Published in final edited form as:
Int Psychogeriatr. 2013 July ; 25(7): 1115–1123. doi:10.1017/S1041610213000367.

CERAD Practice Effects and Attrition Bias in a Dementia Prevention Trial

Melissa Mathews, Ph.D.[1], **Erin Abner, M.P.H.**[1,3], **Allison Caban-Holt, Ph.D.**[1,2], **Richard Kryscio, Ph.D.**[1,3], and **Frederick Schmitt, Ph.D.**[1,2,4]

[1]Sanders-Brown Center on Aging, University of Kentucky, Lexington, KY, USA

[2]Department of Behavioral Science, University of Kentucky, Lexington, KY, USA

[3]Department of Statistics and Biostatistics, University of Kentucky, Lexington, KY, USA

[4]Department of Neurology, University of Kentucky, Lexington, KY, USA

Abstract

Background—The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) set of tests is frequently used for tracking cognition longitudinally in both clinical and research settings. Repeated cognitive assessments are an important component in measuring such changes; however, practice effects and attrition bias may obscure significant clinical change over time. The current

Corresponding Authors Contact Information: Frederick Schmitt, Ph. D., 303 Sanders-Brown Center on Aging, University of Kentucky, 800 S. Limestone Street, Lexington, KY 40536-023, USA, fascom@uky.edu; voice: 859.218.5051; fax: 859.323.1772.
Co-Authors Full Names and Details: Melissa Mathews, Ph.D., 303A Sanders-Brown Center on Aging, University of Kentucky, 800 S. Limestone Street, Lexington, KY 40536-023, USA, melissa.mathews@uky.edu; voice: 859.257.1412 x 240; fax: 859.323.1772
Erin Abner, M. P. H., 207 Sanders-Brown Center on Aging, University of Kentucky, 800 S. Limestone Street, Lexington, KY 40536-023, USA, elabne0@email.uky.edu; voice: 859.257.1412 x 489; fax: 859.323.1772
Allison Caban-Holt, Ph.D., 303B Sanders-Brown Center on Aging, University of Kentucky, 800 S. Limestone Street, Lexington, KY, USA, 40536-023, amcaba2@email.uky.edu; voice: 859.257.1412 x 322; fax: 859.323.1772
Richard Kryscio, Ph.D., 230 Sanders-Brown Center on Aging, University of Kentucky, 800 S. Limestone Street, Lexington, KY 40536-023, USA, kryscio@email.uky.edu; voice: 859.257.4064; fax: 859.323.1772

Conflict of Interest: Dr. Melissa Mathews: NIA Grant (Post-doctoral researcher)
Erin Abner, M.P.H: Alltech (SELPLEX clinical trial) Statistician
NIA – ADC Grant (Data Manager, Bio-Data Management Core)
NIA – PREADVISE Grant (Data Manager, Statistician)
NIA – R01 Grant (Data Manager)
Dr. Allison Caban-Holt
NIA – PREADIVSE Grant (Co-Investigator)
NIA-ADCS Home-Based Assessment (Site Investigator)
NICHD- R01 Frontal Aging and Down Syndrome (Co-Investigator)
Dr. Richard Kryscio: Alltech Data Safety Monitoring Board (SELPLEX clinical trial)
NIA (2 R01 Grants: co-PI)
NIA- ADC Grant (Leader, Bio-Data Management Core)
NCRR-CTSA (Leader, BERD Key Function)
NCRR – COBRE (Leader, Biostat/Bioinformatics Core)
NICHD, NIA (R21, R03, PPG), NIDA, NINDS: co-invest.
Dr. Frederick Schmitt
NIH: R01 AG038651; R01HD064993; P30 AG028383; R01AG019241

Description of Authors' Roles: Melissa Mathews – developed research question, data interpretation, manuscript preparation
Erin Abner – developed research question, data analysis and interpretation, manuscript revisions
Allison Caban-Holt – data acquisition, data interpretation, and manuscript revisions
Richard Kryscio – developed research question, data analysis and interpretation, manuscript revisions, study supervision
Frederick Schmitt – developed research question, data interpretation, manuscript revisions, study supervision

Supplementary material (Supplemental Figure 1and Supplemental Table1) is available.

study sought to examine the presence and magnitude of practice effects and the role of attrition bias in a sample of cognitively normal older men enrolled in a prevention trial.

Method—Participants were grouped according to whether they completed five years of follow-up (n = 182) or less (n = 126). Practice effects were examined in these participants as a whole (n = 308) and by group.

Results—Findings indicate that moderate practice effects exist in both groups on the CERAD T-score and that attrition bias likely does not play a contributing role in improved scores over time.

Conclusion—The current study provides additional evidence and support for previous findings that repeated cognitive assessment results in rising test scores in longitudinally collected data and demonstrates that these findings are unlikely to be due to attrition.

