UTILITY OF THE CONSORTIUM TO ESTABLISH A REGISTRY FOR ALZHEIMER'S DISEASE (CERAD) NEUROPSYCHOLOGICAL BATTERY TOTAL SCORE IN THE PROGRESSION OF ALZHEIMER'S DISEASE

APPROVED BY SUPERVISORY COMMITTEE

Laura Lacritz

Linda Hynan

C. Munro Cullum

Dedicated in loving memory to

Reverend Robert Swanson

and

Arnold & Irene Richardson

UTILITY OF THE CONSORTIUM TO ESTABLISH A REGISTRY FOR ALZHEIMER'S DISEASE (CERAD) NEUROPSYCHOLOGICAL BATTERY TOTAL SCORE IN THE PROGRESSION OF ALZHEIMER'S DISEASE

by

HEIDI CHRISTINE ROSSETTI

THESIS

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August, 2007

ACKNOWLEDGEMENTS

There are a number of individuals I would like to acknowledge, for the successful completion of this thesis would not have been possible without their support, encouragement, and assistance. First and foremost, I extend my sincerest appreciation to the members of my committee. Many thanks are owed to my thesis chair, Laura Lacritz, who provided a limitless store of guidance and expertise. She challenged me to think as both a researcher and a clinician, and her professionalism, dedication, and enthusiasm have made her an amazing role model and mentor. I owe special thanks to Linda Hynan, who generously contributed her time and effort to this project. She graciously offered her statistical acumen to a student who was very much a statistical neophyte. I would also like to express my gratitude to Munro Cullum, whose extensive knowledge of neuropsychology in general and the CERAD battery in particular has proven invaluable. His ready supply of encouragement and humor are much appreciated. I would also like to thank my program chair, Cheryl Silver, for her excellent guidance and support throughout this pursuit of my Master's degree.

I would like to thank my fellow students, who have also become treasured friends: Sarah Spector, Alyssa Parker, Jenny Hughes, Anu Dhingra, and Keith Bernardo. They have provided warm support, sharp intellect, gentle humor, and endless encouragement. Finally, I would like to express my love and gratitude to my friends and family. My parents, John and Judy Richardson, have always supported my goals and dreams, and inspired me through struggles and stressors along the way. I am

blessed to have their example to follow. My best friend and sister-in-law, Ashley Morrow, helped me keep things in perspective, retain some semblance of a social life, and was unfailingly caring throughout this experience. Last, but most definitely not least, I owe my deepest thanks to my incredible husband, Casey. He is the strongest supporter and champion of my every endeavor, and I am continually impressed with his unwavering love and patience. I count myself lucky to have found such a partner – my dearest friend, my confidante, and my ally.

UTILITY OF THE CONSORTIUM TO ESTABLISH A REGISTRY FOR ALZHEIMER'S DISEASE (CERAD) NEUROPSYCHOLOGICAL BATTERY TOTAL SCORE IN THE PROGRESSION OF ALZHEIMER'S DISEASE

Publication No. _____

Heidi Christine Rossetti, M.S.

The University of Texas Southwestern Medical Center at Dallas, 2007

Supervising Professor: Laura H. Lacritz, Ph.D., ABPP

ABSTRACT

The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) created a neuropsychological battery that is both brief and sensitive to dementia (Morris et al., 1989). Chandler et al. (2005) put forth a method of calculating a Total Score for the CERAD along with normative data. The objective of this study was to determine the utility of the Total Score as a measure of progression of Alzheimer's disease (AD).

Subjects included CERAD registry normal controls (NC; *N* = 383) and AD subjects (*N* = 655) with a baseline assessment and at least one follow-up assessment. Change Scores were calculated along with Reliable Change Indexes (RCI). The AD sample declined an average of -7.2 points per year, compared to a 1.0 point annual increase obtained by the NC sample. By the third annual assessment, the majority of AD subjects (65.2%) exceeded the confidence interval established by the RCI. Annualized CERAD Change Scores significantly correlated with change scores on the MMSE (*r* = .66), CDR Sum of Boxes (*r* = -.42), and BDRS (*r* = -.38). The impact of race, gender, education, and age-at-baseline on AD progression was examined with analysis of covariance and multiple regression. Demographic variables accounted for only 4% of the variance in annualized change in CERAD performance, with greater annualized decline in Total Score observed in Caucasians (*M* = -7.64, *SD* = 6.82) versus African-Americans (*M* = -4.60, *SD* = 7.03); males (M = -8.22, SD = 6.70) versus females (M = 6-.44, SD = 7.04); and younger age-at-baseline (M = -8.72, SD = 6.44) versus older age-at-baseline (M = -6.85, SD = 7.01). Neither education nor dementia severity significantly impacted annualized Change Scores. The current study provides support for the validity of the CERAD Total Score as a measure of progression in AD.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AD	Alzheimer's Disease
ADRC	Alzheimer's Disease Research Center
ADL	Activities of Daily Living
BDRS	Blessed Dementia Rating Scale
BNT	Boston Naming Test
CI	Confidence Interval
CDR	Clinical Dementia Rating
CERAD	Consortium to Establish a Registry for Alzheimer's Disease
DRS	Dementia Rating Scale
FSIQ	Full Scale Intelligence Quotient
IADL	Instrumental Activities of Daily Living
M	Mean
MCI	Mild Cognitive Impairment
MMSE	Mini Mental State Examination
NC	Normal Control
NINCDS-ADRDA	National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association
RCI	Reliable Change Index
SBT	Short Blessed Test

SE_{diff}	Standard Error of the Difference
SE_m	Standard Error of Measurement
SD	Standard Deviation
SPSS	Statistical Package for the Social Sciences

INTRODUCTION

Alzheimer's disease (AD) is a chronic, debilitating illness with a hallmark of progressive cognitive decline. It is the most common cause of dementia, comprising 50% to 80% of cases (Corey-Bloom, 2004). It has been estimated that 5% of individuals aged 65 to 74 years and nearly 50% of people over age 85 have AD (Desai & Grossberg, 2005). The fastest growing sector of the United States population is over 65 years of age, and this segment is expected to number approximately 58.9 million by the year 2025 (Fhsani, Bedlich, McNicoll, & Fly, 2003). The prevalence and incidence of dementia will undoubtedly rise considerably as the United States population continues this aging trend.

