish-speaking patients with Alzheimer's disease. *Alzheimer Dis Assoc Disord.* 1997; 11(suppl 2):S57–S64. [PubMed: 9236954]
10. Mackell JA, Ferris SH, Mohs R, et al. Multicenter evaluation of new instruments for Alzheimer's disease clinical trials: summary of results. *Alzheimer Dis Assoc Disord.* 1997; 11(suppl 2):S65–S69. [PubMed: 9236955]
11. Saxton J, McGonigle-Gibson K, Swihart A, et al. Assessment of the severely impaired patient: description and validation of a new neuropsychological test battery. *Psychol Assess.* 1990; 2:298–303.
12. McKhann G, Drachmann D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology.* 1984; 34:939–944. [PubMed: 6610841]
13. 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]
14. Berg L. Clinical dementia rating (CDR). *Psychopharmacol Bull.* 1988; 24:637–639. [PubMed: 3249765]
15. Reisberg B, Ferris SH, de Leon MJ, et al. The global deterioration scale (GDS). *Psychopharmacol Bull.* 1988a; 24:661–663. [PubMed: 3249768]
16. Reisberg B. Functional assessment staging (FAST). *Psychopharmacol Bull.* 1988; 24:653–659. [PubMed: 3249767]
17. Tariot PN, Mack JI, Patterson MB, et al. The CERAD behavior rating scale for dementia (BRSD). *Am J Psychiatry.* 1995; 152:1349–1357. [PubMed: 7653692]
18. Cohen-Mansfield J. Assessment of disruptive behavior/agitation in the elderly: function, methods, and difficulties. *J Geriatr Psychiatry Neurol.* 1995; 8:52–60. [PubMed: 7710649]
19. Jacobs DM, Sano M, Albert S, et al. Cross-cultural neuropsychological assessment: a comparison of randomly selected, demographically matched cohorts of English- and Spanish-speaking older adults. *J Clin Exp Neuropsychol.* 1997; 19:331–339. [PubMed: 9268808]
20. Mungas D, Reed BR, Tomaszewski Farias S, et al. Criterion-referenced validity of a neuropsychological test batter: equivalent performance in elderly Hispanics and non-Hispanic Whites. *J Int Neuropsychol Soc.* 2005; 11:620–630. [PubMed: 16212690]
21. Sano M, Aisen PS, Dalton AJ, et al. Assessment of aging individuals with Down syndrome in clinical trials: results of baseline measures. *J Policy Practice Intell Disabilities.* 2005; 2:126–138.

TABLE 1

Comparison of Demographic and EDSVs Between AD Patients who Either Completed or did not Complete the 12-month Protocol in the Spanish-speaking and English-speaking Cohorts

	Spanish-speakers: AD (n = 77)		English-speakers: AD (n = 242)	
	12-month Completers (n = 57)	12-month Noncompleters (n = 20)	12-month Completers (n = 213)	12-month Noncompleters (n = 29)
Age	71.6 (10.5)	76.8 (7.9)	71.8 (9.0)	76.1 (8.6)
% Female	70	65	62	55
Education	7.6 (5.1)	7.4 (4.7)	13.2 (2.9)	12.9 (2.9)
Informant, % female	88	75	57	59
Baseline MMSE	12.3 (8.5)	13.3 (6.6)	13.2 (7.9)	10.1 (7.5)
Baseline CDR	1.7 (0.7)	1.7 (0.7)	1.7 (0.6)	1.9 (0.7)
Baseline CDR-SB	10.2 (4.2)	9.5 (4.0)	10.4 (3.8)	11.2 (3.7)
Baseline GDS	5.1 (1.0)	4.8 (0.9)	5.1 (0.8)	5.0 (0.7)
Baseline FAST	5.4 (1.2)	5.1 (1.0)	5.4 (1.1)	5.5 (1.1)

Values are given as group mean (SD) or percent (see text for details of specific comparisons between groups).

TABLE 2

Correlation Coefficients at 12 Months Among EDSVs^{*}: Spanish-speaking Subjects (Values From the English-speaking Subjects Appear in Parentheses[†])

	CDR	CDR-SB	GDS	FAST (n = 72)
MMSE (n = 64)	-0.89 (-0.91)	-0.90 (-0.94)	-0.83 (-0.88)	-0.79 (-0.89)
CDR (n = 66)		0.97 (0.98)	0.90 (0.91)	0.87 (0.92)
CDR-SB (n = 66)			0.92 (0.94)	0.90 (0.95)
GDS (n = 72)				0.96 (0.96)

* $P < 0.001$.