Keywords

practice effects; CERAD; attrition bias; aging

Introduction

Many clinical trials and epidemiologic studies of neurodegenerative disease use repeated cognitive assessments for case ascertainment and quantifying treatment effects. However, recognition of material or familiarity with testing procedures may result in participants maintaining scores above recommended cut-points for impairment, thus escaping detection. Further complicating this picture, practice effects could differ by measure (Caban-Holt *et al.*, 2005; Hickman *et al.*, 2000), time between administrations (Cooper *et al.*, 2001), baseline cognitive status (Cooper *et al.*, 2004), and presence of neuropathology in the absence of dementia (Galvin *et al.*, 2005). A recent meta-analysis of approximately 1600 different effect sizes (Calamia *et al.*, 2012) found multiple variables accounting for the degree of change in cognitive test scores over time. More specifically, they found that age is negatively correlated with practice gain over time such that older adults benefit less from repeated administrations. Length of the test-retest interval and its relationship to practice effects also varied by measure, with some practice effects being eliminated after two-to-three years and others taking as long as seven years to extinguish. Effects of alternate forms were found to be inconsistent. For example, alternate forms were effective in reducing practice effects on verbal list learning measures. Other tests, such as verbal fluency, continued to show practice effects despite alternate form usage.

The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) neuropsychological battery (Morris *et al.*, 1989) has been used in epidemiological studies and clinical trials internationally (Fillenbaum *et al.*, 2008). As a result, practice effects in this battery may have tremendous significance. However, only a few studies have examined practice effects on the CERAD battery (Burkhart *et al.*, 2011; Rosetti *et al.*, 2010; Stein *et al.*, 2012; Zehnder *et al.*, 2007). Although Stein et al. (2012) examined practice effects in cognitively normal adults, only the selected subtests of verbal fluency, word list recall, word list memory, and word list recognition were studied. They found no practice effects for verbal fluency and “small but significant gains” on word list recall, memory, and recognition. Zehnder et al. (2007) also examined practice effects after one year on the

individual components of the German version of the CERAD battery in 374 normal controls and 95 patients with mild Alzheimer's disease (AD). Similar to Stein's (2012) results, they found small increases over time on word list learning (0.41 ± 0.99), word list memory/ delayed recall (0.33 ± 0.94), and word list recognition (0.33 ± 1.14). Figure copy also showed small but significant decreases over one year (-0.18 ± 1.24). No improvement was evident in those patients with mild AD.

Of the previously listed studies examining practice effects on the CERAD, only two investigated the total score (Burkhart *et al.*, 2011; Rosetti *et al.*, 2010) developed by Chandler et al (Chandler *et al.*, 2005) as an indicator of global cognitive status. Rosetti et al. (2010) examined 383 normal controls and 655 participants with AD to learn more about progression in AD over time. Participants contributed a baseline assessment and up to four annual follow-up visits. They found that, over a period of four years post-baseline, normal control participants gained an average of 2.8 points on the total score whereas participants with AD showed a score decline of 22.2 points on average. Although the effect of attrition on degree of change over time was evaluated, this study only investigated attrition in the dementia group and did not assess drop-out in normal controls. Additionally, Burkhart et al. (2011) examined 57 healthy older adults who were at least 65 years of age using the German version of the CERAD on study entry and then again at seven and 90 day follow-up. Short intervals were used to approximate the typical time intervals utilized in postoperative cognitive dysfunction (POCD) research. On reassessment at day seven, significant practice effects on word list learning (4 points) and word list recall (1 point) were observed. Mean scores for each word list subtest (i.e., recall, memory, recognition) were stable at the 90 day follow-up and did not change from the scores obtained on the seven day follow-up. Other CERAD subtests did not show a practice effect likely due to ceiling effects as large proportions of their sample performed within one point of the maximum score on the Boston Naming Test (97%) and constructional praxis (89%). CERAD total scores increased significantly and were five points higher than baseline on seven day follow-up and seven points higher than baseline on 90 day follow up. Although Rosetti et al (2010) and Burkhart et al (2012) examined practice effects on the CERAD total score, those studies did not follow participants beyond three months (Burkhart *et al.*, 2011), nor did they examine the relative contribution of the individual subtest components (Rosetti *et al.*, 2010), or the effect of participant dropout in normally aging individuals (Burkhart *et al.*, 2011; Rosetti *et al.*, 2010). People experiencing cognitive difficulties may withdraw from study participation which may also serve to skew group performance.

In an effort to replicate and synthesize previous research findings on the CERAD test battery (Morris *et al.*, 1989), this study investigates the effect of practice on both the CERAD total score and on the individual CERAD subtests administered yearly to a group of cognitively normal older men enrolled in a clinical primary Alzheimer's prevention trial (Kryscio *et al.*, 2004). Similar to previous studies, we expected to find practice effects on the total score and on the word list subtests. To address the potential effect of attrition on increasing scores in longitudinally collected data, we examined the performance trajectory for those individuals who participated in all five assessment points (n = 182) and for those persons who completed at least two assessments (n = 308).

Methods

PREADViSE Participants

The Prevention of Alzheimer's Disease by Vitamin E and Selenium (PREADViSE) trial is a companion study to the Selenium and Vitamin E Cancer Prevention Trial (SELECT) (Klein *et al.*, 2003) funded by the National Cancer Institute examining the prevention of prostate cancer in men over the age of 50. The methods used in the PREADViSE trial have been published previously (Caban-Holt *et al.*, 2006 ; Kryscio *et al.*, 2004) and are described briefly here.