AD is defined by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) as a syndrome that is insidious in onset and characterized by a progressive deterioration of memory and deficits in two or more areas of cognition [e.g., language, motor, and/or visuospatial skills] (McKhann et al., 1984). Although the presentation and course of AD is often variable, cognitive changes tend to accumulate before a notable decline in functioning results in evaluation and diagnosis. As a result, individuals with AD typically seek medical attention several years after disease onset (Cullum & Rosenberg, 1998).

Corey-Bloom (2004) provided a summary of the cognitive impairments observed during the course of AD. The earliest symptom of AD is typically memory loss. This chief complaint initially involves difficulty learning and recalling new information such as names or conversation details. Eventually, the memory deterioration affects remote memory. Language deficits primarily involve naming difficulty and decreased fluency, progressing to aphasia and near-mutism. The effects of AD on visuospatial ability are evidenced by misplacing objects or getting lost, difficulty recognizing and drawing complex figures, and impaired driving. Tasks involving problem solving, abstraction, reasoning, decision making, and judgment become increasingly difficult. Individuals may have trouble with calculation and organizing complex tasks. Difficulty with activities of daily living such as handling money, operating appliances, or dressing may become apparent.

With the continuing advances in the treatment of AD, routine screening and early diagnosis are of crucial importance. Neuropsychological screening instruments serve several valuable functions for both the clinical care of patients and for enhancing our understanding of this disease. The information obtained from screening measures can inform caretakers about the need for intervention in areas such as driving or finances, and prompt families to plan long-term illness management (Kilada et al., 2005). A screening measure of cognitive function can identify patients with dementia and/or indicate the need for further examination

(Jacobs et al., 1995). Screening measures are used to track progression of the disease, determine the degree of cognitive and functional decline, and predict time to institutionalization or death (Lopez et al., 2000). Screening instruments can also be used to measure changes associated with various interventions (e.g. cholinesterase inhibitors).

Despite the usefulness of cognitive screening measures, only a minority of general practitioners routinely employ them, most citing lack of time as the major obstacle (Mendiodo et al., 2003). In order to be both valuable and utilized, a screening instrument must be brief, standardized, and reliable (Relkin, 2000). It is equally important that assessment tools adapt as the dementia progresses and the patient's condition deteriorates. AD progression varies per individual, with some worsening quickly while others experience a more protracted decline (Adak et al., 2004). The typical duration of illness following diagnosis is variable, with studies reporting median survival rates ranging from 3.4 to 8.3 years (Brookmeyer, Currada, Curriero, & Kawas, 2002; Larson et al., 2004; Neumann et al., 2001). The ability of a screening measure to detect the presence of impairment and measure cognitive change over time is integral to the prognosis and management of AD.

The following section reviews various cognitive screening measures that have been used in identification of dementia and assessing change over time, focusing on the Consortium to Establish a Registry for Alzheimer's Disease

(CERAD) neuropsychological test battery. A total score for the CERAD has recently been developed and has not yet been examined with respect to its association with other established measures of cognitive functioning and change, which was the goal of the current study.

Cognitive Screening Measures

Some of the most common screening tools and measures of progression include the Mini Mental State Examination (MMSE), Clinical Dementia Rating (CDR), Blessed Dementia Rating Scale (BDRS), and the Short Blessed Test (SBT).

MMSE

The MMSE is a brief, well-accepted formal measure of cognitive functioning that was originally created to identify the presence of cognitive impairment and document cognitive changes over time (Folstein, Folstein & McHugh, 1975). It consists of 22 items which assess orientation to time and place, attention/concentration, immediate and delayed recall, naming ability, ability to follow simple commands, and constructional skills. The MMSE can be administered in approximately 5-10 minutes, with a maximum score of 30 points. A score of 23 or less is generally used as a traditional cut-off for impairment (Tombaugh & McIntyre, 1992). The MMSE is reliable, with inter-rater reliability demonstrated at $r > .65$ and two-month test-retest reliability at $r > .80$ (Folstein et al., 1975).

Although the MMSE is the most widely used dementia screening tool in research and clinical settings (Foldi, Majerovitz, Sheikh, Rodriguez, 1999), it has limitations. Due to a lack of sensitivity, the MMSE has difficulty differentiating between very mild AD patients and normal subjects (Tombaugh & McIntyre, 1992). The measure consists of primarily verbal items, which may lead to inaccurate estimation of cognitive abilities when language capability is affected (Foldi, Majerovitz, Sheikh, Rodriguez, 1999). Because of this dependence upon language, the MMSE fails to adequately assess nonverbal skills such as constructional praxis (Tombaugh & McIntyre, 1992). In addition, the measure does not assess executive functioning. The MMSE requires that cognitive deficits in highly educated people be relatively pronounced in order to meet the level of clinical significance, thereby increasing the likelihood of misclassifying some cognitively impaired patients as normal (Tombaugh & McIntyre, 1992). Woodbury and Fillenbaum (1996) found that the MMSE does not identify specific areas of cognitive loss but may be a good indicator of the severity of overall impairment.

Given the progressive decline of cognitive functioning in AD, performance on screening tests should also decline with repeated testing over time. A longitudinal study using 1 month to 3 year test-retest intervals found significant decline in AD over time, with MMSE scores falling on average between 2 to 5 points per year (Van Belle, Uhlmann, & Hughes, 1990). In similar

studies, Becker et al. (1988) reported a mean drop of 2 to 3 points per year, and Holmes and Lovestone (2003) found an average decline of 2.2 points per year. Morris et al. (1993) determined that the subjects with a baseline MMSE of 25 dropped 1 point annually, while subjects with a baseline of 15 dropped 4 points per year. Clark et al. (1999) examined the variability in annual MMSE scores in CERAD study patients reported that a change in MMSE score must exceed 3 points in order to be clinically meaningful. Larson et al. (2004) conducted an investigation of survival after the initial diagnosis of AD, and defined "greater cognitive decline" over 1 year as a decrease of 5 or more points in MMSE score.

However, in their comprehensive review of the MMSE, Tombaugh and McIntyre (1992) found that the MMSE was less sensitive to the progressive decline of functioning associated with severe AD. In their study of moderate to severe AD individuals, Feldman and Woodward (2005) also noted that the MMSE encounters floor effects which decreased its sensitivity and reliability. Clark et al. (1999) suggested that the MMSE's ability to document AD changes over time may be somewhat limited due to high variability in individual rates of change and high measurement error. Therefore, the use of the MMSE as a tool to assess progression in certain stages of AD may be problematic.