[†] English-speaking n = 241–275.

MMSE: lower score = greater impairment; CDR, CDR-SB, GDS: higher score = greater impairment.

TABLE 3

Mean 12-month Decline (SD) on EDSVs: Spanish-speaking and English-speaking Cohorts

Variable/Group	Baseline Scores (n)	12-month Change (SD)	Effect Size: 12-Month Change/ Baseline SD	P: SSI AD vs. ESI AD Change [*]
MMSE				
SSI Subjects				
Controls	26.93 (15)	0.7 (1.6)		0.008
Patients	13.51 (49)	1.6 (3.8)	0.20	
ESI Subjects				
Controls	29.42 (58)	0.1 (1.2)		0.48
Patients	13.77 (183)	3.8 (4.2)	0.48	
CDR				
SSI Subjects				
Controls	0.0 (15)	0.1 (0.2)		0.001
Patients	1.71 (51)	0.1 (0.4)	0.16	
ESI Subjects				
Controls	0.0 (62)	0.0 (0.1)		0.65
Patients	1.69 (206)	0.4 (0.5)	0.65	
CDR-SB				
SSI Subjects				
Controls	0.27 (15)	0.0 (0.5)		0.028
Patients	10.08 (51)	1.4 (2.1)	0.36	
ESI Subjects				
Controls	0.04 (62)	0.0 (0.3)		0.59
Patients	10.39 (206)	2.2 (2.2)	0.59	
GDS				
SSI Subjects				
Controls	1.13 (15)	-0.1 (0.4)		0.055
Patients	5.09 (57)	0.3 (0.7)	0.30	
ESI Subjects				
Controls	1.19 (62)	-0.1 (0.5)		0.73
Patients	5.09 (213)	0.6 (0.6)	0.73	
FAST				
SSI Subjects				
Controls	1.13 (15)	0.0 (0.5)		0.202
Patients	5.36 (57)	0.4 (0.8)	0.33	
ESI Subjects				
Controls	1.11 (62)	-0.0 (0.3)		0.47
Patients	5.38 (213)	0.5 (0.8)	0.47	

Note: For all measures, a positive score reflects a decline in performance (MMSE change scores have been reverse coded).

* *P* values reported for series of ANCOVAs: differences between Spanish-speaking and English-speaking AD patients on 12-month scores, covarying baseline scores and education level.

TABLE 4
 Test-retest Reliability for Spanish and English-speaking Subjects on New Measures: Baseline to 1-month Pearson Correlations

Variable	Spanish-speakers			English-speakers		
	r	n	P	r	n	P
Word List Learning	0.925	26	<0.0001	0.910	100	<0.0001
Delayed Recall	0.877	26	<0.0001	0.933	100	<0.0001
Cancellation	0.896	26	<0.0001	0.897	98	<0.0001
No. Mazes Completed	0.888	26	<0.0001	0.791	97	<0.0001
Maze Speed Score	0.844	26	<0.0001	0.806	97	<0.0001
SIB	0.546	28	0.0026	0.797	96	<0.0001
ADL	0.925	39	<0.0001	0.980	146	<0.0001
BRSD	0.916	39	<0.0001	0.826	145	<0.0001
CMAI	0.810	39	<0.0001	0.866	146	<0.0001

\$watermark-text

\$watermark-text

\$watermark-text

TABLE 5

Results of New Measures for 12-month Completers

Variable/Group	Baseline Scores (n)	12-month Change Score (SD)	Effect Size: 12-month Change/Baseline SD	P: SSI C vs. AD at Baseline	P: SSI AD 12-month Change	P: SSI AD vs. ESI AD Change
Cognitive Domain and SIB						
Word List Learning (– = decline)						
SSI subjects						
Controls	29.68 (15)	-4.13 (3.82)				
Patients	15.52 (31)	-2.74 (3.52)	0.43	0.001	0.000	0.046
ESI subjects						
Controls	30.87 (62)	0.87 (3.98)				
Patients	14.31 (113)	-4.48 (4.85)	0.68			
Delayed Recall (– = decline)						
SSI subjects						
Controls	7.93 (15)	-1.47 (1.85)				
Patients	2.03 (32)	-1.00 (1.63)	0.48	0.001	0.002	0.783
ESI subjects						
Controls	8.07 (61)	0.41 (1.16)				
Patients	1.52 (116)	-0.70 (1.45)	0.38			
Cancellation (+ = decline)						
SSI subjects						
Controls	1.68 (15)	-0.27 (0.80)				
Patients	3.31 (29)	0.21 (0.94)	0.17	0.001	0.246	0.490
ESI subjects						
Controls	1.00 (62)	-0.18 (0.74)				
Patients	3.01 (104)	0.44 (0.81)	0.33			
No. Mazes Completed (– = decline)						
SSI subjects						
Controls	6.27 (15)	0.13 (0.83)				
Patients	4.60 (30)	-0.50 (1.41)	0.28	0.001 *	0.062	0.586
ESI subjects						
Controls	6.65 (62)	0.24 (0.69)				