Recruitment for the PREADViSE trial took place from 2002-2008 at 130 clinical research sites also participating in the SELECT trial in the United States, Canada, and Puerto Rico. Men over the age of 60 were invited to enroll in PREADViSE. Baseline exclusion criteria included neurological, psychological, and medical conditions or medications affecting cognition or physical ability to participate in evaluation. A subset of cognitively normal participants from eight SELECT and PREADViSE sites was recruited to a validation sub-study in order to assess the usefulness of a brief memory screening tool (Memory Impairment Screen) that was used in the overall study. Participants in the validation sub-study are collectively referred to as the Normal Aging Group (n = 563), and they comprise the sample from which the current analyses were drawn. PREADVISE participants were continuously enrolled in the Normal Aging Group throughout this period. Exams were administered annually for a total of five years. Alternating versions of the word list (Ferris *et al.*, 1997; Rosen *et al.*, 1984) were used each year in an attempt to mitigate practice effects.

Participants in the current analysis

Of the 563 participants in the Normal Aging Group, those who completed only one assessment (n = 249) were excluded from the current study as practice effects could not be determined. Participants with more than one assessment who were diagnosed with Mild Cognitive Impairment (MCI) or dementia as a result of their baseline Normal Aging Group evaluation (n = 6) were also excluded. Reasons participants contributed a single data point include a planned cross-sectional validation (n=172). Specifically, some sites chose to participate only cross-sectionally and thus contributed only one assessment. Other reasons a single data point may have been contributed include withdrawal from the study after first assessment (n = 21) and clinical site closure (n = 56). Information was not collected regarding reasons for study withdrawal among those 21 participants who withdrew following their first assessment. Information regarding reason for withdrawal was also not collected throughout the study. Thus, the current sample consists of 308 participants who met criteria for inclusion in the Normal Aging Group sample, had at least two assessments, and were not clinically impaired at baseline. To investigate potential changes in the magnitude of practice effects related to study completion, a sensitivity analysis was conducted with the participants from this sample who provided data at baseline and all four follow-up visits (n = 182; “completers”) as well as those participants who did not complete all assessments (n = 126; “non-completers”). Among those classified as non-completers, participants contributed two (n = 27), three (n = 39), or four (n = 60) assessments.

Most non-completers provided annual consecutive assessments until withdrawing from the study; however, 10 individuals labeled non-completers did not have consecutive visits. Two individuals missed follow-up one but reappeared at the second follow-up, three participants missed the second follow-up two but completed the third follow-up, and five individuals missed the third follow-up but participated in follow-up four. Visits were numbered consecutively to indicate number of visits completed (and thus exposures to the tests), not as they occurred. All of the missed consecutive follow-up visits were due to clinical site scheduling conflicts.

Standard protocol approvals, registrations, and patient consents

Written informed consent was obtained from all study participants. All procedures were conducted with the approval of the University of Kentucky Institutional Review Board (IRB) as well as the IRB at each study site.

Analyses

A total score summing the raw subtest scores with the exception of the MMSE was developed by Chandler et al.(Chandler *et al.*, 2005) and is used in the current analysis. The maximum possible total score is 100 and corrections for age, education, and gender are made to the total raw score. The raw corrected score is transformed to a T-score (i.e., a standardized score with a mean of 50 and standard deviation of 10) according to the protocol provided in Chandler et al.(Chandler *et al.*, 2005)

To examine change over time, mean CERAD T-scores were analyzed using linear mixed models with visit as the only covariate; a first-order autoregressive covariance structure and a random effect for subjects nested within study site account for the repeated measures. Differences in baseline characteristics between the completers and non-completers were examined with chi-square tests and two sample *t*-tests. All analyses were conducted using PC-SAS 9.3®.

Results

Participants

Participant characteristics are summarized by group in Table 1. There were no significant differences at baseline in age, education, CERAD T-Score, or estimated full-scale IQ between the completers and non-completers. The 21 men who withdrew following their first assessment differed significantly from men who completed two or more assessments in terms of education ($p = 0.0426$) and estimated full-scale IQ ($p = 0.009$). Those who withdrew from the study after the first assessment had approximately one year less education and their estimated full-scale IQ was approximately 6.5 points lower than those who did not withdraw after the first assessment. It is unclear why education level and estimated full scale IQ would be different between those who withdrew and those who did not withdraw. These differences are not explained by age or baseline global cognitive functioning given that the CERAD T-score did not differ significantly between the two groups. Men who were not included due to participation in the cross-sectional study or site

closure were not included in this analysis because they did not actually withdraw from the study.

Change in Practice Effects Related to Study Completion

Baseline T-Scores for all included participants (n = 308) were approximately normally distributed, with 3.9% of participants scoring ≥65 and 7.5% of participants scoring ≤35. At the fifth assessment (n = 182), 11.5% of participants scored ≥65 and 4.4% of participants scored ≤35, representing a clear shift in the distribution with a medium effect size. Mean T-Scores increase significantly between Visits 1 and 4 (Figure 1) with mean scores at Visit 4 0.53 standard deviations (4.8 points, Cohen's d = 0.44) above the baseline mean.