CDR

The CDR is a widely accepted rating scale of AD severity based on the results of clinician ratings of a patient's cognitive performance and information

gathered during separate semi-structured interviews with an informant and the patient (Berg, 1988; Feldman & Woodward, 2005). The CDR is composed of six categories or "boxes" that assess cognitive function in the domains of Memory, Orientation, Judgment and Problem Solving, Community Affairs, Home and Hobbies, and Personal Care (Berg, 1988; Hughes, Berg, Danziger, Coben, & Martin, 1982). Ratings are assigned in which 0 = no dementia, 0.5 = questionable, 1 = mild, 2 = moderate, 3 = severe, 4 = profound, and 5 = terminal (Fillenbaum, Peterson, & Morris, 1996). Scores for each of the domains can be totaled to derive a total score, the CDR Sum of Boxes score, ranging from 0 to 30 points (Morris, 1993). That data is entered into a scoring program to derive the overall CDR Stage (range 0-5).

By incorporating the assessment of behavior and functioning in daily activities, the CDR examines more than cognitive performance. It was found to be predictive of time of death in a sample of AD subjects, and to be significantly negatively correlated with the MMSE (Fillenbaum, Peterson, & Morris, 1996). The reliability and validity of the CDR has been established, and the Sum of Boxes score offers clinicians and researchers a single data point that summarizes dementia severity across multiple domains (Fillenbaum, Peterson, & Morris, 1996; Morris, 1993). Training in CDR assessment is not standardized, and focuses on instructing health care professionals to rate the level of impairment in each domain in agreement with the rating of a clinician highly experienced with

CDR assessment (Tractenberg, Schafer, & Morris, 2001). Inter-rater reliability of clinicians trained on the CDR has been reported at $r = .90$ using the Sum of Boxes score (Burke et al., 1988). However, the accuracy of the CDR is largely dependent upon the availability and reliability of a well-informed caregiver (Perneczky et al., 2006). Furthermore, its use in general practice is often limited because it requires sufficient time to conduct the caregiver and patient interviews (Feldman & Woodward, 2005).

BDRS and SBT

The BDRS evaluates memory and functional impairment in the performance of everyday activities (Blessed et al., 1968). The CERAD utilizes a variant of the BDRS known as the Short Blessed Test (SBT) that is a measure of orientation, memory, and attention (Katzman, Brown, & Fuld, 1983). Compared to the BDRS, the SBT consists of the first 11 of the original 22 items. The questions assess various activities, such as how well the person can perform household tasks, find their way around, understand what is going on around them, eat without assistance, dress themselves, and use the toilet by themselves (Pisani, Inouye, McNicoll, & Relich, 2003). The test is normally given to the patient; however, if the patient is too impaired the test is conducted with a primary caregiver. Therefore, the subject of the administration varies, which may negatively affect reliability. Points ranging from zero to one for the first eight items and zero to three for the last three items are given for each response. The

points are added to calculate a maximum total score of 17, with a higher score indicating greater impairment. The BDRS has been shown to effectively discriminate between demented and non-demented subjects (Juva, Makela, & Erkinjuntti, 1997), and it has shown good test-retest reliability (r = .88; Villardita & Lomeo, 1992). Holmes & Lovestone (2003) reported an average rate of progression of 1.3 points per year.

CERAD

In order to examine the progression of AD, clinicians and researchers require a standardized, brief, and reliable means of evaluating the presence and degree of cognitive impairment in AD cases, with an emphasis on longitudinal study. Repeated assessment of the same individuals over time is necessary to examine individual variation, risk factors, and disease trajectories (Fisher, Rourke, & Bieliauskas, 1999; Anstey & Hofer, 2004). Although the above measures are well-known and widely used, it has been argued that none are entirely adequate (Morris et al., 1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) neuropsychological battery was developed to meet these needs by providing uniform longitudinal data on a large sample of patients with AD (Welsh-Bohmer & Mohs, 1997; Fisher et al., 1999).

Funded by the National Institute on Aging (NIA) for a ten year period beginning in 1986, a CERAD registry database was compiled using longitudinal data from 24 participating Alzheimer's Disease Research Centers (ADRC) across

the United States (Edland, Beekly, Barnhart, & van Belle, 1997). The CERAD registry consisted of four main areas of AD assessment: clinical, neuropsychological, neuropathological, and neuroimaging (Morris et al., 1989). A key criterion for inclusion into the CERAD project was the likelihood of annual follow-up (Fillenbaum et al., 1998). As a result, a large database with rich clinical information was amassed.

Although grant funding for the CERAD project ended in 1995, the CERAD neuropsychological battery continues to be utilized by AD research centers, hospitals, private clinics, and major universities for both clinical and research purposes (Heyman & Fillenbaum, 1997). The CERAD battery has been translated into Bulgarian, Czech, French, Spanish, Italian, German, Japanese, Chinese, Hebrew, Portuguese, Korean, and Dutch (Heyman et al., 1997). As a result of the widespread use of the CERAD neuropsychological battery, normative data have been provided for populations outside of the United States. Additional populations include Australian (Collie, Shafiq-Antonacci, Maruff, Tyler, & Currie, 1999), African Caribbean (Stewart, Richards, Brayne, & Mann, 2001), Finnish (Karrasch & Laine, 2003), and German (Satzger et al., 2001), among others.

The battery is comprised of four subtests that are well-established [Animal Category Verbal Fluency (Isaacs & Kennie, 1973), Modified Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 1978), Word List Memory (Atkinson &

Shiffrin, 1971), and Constructional Praxis (Rosen, Mohs, & Davis, 1984)]. These tests were selected in order to assess the primary cognitive areas affected by AD, including memory, language, visuoconstructional ability, and general cognitive functioning (Morris et al., 1989; Morris et al., 1988). The chosen tasks were also intended to be useful in documenting progression of illness at annual intervals throughout the course of AD (Heyman et al., 1997). The battery can be administered within approximately 30 minutes (Morris et al., 1989).

Reliability

The reliability of the CERAD neuropsychological battery was initially investigated by Morris et al. (1989) using data from 16 CERAD centers. Reliability for each of the measures was statistically significant. Inter-rater reliability in a mixed sample of AD and control subjects was high [range r = .92 (Constructional Praxis) to r = 1.0 (Word List Recall)]. The lowest one-month test-retest correlations in AD subjects were Word List Recall (r = .43) and Recognition (r = .44), which was attributed to the restricted range of scores and floor effects. All other test-retest correlations were high [range r = .68 (Word List Memory) to r = .90 (BNT)]. Lower correlations for test-retest were found in control subjects, likely due to restricted range of scores and ceiling effects [range r = .36 (Word List Recognition) to r = .67 (Verbal Fluency)].