\$watermark-text

\$watermark-text

\$watermark-text

Variable/Group	Baseline Scores (n)	12-month Change Score (SD)	Effect Size: 12-month Change/Baseline SD	P: SSI C vs. AD at Baseline	P: SSI AD 12-month Change	P: SSI AD vs. ESI AD Change
Patients	4.43 (106)	-0.18 (1.91)	0.10			
Maze Speed Score (+ = decline)						
SSI Subjects						
Controls	0.13 (15)	-0.13 (0.35)				
Patients	0.40 (30)	0.23 (1.46)	0.23	0.200*	0.387	0.632
ESI subjects						
Controls	0.00 (62)	0.00 (0.00)				
Patients	0.35 (106)	0.39 (1.57)	0.41			
SIB (- = decline)						
SSI subjects						
Controls	—	—				
Patients	65.89 (46)	-13.07 (32.38)	0.42	—	0.009	0.026
ESI subjects						
Controls	—	—				
Patients	66.04 (180)	-23.72 (25.96)	0.84			
Functional Domain ADCS-ADL (- = decline)						
SSI subjects						
Controls	74.60 (15)	-4.20 (4.89)				
Patients	35.70 (59)	-5.36 (8.88)	0.17	0.001*	0.001	0.012
ESI subjects						
Controls	73.52 (62)	0.18 (2.71)				
Patients	39.91 (210)	-10.82 (10.19)	0.36			
Behavioral Disturbance Domain						
BRSB (+ = decline)						
SSI subjects						
Controls	3.60 (15)	-0.80 (5.00)				
Patients	37.55 (58)	-3.93 (21.00)	0.16 [†]	0.001*	0.159	0.678
ESI subjects						
Controls	5.05 (62)	-1.52 (4.71)				
Patients	29.39 (207)	1.76 (16.45)	0.09			

Variable/Group	Baseline Scores (n)	12-month Change Score (SD)	Effect Size: 12-month Change/Baseline SD	P: SSI C vs. AD at Baseline	P: SSI AD 12-month Change	P: SSI AD vs. ESI AD Change
CMAI (+ = decline)						
SSI subjects						
Controls	1.00 (15)	-0.20 (2.91)				
Patients	33.58 (59)	-0.02 (22.83)	0.00 [‡]	0.001*	0.996	0.637
ESI subjects						
Controls	3.63 (62)	-0.86 (3.56)				
Patients	26.11 (212)	1.15 (15.39)	0.04			

* Unequal variance estimate used in calculating *t* test.

[‡] Direction of 12-month change is improved.

TABLE 6

Frequency of Most Common Agitated Behaviors Identified by the CMAI Among Spanish and English AD Patients

Most Common Agitated Behaviors (%)			
Spanish AD patients		English AD Patients	
Repeating sentences	85.5	Repeating sentences	80.9
Restlessness	57.9	Restlessness	63.5
Uncooperativeness	57.1	Uncooperativeness	53.1
Complaining	56.6	Pacing	49.0
Pacing	56.0	Complaining	48.1

\$watermark-text

\$watermark-text

\$watermark-text

TABLE 7
 Frequency Distribution (Percentage) of CGIC Change in Spanish-speaking Cohort From Baseline Over 12 Months for AD and Control Subjects

Follow-up	Marked Improvement	Moderate Improvement	Minimal Improvement	No Change	Minimal Worsening	Moderate Worsening	Marked Worsening
<i>AD subjects</i>							
1mo (n = 28)	0	0	0	75	25	0	0
2mo (n = 29)	3	3	0	59	24	10	0
6mo (n = 59)	0	0	8	34	41	15	2
12mo (n = 49)	0	0	0	31	29	27	14
<i>Control subjects</i>							
1mo (n = 9)	0	0	0	100	0	0	0
2mo (n = 9)	0	0	0	89	11	0	0
6mo (n = 16)	0	0	6	94	0	0	0
12mo (n = 15)	0	0	0	93	7	0	0