In order to examine whether the trajectory of performance differed for those who completed the study, the analysis was repeated on completers. The results obtained were quite similar to those of the full sample. Their baseline T-scores followed a similar distribution (n = 182) with 2.2% of participants scoring ≥65 and 7.7% of participants scoring ≤35 (Table 1). Mean T-Scores increased significantly between Visits 1 and 5 (Figure 1). Visit 4 showed the greatest difference from baseline with mean scores 0.56 standard deviations (4.9 points) above the baseline mean. Mean scores at Visits 1 through 4 were examined between completers and non-completers to determine whether group performance differed by visit. No significant differences were observed (see table S1 published as supplementary material online attached to the electronic version of this paper at http://www.journals.cambridge.org/ipg).

In both completers and non-completers, the performance trajectory changed slightly in our sample between the fourth and fifth assessments, with mean CERAD T-scores falling approximately one point (p = 0.056). This change was not statistically or clinically significant.

Contribution of Individual Subtest Components to Overall Practice Effects

Since study completion did not significantly impact performance trajectory, the sample including both completers and non-completers was used to examine the individual components of the CERAD T-score (Figure 2). Although statistically significant practice effects were observed on each subtest, increases were limited in terms of clinical significance and generally showed small effect sizes (i.e., Cohen's d ≤ 0.2,) (Cohen, 1992). Performance on Animal Naming increased significantly, though not monotonically, between Visits 1 and 4 (p = 0.0298); however, the observed difference was less than one point (0.66 ± 0.30 points, Cohen's d = 0.042). Performance on the 15-item Boston Naming Test also increased significantly (p = 0.0195) between visits 1 and 5. Again, however, the observed difference was a fraction of a point (0.15 ± 0.06 points, Cohen's d = 0.15) as the vast majority of participants performed at or near the ceiling at all visits. Scores also increased significantly on Praxis Drawing between Visits 1 and 2 (p = 0.0025), again by less than one point (0.21 ± 0.07 points, Cohen's d = 0.19). Praxis Drawing scores did not increase significantly at any other time point with all participants performing at or near the ceiling on this subtest as well.

Performance on Word List Learning was associated with a significant increase in scores ($p < 0.0001$) between baseline and the 4th assessment (1.15 ± 0.28 points, Cohen's d = 0.26). Delayed Recall scores changed by less than one point across all five assessments; however, a statistically significant increase was observed from visit 1 to 4 ($p < 0.0001$, 0.66 ± 0.16, Cohen's d = 0.26). Finally, the Word List Recognition discrimination score also changed less than one point from baseline across the subsequent four assessments, yet significant differences were also observed from baseline to visit 4 ($p = 0.0117$, 0.27 ± 0.11 points, Cohen's d = 0.16).

Between visits four and five, a slight decrease was observed on some subtests. Notably, three of the six subtests (i.e., Boston Naming Test, Praxis Drawing, and Word List Recognition) are subject to ceiling effects and performances on these tests were close to their respective ceilings throughout the entire assessment period. Of the remaining subtests, slight and non-significant decreases were observed in Animal Naming (-0.40 ± 0.26, $p = 0.120$) and Word List Learning (-0.40 ± 0.24, $p = 0.0993$). The only subtest showing a significant decline between visits four and five was Word List Delayed Recall (-0.52 ± 0.13, $p < 0.0001$, Cohen's d = 0.25).

To further assess the influence of Word List Delayed Recall on the overall CERAD-T score, we examined differences in performance based on how participants' baseline scores compared to previously reported normative data (Welsh *et al.*, 1994). Impairment was defined as 1.5 SDs below the mean. The analyses show that those persons who were impaired at baseline (n = 33) demonstrated a practice effect on the CERAD T-score that dropped significantly between visits four (51.01 ± 1.69) and five (43.38 ± 2.11, $p < 0.0001$, Cohen's d = 0.78). In contrast, participants who had normal scores at baseline (n = 275) showed a plateau at visit four with no significant changes between visits four (56.08 ± 0.61) and five (55.47 ± 0.68, $p = 0.30$) (See Figure 3).

Discussion

Multiple factors must be accounted for when interpreting the findings obtained from longitudinal research studies. For example, repeatedly administered cognitive assessments may be subject to practice effects and multi-year longitudinal studies naturally suffer from participant attrition. Practice effects may obscure subtle but important cognitive decline and negatively affect case ascertainment. Furthermore, attrition is often not random and may be related to a variety of participant-specific factors that also have some effect on cognitive performance such as health status or education (Chang *et al.*, 2009).

Given the prevalent use of the CERAD battery in longitudinal research, this study examined practice effects on the individual subtests and the T-score derived from the CERAD tests (Morris *et al.*, 1989). Additionally, analyses were repeated between completers and non-completers to determine whether attrition bias may erroneously contribute to longitudinally increasing group mean scores. We observed significant increases in CERAD T-score performance across five yearly testing sessions with scores increasing by half of a standard deviation by the fourth assessment. It does not appear to be the case that attrition bias resulted in rising scores over time due to drop-out of poor performers. Age, estimated full-

scale IQ, and baseline CERAD T-score were not significantly different between completers and non-completers.

Our findings based upon the sample including both completers and non-completers was remarkably similar to that found in the Rosetti et al (Rosetti *et al.*, 2010) study. In their sample of 383 normal control participants, they observed a gain of 2.8 points in the CERAD raw (unadjusted) total score over a period that also included baseline assessment through four years of follow-up. This is reflected in the current study which saw a raw score increase slightly less than three points (see Figure 4) over three years in this sample. Similar to Burkhart et al. (Burkhart *et al.*, 2011), we also found evidence for ceiling effects on many of the subtests such as Praxis Drawing and Boston Naming. Given that the follow-up period in their study did not extend beyond three months and that they used a German algorithm for providing age, gender, and education adjustments, further comparisons beyond raw subtest scores are not possible.