Validity

Content Validity

The CERAD neuropsychological battery consists of tests involving language, praxis, general cognitive function, and memory. These tasks were selected because the literature of the 1980's suggested they were the primary domains affected in AD (Morris et al., 1989). The prominence of such deficits in AD has been supported in more recent research, as well as deficits in executive functioning which often occur early in the course of the disease (Corey-Bloom, 2004). Although the CERAD battery does not include a subtest of executive functioning, multiple studies have supported the validity of the CERAD battery in AD assessment (Morris et al., 1989; Morris et al., 1993; Welsh, Butters, Hughes, Mohs, & Heyman, 1992: Welsh-Bohmer et al., 1997).

The subtests on the CERAD battery are significantly inter-correlated (Morris et al., 1989), with correlations ranging from $r = .78$ (Word List Learning and Verbal Fluency) to $r = .16$ (modified BNT with Constructional Praxis and Word List Recall). The relationships between the subtests and measured cognitive domains have been investigated by three factor analytic studies using subjects from the CERAD registry. Morris et al. (1989) studied test scores of subjects with mild to severe AD and determined that three factors (memory, language, and construction) accounted for 73% of the variance. Henderson and Buckwalter (1994) conducted separate factor analyses on AD and control subjects

and identified three key factors (memory and attention, visuospatial ability, and language). Principle component analysis of the subtests using control subjects identified five factors (memory and learning, verbal fluency/naming, praxis, orientation/registration, and executive function) that accounted for 64.5% of the total variance (Collie, Shafiq-Antonacci, Maruff, Tyler, & Currie, 1999). These three studies support the original aim of CERAD to compile a neuropsychological battery with tests of memory, language, praxis, and general cognitive function (Morris et al., 1989).

Discriminant Validity

Subtests of the CERAD battery have been shown to differentiate between controls and AD subjects and discriminate mild, moderate, and severe cases of AD. Morris et al. (1989) determined that control subjects performed significantly better than AD subjects. Their analysis also found that subjects with mild AD (CDR = 1) performed significantly better than subjects with moderate AD (CDR = 2). Welsh, Butters, Hughes, Mohs, & Heyman (1991) found that control subjects performed significantly better than AD subjects on all aspects of the CERAD Word List task (learning, recall, recognition, intrusion errors, and savings scores). Welsh et al. (1992) conducted a discriminant function analysis and determined that delayed recall best discriminated controls from mild AD subjects. Their analysis also revealed that the addition of Constructional Praxis improved discrimination of mild and moderate AD. Furthermore, the

combination of true positives, intrusion errors, Verbal Fluency, and Constructional Praxis best discriminated moderate and severe AD.

Concurrent Validity

Few comparison studies of CERAD subtests to other accepted measures have been conducted. The CERAD Word List Learning, Recall, and Recognition were compared to the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987), a well established 16-item word list learning test (Kaltreider et al., 2000). The number of words learned was modestly correlated between tests ($r = .65$, $p < .001$). The correlations for the initial learning trial ($r = .39$), last trial ($r = .48$), true positives ($r = .40$), recognition discriminability ($r = .35$), and intrusion errors ($r = .30$) were lower but significant ($p < .001$). The concurrent validity of many of the tests from which the CERAD subtests were derived have been previously examined (Spreen & Strauss, 1998).

Normative Data

Welsh et al. (1994) analyzed the performance of 413 normal elderly controls from 23 participating CERAD sites and provided normative data for the CERAD neuropsychological battery as well as mean test results for gender, age, and education. The sample was well educated ($M = 14.0$ years), with smaller numbers of subjects with less education (< 12 years) compared to those with more education (≥ 12 years). Since minority members were under-represented, the normative data was limited to Caucasians.

Demographic Variables and CERAD

Gender, Education, Age of Onset

In their establishment of normative data for the CERAD neuropsychological battery, Welsh et al. (1994) compared a higher education group (≥ 12 years) with a lower education group (< 12 years) and reported significant effects for gender, education, and age of onset across subtests. Women performed better than men on the MMSE and Word List Learning, Recall, and Recognition in the higher education group. Also in the higher education group, younger individuals (50-69 years) outperformed those in the older age group (70-89) on the MMSE, Modified BNT, Word List Learning, Recall, Recognition, and Savings Score. In the lower education group, age differences were found only on Constructional Praxis. The higher education group performed better than the lower education group on the MMSE, Modified BNT, Constructional Praxis, and Word List Recognition True Positives. No education effects were found for Verbal Fluency, Word List Learning, Recall, Savings Score, or recognition false positive errors. This study of demographic variables was limited by the small sample size within each age groups (n = 23) and a disproportionate number of women (n = 30) to men (n = 16) in the education samples. Additionally, the authors did not report interaction effects among the demographic factors, which makes it unclear how much of the variance was due to each demographic variable.

The impact of demographic variables upon dementia and its progression is a widely researched topic, and much of the Welsh et al. (1994) findings have been supported by other studies using either the CERAD battery or other measures. Prior literature suggests that there may be gender differences in cognitive function and risk of AD (Barnes et al., 2003). Lapane et al. (2001) studied gender differences and mortality in nursing home residents with AD. Their results showed that women experienced more severe impairment and faster cognitive decline than men. However, they also found that men had an increased risk of mortality. This is supported by Heyman, Peterson, Fillenbaum, and Pieper (1996), who found that men with AD had a median survival time of 5.7 years, compared with 7.2 years for women. These findings raise the possibility that AD has underlying mechanisms that vary according to gender. While women generally outperform men on cognitive testing in normal control samples, males with AD have shown better cognitive performance than AD females on neuropsychological tests (Doraiswamy et al., 1997), and women with AD performed worse than men with AD on certain CERAD subtests (Henderson & Buckwalter, 1994). Although women hold a premorbid advantage over men in verbal skills, this study found gender-associated differences in naming, verbal fluency, and language factors. Women were more impaired in subtests involving verbal skills and long-term memory than men. However, given that the noted

differences were relatively modest, the clinical significance of these findings is uncertain.