A non-significant decline in CERAD T-scores was observed between visits four and five. However, upon closer inspection, Word List Delayed Recall was the only subtest that significantly influenced the change in trajectory. When the trajectory change was examined in terms of whether participants were considered impaired or unimpaired on the Word List Delayed Recall subtest at baseline, a much more striking contrast is evident. Participants who were not impaired showed a practice effect that plateaued at visit four whereas participants who were impaired improved dramatically by visit four only to have a large drop in their scores at visit five. The question of why those who performed poorly on the Word List Delayed Recall subtest at baseline showed such a different pattern of performance on the CERAD T-score cannot be answered with the current data. One may speculate that a cognitive profile including memory impairments in the context of otherwise normal cognition functioning may be indicative of the early stages of a neurodegenerative disease process. Should that be the case, it is possible that the large drop in CERAD T-scores is due to disease pathology reaching a critical threshold that overwhelms the practice effect. Further research with clinically defined or autopsy confirmed cases of neurodegenerative disease will be needed to shed more light on this possibility.

The current findings have several implications for clinical prevention trials. Longitudinal studies using the CERAD battery should be aware that moderate practice effects were observed in cognitively normal individuals who were screened into the PREADViSE dementia prevention trial. Thus, all studies using the CERAD battery longitudinally may wish to consider whether practice effects were influential in their own findings. Given that cut-scores are frequently the standard for determining whether an individual meets criteria as cognitively impaired, it is important to consider the effect of even small increases in performance over time. Although not evident in our cognitively normal sample, it is easy to envision a scenario in which practice-driven increases lead to maintenance of scores just above conventional cut points even though the participant may be experiencing early stage cognitive decline that is below the level of clinical detection. In clinical prevention trials that depend heavily upon these screening measures for case ascertainment, this could have deleterious consequences in terms of obtaining a suitable number of participants who meet criteria for impairment.

A separate but related issue concerns the consequences of relying on cut scores derived from a summary score of performance in multiple cognitive domains. When dividing our sample into two groups on the basis of whether they were impaired or unimpaired on the Word List Delayed Recall subtest relative to a normative sample, some differences in performance trajectory were observed. Both groups continued to show practice effects; however, the impaired group showed a steep drop in CERAD T-score and loss of the practice effect after five years. Given the small number of participants in the impaired group, it is difficult to say whether this finding is truly related to differences in memory performance or whether it is driven primarily by the variability of a small sample. It may be that poor performance on a measure of delayed recall relative to one's normative group is indicative of future global cognitive decline and, potentially, transition to a diagnosis of neurodegenerative disease. Previous work has shown that verbal delayed recall performance is a significant predictor of transition to dementia over five and 10 years of follow-up in a non-demented sample (Tierney *et al.*, 2005) and in a large sample (N = 3,055) of non-demented primary care patients over 75 years of age followed over three assessment periods occurring at 18-month intervals (Jessen *et al.*, 2011). Thus, it is certainly possible that the memory-impaired group in our sample is actually manifesting early symptoms of Alzheimer's disease.

The magnitude of the practice effect was most apparent in the CERAD demographically adjusted T-score, with lesser increases observed in the subtest raw scores and the unadjusted raw total score. CERAD T-scores increased approximately one-half of a standard deviation across four years of assessment whereas CERAD raw total scores increased 0.40 standard deviations over a four-year period. Mean CERAD total scores, both raw and adjusted, increase significantly between visits one and five, which suggests that participants are benefitting from repeated exposure to the testing material.

Strengths of this study include a large, geographically diverse sample, including a number of individuals (n = 182) who participated in the study for five consecutive years. This allowed for a careful analysis of whether attrition bias played a role in the findings of a practice effect in our study. Attrition bias did not account for these findings since there were no significant differences between completers and non-completers. Additionally, following individuals for five years provided adequate data to fully discern a moderate practice effect on the CERAD T-score. Even small practice effects may have large consequences for case ascertainment in longitudinal clinical trials. The finding of a moderate practice effect on CERAD T-scores with increases of, on average, five points (i.e., half of a standard deviation) over a follow-up period of four years suggests that corrections may be needed each year, especially when relying on cut-scores to screen for impairment. With improvements of such magnitude, participants may have to experience substantial decline before it becomes evident on longitudinal screening, and this could have serious consequences for intervention studies targeting mild cognitive impairment as an entity for intervention.

This study also has some weaknesses to be addressed in future research. First, our findings are limited to men given that the study sample is drawn from participants in a prostate cancer prevention trial. Additionally, the screening process is by necessity rather minimal in clinical trials; thus, it cannot be concluded with certainty that some participants recruited as

normal were actually experiencing mild impairments and were perhaps in the preclinical stage of dementia.

In summary, the current study demonstrated moderate practice effects on the CERAD T-score in a cohort of older men followed longitudinally for four years. Additionally, the practice effect was not due to attrition resulting from drop-out of poor performers given that no differences were observed in those variables typically affecting performance over time (i.e., age, estimated full-scale IQ, and baseline CERAD T-score performance). One significant result that should be expanded upon in future research is the finding that participants who are unimpaired on a global measure of cognition may actually be experiencing clinically significant decrements in specific domains of cognitive functioning. For example, a subset of our cognitively “normal” participants performed in the impaired range on the CERAD Word List Delay subtest at baseline. Those participants also demonstrated a different performance trajectory than those who were not impaired at baseline on the Word List Delay subtest. Specifically, impaired participants showed significant practice effects on the CERAD T-score across three follow-up visits followed by a steep decline between visits four and five. Although we are unable as yet to explore this finding in terms of whether these participants eventually transitioned to a diagnosis of dementia, it is possible that the CERAD Word List Delay subtest is a stronger marker of future cognitive decline than the global CERAD T-score. Future longitudinal research following participants until the point of clinical diagnosis may be useful in determining the prospective utility of the individual CERAD subtests.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study is not industry sponsored.

NIA Grant Support: R01AG019241 and R01AG038651

References

Boekamp JR, Strauss ME, Adams N. Estimating Premorbid Intelligence in African-American and White Elderly Veterans Using the American Version of the National Adult Reading Test. Journal of Clinical and Experimental Neuropsychology. 1995; 17:645–653. [PubMed: 8557806]

Burkhart CS, et al. Evaluation of a Summary Score of Cognitive Performance for Use in Trials in Perioperative and Critical Care. Dementia and Geriatric Cognitive Disorders. 2011; 31:451–459. [PubMed: 21778726]

Caban-Holt A, Bottiggi K, Schmitt FA. Measuring treatment response in Alzheimer's disease clinical trials. Geriatrics. 2005:3–8. [PubMed: 16025769]

Caban-Holt, A., et al. Studying the effects of vitamin E and selenium for Alzheimer's disease prevention: the PREADVISE model. In: Vellas, B.; Fitten, L.; Winblad, B.; Feldman, H.; Grundman, M.; Giacobini, E.; Kurz, A., editors. Research and Practice in Alzheimer's Disese. Paris: Serdi Publisher; 2006. p. 124-130.

Calamia M, Markon K, Tranel D. Scoring Higher the Second Time Around: Meta-Analyses of Practice Effects in Neuropsychological Assessment. Clinical Neuropsychologist. 2012; 26:543–570. [PubMed: 22540222]

Chandler MJ, et al. A total score for the CERAD neuropsychological battery. Neurology. 2005; 65:102–106. [PubMed: 16009893]

Chang CCH, Yang HC, Tang G, Ganguli M. Minimizing attrition bias: a longitudinal study of depressive symptoms in an elderly cohort. International Psychogeriatrics. 2009; 21:869–878. [PubMed: 19288971]

Cohen J. A Power Primer. Psychological Bulletin. 1992; 112:155–159. [PubMed: 19565683]

Cooper DB, et al. Effects of practice on category fluency in Alzheimer's disease. Clinical Neuropsychologist. 2001; 15:125–128. [PubMed: 11778573]

Cooper DB, Lacritz LH, Weiner MF, Rosenberg RN, Cullum CM. Category fluency in mild cognitive impairment - Reduced effect of practice in test-retest conditions. Alzheimer Disease & Associated Disorders. 2004; 18:120–122. [PubMed: 15494616]

Ferris SH, et al. A multicenter evaluation of new treatment efficacy instruments for Alzheimer's disease clinical trials: Overview and general results. Alzheimer Disease & Associated Disorders. 1997; 11:S1–S12. [PubMed: 9236947]

Fillenbaum GG, et al. Consortium to establish a registry for Alzheimer's disease (CERAD): The first twenty years. Alzheimer's and Dementia. 2008; 4:96–109.

Galvin JE, et al. Predictors of-preclinical Alzheimer disease and dementia - A clinicopathologic study. Archives of Neurology. 2005; 62:758–765. [PubMed: 15883263]

Hickman SE, Howieson DB, Dame A, Sexton G, Kaye J. Longitudinal analysis of the effects of the aging process on neuropsychological test performance in the healthy young-old and oldest-old. Developmental Neuropsychology. 2000; 17:323–337. [PubMed: 11056847]

Jessen F, et al. Prediction of Dementia in Primary Care Patients. PLoS One. 2011; 6

Klein EA, et al. The selenium and vitamin E cancer prevention trial. World Journal of Urology. 2003; 21:21–27. [PubMed: 12756490]

Kryscio RJ, Mendiondo MS, Schmitt FA, Markesbery WR. Designing a large prevention trial: statistical issues. Statistics in Medicine. 2004; 23:285–296. [PubMed: 14716729]

Morris JC, 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]

Rosen WG, Mohs RC, Davis KL. A New Rating-Scale for Alzheimers-Disease. American Journal of Psychiatry. 1984; 141:1356–1364. [PubMed: 6496779]

Rosetti H, Cullum C, Hynan L, Lacritz L. The CERAD neuropsychological battery total score and the progression of Alzheimer disease. Alzheimer Disease & Associated Disorders. 2010; 24:138–142. [PubMed: 20505431]

Stein J, et al. The assessment of changes in cognitive functioning: Age, education, and gender-specific reliable change indices for older adults tested on the CERAD-NP battery: Results of the German study on ageing, cognition, and dementia in primary care patients (AgeCoDe). American Journal of Geriatric Psychiatry. 2012; 20:84–97. [PubMed: 22183013]