Lack of education may be a risk factor for poorer cognitive performance (Corey-Bloom, 2004). Stewart, Richards, Brayne, and Mann (2001) studied African Caribbean subjects in the United Kingdom, and found that less education was associated with poorer performance on all CERAD subtests except verbal fluency. Karrasch and Laine (2003) reported that most CERAD subtests were affected by educational level, as are most cognitive tests. However, it is important to note that the CERAD cohort consisted mainly of highly educated NC and AD subjects and relatively few subjects with less education. In their establishment of Australian CERAD norms, Collie et al. (1999) noted that both education and gender had significant effects on performance. However, some studies have suggested that AD progresses more quickly in subjects with higher education, and that those with more education actually have an increased risk of mortality (Stern et al., 1999; Stern, Tang, Denaro, & Mayeux, 1995). Another study of demographic factors, including education and gender, found that neither significantly influenced AD progression (Bowler, Munoz, Merskey, & Hachinski, 1998).

Studies of the relationship between age of onset and AD progression have yielded mixed results. There is evidence that the progression of AD is not significantly affected by age of onset (Brocco et al., 1994; Heyman, Peterson,

Fillenbaum, & Pieper, 1996). However, Corey-Bloom (2004) reported that earlier age of onset predicts faster deterioration. Koss et al. (1996) investigated the inconsistencies found in the existing literature concerning the effect of age of onset on AD progression. The CERAD database was selected as the data source because of the sizeable amount of carefully evaluated subjects. Due to the insidious nature of dementia, a precise time of onset is difficult to establish, so these researchers used age at entry as a substitute for age of onset. Using ten-year odds ratios, the results indicated that while older subjects had poorer performance on testing, younger subjects declined at a significantly faster rate on all CERAD neuropsychological measures.

Race

With respect to racial variables, the majority of AD studies have been conducted with a limited, nonrepresentative number of ethnicities (Fillenbaum et al., 1998; Johnson, Hughes, Bullock, & Hindmarch, 2003; Stewart, Richards, Brayne, & Mann, 2001). African-American subjects have been shown to perform more poorly than Caucasian counterparts on the MMSE, constructional praxis, and naming (Welsh et al., 1995). In a study comparing the performance of community dwelling, healthy elderly African-American and white subjects on the CERAD battery, significant effects for race were not found when age, education, and gender were controlled (Fillenbaum et al., 2001). Ripich, Carpenter, and Ziol (1997) compared a small sample of African-American and Caucasian subjects

with AD on language measures (i.e. Boston Naming Test, Peabody Picture Vocabulary Test-Revised, Shortened Token Test) and found no significant differences between groups after controlling for education and MMSE score. In a study comparing the performance of Native Americans and Caucasians on the CERAD battery, researchers found no significant differences between the two groups (Whyte et al., 2005). While African Americans exhibited a slower rate of decline in score performance on CERAD neuropsychological subtests than Caucasian subjects over a one-year period, the differences were small enough to indicate that race did not strongly influence the progression of AD (Fillenbaum et al., 1998). Consistent with those findings, Heyman, Peterson, Fillenbaum, and Pieper (1996) also found no significant difference between African-American and Caucasian subjects in the rate of AD progression. However, in their 4-year study, Barnes et al. (2005) examined racial differences on a battery of nine cognitive tests, including the MMSE, Boston Naming Test, and Category Fluency. They determined that although African-Americans were more impaired at baseline than non-African Americans, African Americans declined at a 25% slower rate on average.

As of yet, no consensus has been reached regarding the influences of demographic variables such as gender, education, age of onset, or race on AD progression. Welsh et al. (1995) proposed that AD may be a “leveler” which causes pre-existing differences to be outweighed by the effects of cognitive

deterioration. Some researchers suggest that demographic variables may affect the risk of AD but not necessarily influence its progression (Bowler, Munoz, Merskey, & Hachinski, 1998; Johnson, Hughes, Bullock, & Hindmarch, 2003).

Progression of AD and CERAD

A substantial amount of AD research has been devoted to the study of its progression (Storandt, Grant, Miller, & Morris, 2002). Researchers have assessed AD progression in a variety of ways, including utilizing survival analysis to measure time to institutionalization and death (Heyman, Peterson, Fillenbaum, & Pieper, 1996; Larson et al., 2004), tracking the deterioration of clinical indicators (Galasko et al., 1995), stage transition probabilities (Neumann et al., 2001), and monitoring changes in cognitive functioning (Stern et al., 1994).

Investigation into predicted survival with AD is an important part of AD research. Research by Larson et al. (2004) utilized an AD patient registry to assess the course of illness and found strong predictors of reduced survival to be male sex, a MMSE score of 17 or less, and a BDRS score of 5.0 or greater. The male subjects had a median survival of 4.2 years from initial diagnosis, and female subjects had a median survival of 5.7 years. They determined that a 5-point or greater decline in MMSE scores after the first year of follow-up was related to a 60% increased risk of mortality, and recommended further study to assess features related to prognosis, such as annual progression. CERAD data have been studied to assess progression to institutionalization (or death as the first

event), with median time from enrollment to first event being 3.1 years. Following institutionalization, men survived 2.1 years, compared with 4.5 years for women (Heyman, et al.. 1996). However, survival analyses of this nature are limited by confounding variables such as socioeconomic status and the availability and quality of caregiving (Doody, Massman, & Dunn, 2001).

Galasko et al. (1995) investigated the use of clinical rating scales and activities of daily living as milestones in AD progression, using data from CERAD patients over a 3 year period. These researchers established clinical milestones that were unambiguous and related to illness progression rather than social or economic factors (e.g., admission to nursing home). A MMSE score of less than 10, a decline of one or more CDR stages, and loss of instrumental activities of daily living (IADLs) were found to be key milestones for tracking AD progression. Their research suggests using clinical milestones as markers of progression as an alternative to the traditional methods of survival analysis, change scores, and staging measures; however, the viability of this method in mild AD is questionable.

Neumann et al. (2001) measured disease progression in terms of transitional probabilities. Transition probabilities indicate the likelihood that a patient will move from one stage to another within a given timeframe. Using data from CERAD, these researchers estimated community-to-nursing home transitions and stage-to-stage transitions using CDR classification. They found

that the majority of community-based patients with mild AD progressed to severe AD and nursing home placement within 5 years. However, their results indicated that the length of time a patient had spent in a certain stage did not influence transition probability. Therefore, progression occurs regardless of the amount of time spent in any one stage which suggests that individuals with AD can progress rapidly or achieve temporary plateaus. Interestingly, Neumann et al. (2001) also found that 4.3% of the CERAD population experienced a "backward transition," meaning that patients moved from moderate to mild, a phenomenon that should be precluded by the progressive nature of the disease. These changes were attributed to the possible influence of medication effects or remission of depression. Overall, the authors propose using transitional probabilities as an alternative way to characterize the course of dementia.