Tierney MC, Yao C, Kiss A, McDowell I. Neuropsychological tests accurately predict incident Alzheimer disease after 5 and 10 years. Neurology. 2005; 64:1853–1859. [PubMed: 15955933]

Welsh KA, et al. The Consortium-to-Establish-a-Registry-for-Alzheimers-Disease (Cerad) .5. A Normative Study of the Neuropsychological Battery. Neurology. 1994; 44:609–614. [PubMed: 8164812]

Zehnder A, Blasi S, Berres M, Spiegel R, Monsch A. Lack of practice eEffects on neuropsychological tests as early cognitive markers of Alzheimer disease? American Journal of Alzheimers Disease and Other Dementias. 2007; 22:416–426.

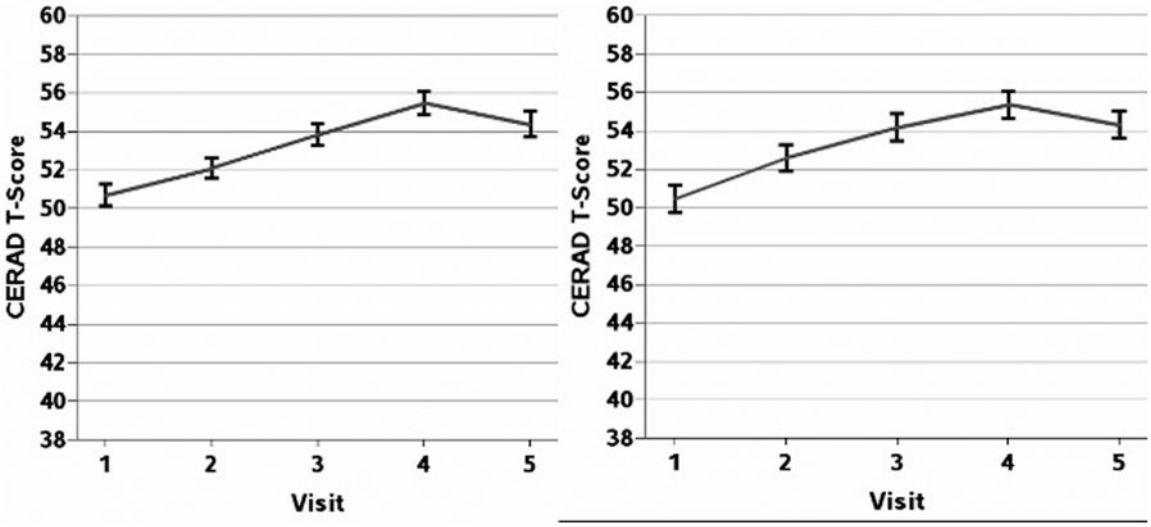

Figure 1.

Mean CERAD T-scores by visit.

At least two visits, N = 308 at V1 and V2

All five visits, N = 182

Note: All pairwise differences are significant ($\alpha = 0.05$) except 3 vs. 5 and 4 vs. 5.

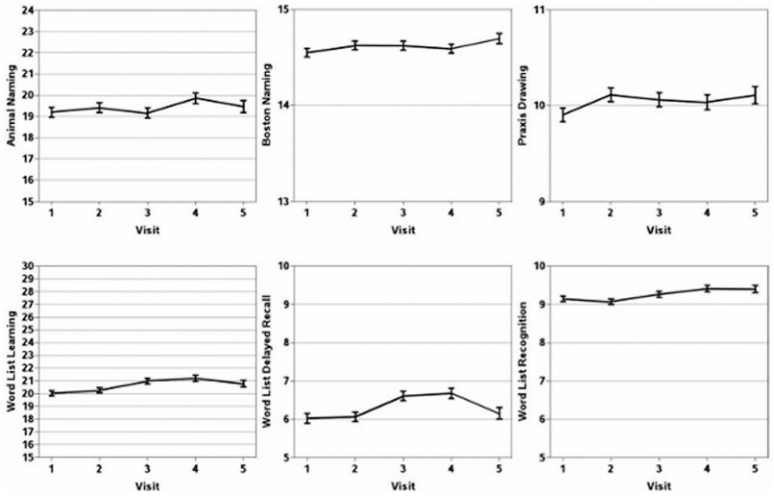

Figure 2.

Mean raw scores from individual CERAD components (at least two visits, n = 308 at V1 and V2).

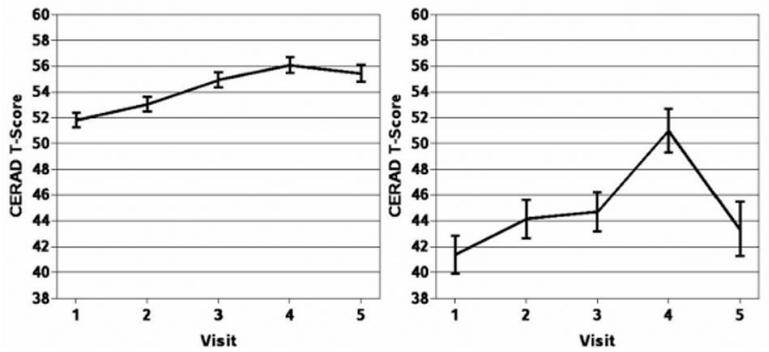

Figure 3.