Progression is frequently measured by differences in psychometric scores over time (Morris et al., 1993), and this approach has certain advantages over the previously discussed methods. Corey-Bloom (2004) noted that although studying AD in terms of time to institutionalization and death is valuable, these outcome measures are complex, variable, and difficult to predict. Using repeated administration of neuropsychological tests avoids the social and domestic factors associated with other outcome variables such as institutionalization, death, or clinical milestones (Bowler, Munroz, Merskey, & Hachinski, 1998). Changes in

test scores offer accurate information about the presence of a progressive dementia and the rate of decline (Locascio, Growdon, & Corkin, 1995).

In order to effectively use change scores to gauge progression, there must be a focus on enrollment, evaluation, and follow-up until death, similar to the focus kept during the development of the CERAD registry (Fillenbaum, Peterson, Welsh-Bomer, Kukull, & Heyman, 1998). Because of its large, diverse population and several years of follow-up visits, CERAD provides researchers with a rich source of data with which to study the progression of AD. Of the CERAD subtests, researchers have determined that the delayed recall task is the best measure for discriminating AD subjects from controls (Welsh et al., 1991). Tasks of verbal fluency and constructional praxis effectively distinguished between mild and moderate or severe AD (Welsh et al., 1992; Locascio, Growdon, & Corkin, 1995), which suggests that tasks of lexical-semantic processing and visuospatial ability are important in AD progression (Welsh-Bohmer et al., 1997). Although the CERAD battery has value for examining cognitive decline, progression research has used the individual subtests, not a total score.

CERAD Total Score

Total scores for measures and testing batteries serve multiple functions. A total score melds data from numerous sources into one informative piece of data that allows for quick reference, streamlined reporting, and a more complete

representation of a patient's clinical status (Sclan & Kanowski, 2001). Summary scores increase power by reducing random variability (Barnes et al., 2005). Composite scores also offer improved test-retest reliability (Olin & Zelinski, 1991) and accuracy (Rojas & Bennett, 1995). For example, the Halstead-Reitan Neuropsychological Test Battery (Reitan & Wolfson, 1993) summary score was found to have better diagnostic accuracy than its individual subtests (Rojas & Bennett, 1995), and the total score on the Dementia Rating Scale (DRS) is more reliable than the single factors (Mattis, 1988). In addition, total scores provide a more objective manner of assessing a patient's level of impairment.

The absence of a composite score for the CERAD battery rendered the results a description of several data points across multiple cognitive domains with no overall picture of level of cognitive impairment. Chandler et al. (2005) examined four methods for tabulating a total score for the CERAD battery. No one method was significantly better than the others in differentiating NC and AD subjects. Therefore, the subtest addition method [found by adding scores from the individual CERAD subtests (excluding the MMSE) into a total composite (maximum score = 100)] was selected as the preferred method due to the simplicity of calculation and absence of additional required variables or score transformations.

The resulting CERAD Total Score differentiated between NC and AD subjects as well as the MMSE, with sensitivity and specificity values above 90%.

In addition, the CERAD battery is more comprehensive and the Total Score results in fewer false negatives than the MMSE when distinguishing normal subjects from those with cognitive impairment (Chandler et al., 2005). A summary score of this nature increases the range of scores, thereby reducing floor and ceiling effects (Anastasi & Urbina, 1997.) The Total Score is likely to be more valid than any individual subtest in differential diagnosis (Chandler et al., 2005). Additionally, demographically corrected normative data for the Total Score have been provided, which may overcome the demographic effects on individual subtests. The Total Score provides a simple summary reference point for global cognitive functioning that is useful in identifying the level of impairment, differentiating between AD, Mild Cognitive Impairment (MCI), and normal aging, and as a screening measure to indicate the need for further clinical evaluation. Now equipped with a composite score, the CERAD neuropsychological battery can be fully compared to other measures, including the MMSE, CDR Sum of Boxes, and the BDRS.

Limitations of CERAD and the Total Score

The presence of ceiling and floor effects has been noted on several of the individual subtests, including the Word List Recall, Constructional Praxis, and Word List Recognition, which restricts their use by limiting accurate assessment of the range of performance (Morris et al., 1989). In addition, the battery does not include a subtest of executive functioning, a domain that is susceptible to decline

in AD (Duke & Kaszniak, 2000). The battery may fail to correctly detect AD patients with nonverbal memory difficulty due to a structure primarily founded on verbal characteristics of memory (Demadura, Delis, Jacobson, & Salmon, 2001). Although research has increased the generalizability of the CERAD normative data (Fillenbaum et al., 1990; Fillenbaum et al., 1998; Kokmen, Ozsarfati, Beard, O'Brien, & Rocca, 1996), a clear understanding of the interaction of age, education, gender, and race has not been achieved.

Only Caucasian subjects were included in the determination of the CERAD Total Score (Chandler et al., 2005); therefore, the application of the Total Score to minority populations should be interpreted with caution. Chandler et al. (2005) also pointed out that due to the average educational attainment of nearly 14 years for the CERAD normative group, subjects with less than 8 years of education were limited in representation. Conclusions drawn based on scores from extreme education levels ($\leq$ 8 years; $\geq$ 19 years) should be made with the knowledge that those with higher education may achieve higher Total Scores, thereby potentially disguising the level of impairment, and those with lower education may score lower, inflating the appearance of impairment.

Current Study

Although the CERAD Total Score is likely to increase the value and utility of the CERAD neuropsychological battery, it has not been examined in terms of AD progression. Just as the Total Score provides clinicians with a

straightforward way to summarize the degree of impairment compared to other available measures, the score could be further employed as a means of tracking AD progression. The purpose of this study was to extend the applicability of the CERAD Total Score by assessing the utility of the CERAD neuropsychological battery as a measure of progression, comparing the annualized Total Change Score to the change scores of other measures such as the MMSE, and determining the impact of demographic variables on the progression of AD as measured by the CERAD Total Change Scores.

HYPOTHESES

Overall Goal: To determine the utility of the Total Score for the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Neuropsychological Battery as a measure of cognitive progression of Alzheimer's Disease.

Aim One: To establish a meaningful degree of cognitive change by comparing the annual change scores of Normal Control subjects to those of AD subjects.

Hypothesis 1: The AD group will exhibit greater annual change in the raw CERAD Total Score than the Normal Control group from baseline to the first annual visit.

Hypothesis 2: The majority of the AD group will exhibit a degree of change from baseline to first annual visit that falls outside the 90% Confidence Interval established by the Reliable Change Index.

Aim Two: To explore how the annual CERAD Total Change Score compares to the annual rate of change in other cognitive measures in the AD group.

Hypothesis 3: The annualized CERAD Total Change Scores will moderately correlate with the annualized change scores of other cognitive measures, with the highest correlation expected with the MMSE, followed by CDR Sum of Boxes, followed by the BDRS.

Aim Three: Determine the impact of demographic variables on AD progression as measured by the CERAD Total Change Score.

Hypothesis 4: African-American AD subjects will exhibit a greater annualized change in the CERAD Total Score than Caucasian AD subjects.

Hypothesis 5: Female AD subjects will exhibit a greater annualized change in the CERAD Total Score than male AD subjects.

Hypothesis 6: AD subjects with lower education (≤ 12 years) will exhibit a greater annualized change in the CERAD Total Score than those with high education (> 12 years).

Hypothesis 7: AD subjects with a younger age at baseline will exhibit a greater annualized change in the CERAD Total Score than those AD subjects with an older age at baseline.

Exploratory Analysis: To explore the impact of dementia severity on AD progression.

Hypothesis 8: AD subjects with mild dementia (CDR = 1) will exhibit greater annualized change in the CERAD Total Score than those AD subjects with moderate to severe dementia (CDR ≥ 2) at baseline.

METHOD

Subjects

Normal Aging

CERAD Registry Controls

The NC (normal control) subjects were drawn from a pool of 465 individuals in the CERAD registry who were recruited from 24 NIA-sponsored Alzheimer's Disease Research Centers (ADRCs) or university programs in the United States between 1986 and 1995 (Division of Neurology, 2003). The control subjects were either spouses of CERAD-participating AD patients or community volunteers. The following criteria were met by all subjects, and will hereafter be referred to as the basic inclusion criteria for the study:

1) Age 50 years or older

2) Able to speak and comprehend English

3) Free of co-morbid conditions that could affect cognition or survival (i.e., major depression, alcoholism, delirium, cancer, heart/pulmonary disease, stroke).

Additional criteria were met by the CERAD registry controls:

1. Free of cognitive impairment as judged by the clinical assessment [Clinical Dementia Rating (CDR; Berg, 1988) = 0].

2. Not institutionalized.

Alzheimer's Disease

CERAD Registry AD Subjects

The AD subjects were drawn from a pool of 1,146 individuals who were diagnosed with probable or possible AD and met the basic inclusion criteria for the study. Using NINCDS-ADRDA guidelines, dementia was established by the following: a clinical examination including the MMSE (score < 25), or the Short Blessed (score > 6), the Blessed Dementia Rating Scale, and supplemented by neuropsychological testing; gradual onset and progression of symptoms, (for at least 12 months), of memory loss and at least one other area of cognitive deficit sufficient to interfere with daily activities; normal consciousness with no signs of delirium; and the absence of systemic disorders or brain pathology that in and of themselves could account for the progressive deficits of memory and cognition (McKhann et al., 1984). AD subjects also met the following inclusion criteria (Heyman, et al., 1997; Morris et al., 1989):

1) Not institutionalized
2) Possessed a caregiver or responsible informant who could provide history on the subject.

The CERAD registry included a limited number of minority subjects. The numbers of NC and AD subjects discussed above include 34 African-American and 2 Hispanic NC subjects, and 215 African-American and 21 Hispanic AD subjects. Previous studies have excluded minority subjects to provide more

uniform samples and to facilitate future comparisons (Welsh et al., 1994; Wiederholt et al., 1993; Chandler et al., 2005). Although there is relatively little consistent information on the progression of AD in minorities, Fillenbaum et al. (1998) found that problems of bias could be highly attenuated in progression studies due to the fact that in such analysis each subject serves as his or her own control. This study included data from African-American subjects; however, Hispanic subjects were excluded from future analyses due to the limited sample size.

CERAD Subjects with Available Repeat Data

Of the original 465 NC and 1146 AD subjects, the current study focused on the Caucasian and African-American subjects with a baseline (entry) assessment and at least one annual follow-up assessment. Descriptive information for these subjects is presented in Tables 1 and 2. The data through the fourth annual visit were included for a total of 5 years of testing data. At the baseline visit, there were 383 NC and 655 AD subjects. There were 342 NC and 594 AD subjects with Visit 1 data (1-year follow-up), 259 NC and 375 AD subjects with Visit 2 data, 214 NC and 204 AD subjects with Visit 3 data, and 178 NC and 98 AD subjects with Visit 4 data.

Insert Tables 1 and 2 about here

Measures

All CERAD registry subjects were administered the CERAD clinical assessment and neuropsychological battery (Morris et al., 1989) by certified personnel at entry and on each return visit. The battery includes the CDR (Berg, 1988), BDRS (Blessed et al., 1968), MMSE (Folstein et al., 1975), Verbal Fluency Test, Modified Boston Naming Test, Constructional Praxis, and Word List Memory, Recall, and Recognition. Descriptions of these individual measures are presented in Appendix A.

Procedures

The standardized administration of the CERAD battery has been described in detail by Morris et al. (1989). Subjects were encouraged to return annually for reassessment. All subjects (or their legal guardians) gave informed consent prior to testing. All testing data were collected prior to the initiation of this study, and patient names were removed from the database to ensure confidentiality.

Statistical Analyses

Calculation of the CERAD Total Score

The Total Score was calculated by summing the raw scores of the CERAD neuropsychological battery subtests, including Verbal Fluency; Modified BNT; Constructional Praxis; and Word List Learning, Recall, and Recognition Discriminability (Chandler et al., 2005). The Recognition Discriminability score was calculated by subtracting the number of false positives from the number of

true positives (range of 0 to 10 as no negative values are allowed). The addition of these subtests resulted in a Total Score with a maximum of 100 points, (with a maximum of 24 points allowed for Verbal Fluency). The resulting Total Score is composed of 39% language (Verbal Fluency and Modified BNT); 30% learning (Word List Learning), 20% memory (Word List Recall and Recognition Discriminability), and 11% construction (Constructional Praxis) (Chandler et al., 2005). The procedure for tabulating the Total Score, including maximum scores for each subtest, can be found in Table 3.