Mean CERAD T-scores by CERAD Word List Delay Performance.

Unimpaired (n = 276)

Impaired (n = 33)

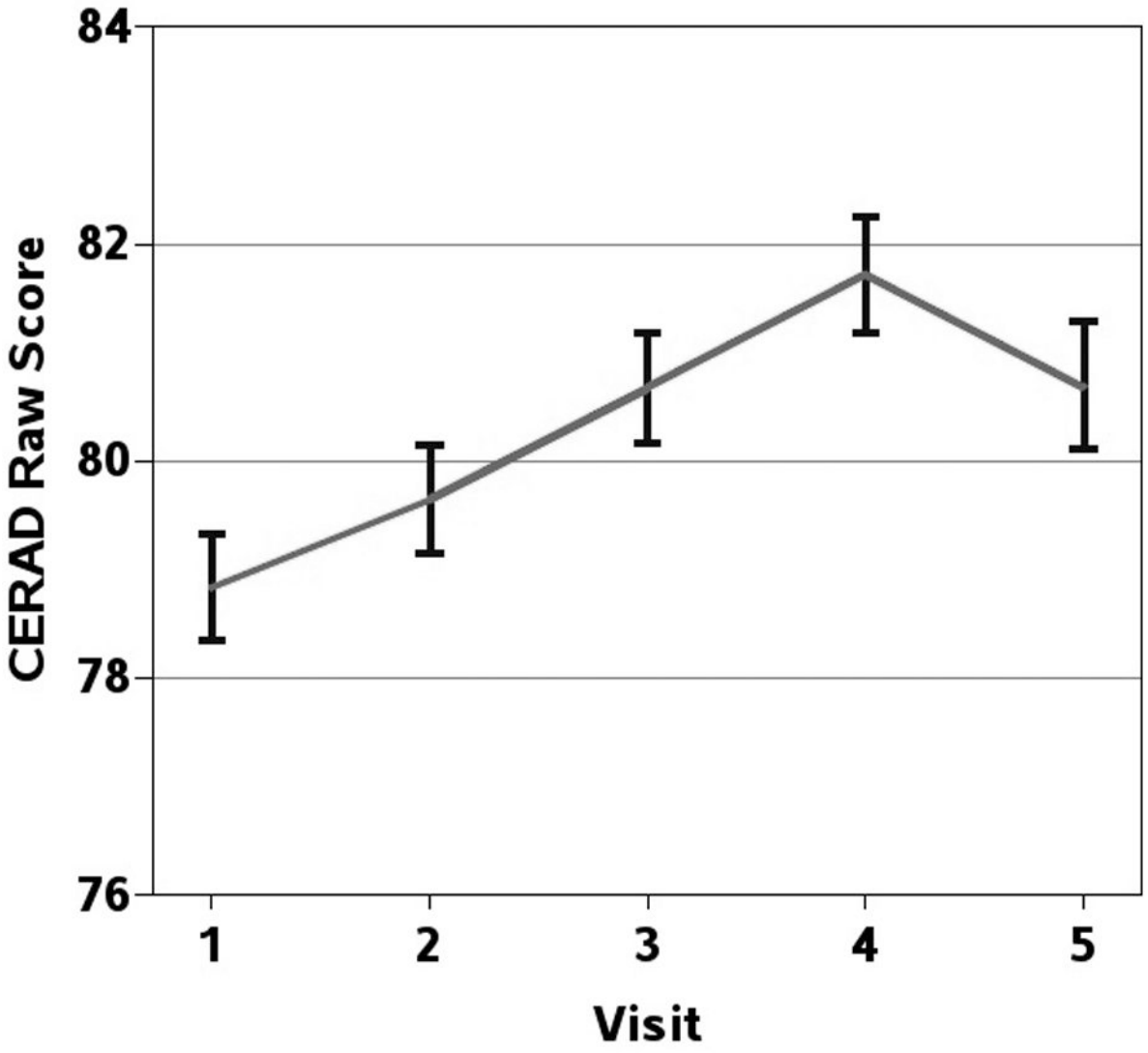

Figure 4. Mean CERAD raw scores for completers (n = 182)

Table 1

Participant characteristics.

Characteristic	2-4 visits (n = 126)	5 visits (n = 182)	All (n = 308)
Age at baseline, y (mean ± SD)	70.4 ± 6.4	70.3 ± 4.8	70.4 ± 5.4
CERAD T-Score at baseline (mean ± SD)	51.0 ± 9.5	50.4 ± 8.7	50.7 ± 9.0
range		27-68	27-70
CERAD T-Score at fourth assessment(mean ± SD)		54.3 ±10.8	
Range		19-77	
Estimated Full-Scale IQ[a]	110.8 ± 8.5	110.2 ± 9.3	110.4 ± 9.0
Education (%)			
More than college	41.3	50.0	46.4
College or less	58.7	50.0	53.7
Race/Ethnicity (%)			
White	89.7	86.3	88.0
African-American	7.1	9.3	8.5
Other	3.2	4.4	3.6

[a] Full Scale IQ estimated with the American Version of the Nelson Adult Reading Test (AmNART)(Boekamp *et al.*, 1995)