Insert Table 3 about here

Aim 1 - Calculation of the CERAD Total Change Score

Simple annual rates of change in test performance were calculated by taking the difference of the Total Score at baseline assessment from the Total Score at the proceeding assessment and dividing by the amount of time between each assessment. Therefore, negative numbers represent a decline in performance. The time period between each visit was assumed to be 12 months; therefore, the difference score was divided by one year. According to CERAD protocol, follow-ups that were done early or late were assigned visit numbers according to a 6-month rule, which stated that a follow-up completed within 6 months of its due date was assigned that visit number. The mean time between

each visit was 13.2 (*3.20*) months. The mean time and standard deviation between the baseline visit and each subject's last visit was 25.67 *(13.37)* months.

In keeping with Aim One, a meaningful degree of cognitive change was established using the repeat data to compute change score cutoff values according to the Reliable Change methodology outlined by Jacobson and Truax (1991). Reliable change indexes "take into account measurement error in a test-retest setting and specify the degree of change required to exceed sources of measurement error" (Hermann et al., 1996, p. 942). Thus, the reliable change index (RCI) provides an objectively defined criterion to characterize meaningful clinical change at the individual level that is preferable to more arbitrary approaches, such as defining change as 1 standard deviation from baseline, which may be either too conservative or too liberal for the measure in question (Sawrie, Chelune, Naugle, & Lüders, 1996).

The RCIs were calculated using data from the NC sample. First, a test-retest reliability coefficient was computed for each Change Score, from which the standard error of measurement (SE_m) was determined. The standard error of difference was derived directly from the SE_m using the formula $SE_{diff} = [2(SE_m)^2]^{1/2}$. The SE_{diff} "describes the spread of the distribution of change scores that would be expected if no actual change had occurred" (Jacobson & Truax, 1991, p. 14). The test-retest change score ($x_2 - x_1$) was divided by the SE_{diff} to create the RCI. In order to establish a 90% Change Score confidence interval (CI), the SE_{diff}

was multiplied by ±1.64 standard deviations. Thus, Change Scores exceeding the confidence interval represent a statistically reliable change that occurs only 10% of the time by chance.

To evaluate Hypothesis 1, between-group comparisons over time were analyzed using independent samples *t* tests to establish that the AD group exhibited greater annual change in raw Total Score than the NC group. To evaluate Hypothesis 2, RCIs were calculated to provide a statistical means for determining if an AD subject's Change Score differed significantly between annual visits. To further evaluate Hypothesis 2, Chi Square Goodness of Fit tests were used to determine if AD subjects that exhibited a degree of change outside the 90% confidence interval were distributed as would be expected in a normal population.

Aim 2 - Relationship with Other Measures

Annualized change scores for the CERAD Total Score, MMSE, CDR Sum of Boxes, and BDRS were calculated by taking the difference between each subject's final assessment score and the baseline assessment score and dividing by the number of total years between these assessments. In order to examine Hypothesis 3, Pearson correlations were used to test for associations between the annualized CERAD Total Change Score and the annualized change scores of the MMSE, CDR, and BDRS. A moderate correlation was defined as ± 0.4 to ± 0.6

(e.g., Guilford, 1956; Hinkle, Wiersma, & Jurs, 2003). Student's *t* test was used to determine the significance of these correlations from zero (or no association).

Aim 3 - Impact of Demographic Variables

The contribution of demographic variables (i.e., race, sex, education, and age at entry) on AD progression was examined using independent samples *t* tests for variables that were continuously and normally distributed and data that were categorical in nature were analyzed using χ^2. Any differences noted between these dichotomous or dichotomized variables at baseline were considered as possible covariates in analysis of covariance (ANCOVA). The homogeneity of slopes assumption was examined in each ANCOVA and only those significant ($p < 0.05$) are described. The impact of demographic variables were further analyzed using multiple regression in order to determine which variables contribute the most to annualized change in CERAD performance.

Exploratory Hypothesis

In order to examine the impact of dementia severity on AD progression, annualized change scores for AD subjects with a baseline CDR Stage ≤ 1 were compared to AD subjects with a baseline CDR Stage ≥ 2 using an independent samples *t* test.

The level of significance for all statistical tests used a two-tailed $p = 0.05$, and assumptions were checked for each analysis. Statistical analyses were conducted using SPSS 13.0 (SPSS Inc, Chicago, IL.)

RESULTS

Demographic Characteristics

Descriptive information for the NC and AD groups from the CERAD registry samples can be viewed in Tables 1 and 2. NC subjects were significantly younger and more educated than AD subjects. Actual mean differences for the groups were small, but moderate effect sizes were observed (age d = .41, education d = .40). There was a larger proportion of females in the NC group compared to the AD group, and in both NC and AD groups, there was a larger proportion of Caucasian subjects. In terms of dementia severity at entry, the majority of the AD sample, approximately 60%, obtained a CDR Stage of 1. The AD group entered the study with a mean MMSE score of 18.48 *(5.1)* compared to the mean MMSE of 28.81 *(1.6)* obtained by the NC group. Overall, the AD sample was largely mildly to moderately impaired, as measured by the CDR and MMSE.

Aim One

The Total Score was calculated according to the methodology set forth by Chandler et al., (2005). Each subject's CERAD Total Score was tabulated at the baseline visit and for each annual follow-up visit. The CERAD Total Score comparisons by group and visit are presented in Table 4.

Insert Table 4 about here

The progression of mean CERAD Change Scores from baseline to each follow-up visit (i.e. the difference from baseline to Visit 1, baseline to Visit 2, etc.) for AD and NC subjects is presented in Figure 1, and illustrates the relatively stable performance of the NC group compared to the steadily declining trajectory of the AD group. The change in mean Total Score from baseline to each visit for the 64 AD and 118 NC subjects with follow-up data at each time interval is shown in Figure 2. This subset of subjects shows a similar result to that found in the larger sample, with steady NC performance and declining performance in AD individuals.

Insert Figure 1 and Figure 2 about here

This pattern of significantly greater annual change in CERAD Total Score for the AD sample compared to controls was also found for each time interval (i.e. the difference between baseline and Visit 1, Visit 1 and Visit 2, etc.), as shown in Table 5.

Insert Table 5 about here

Hypothesis 1 was supported, as independent samples *t* tests for CERAD Total Scores and Change Scores from baseline to first annual visit were significantly different at the p <.001 level, with the AD group showing more annual decline in performance than the NC group. Specifically, AD subjects declined an average of 6.5 points (*SD* = 8.9), whereas the NC subjects gained an average of 2.2 points (*SD* = 6.3) at 1-year follow-up.

Reliable Change Indices

To determine a meaningful degree of change, RCIs were derived from the NC subjects according to the methodology outlined in Jacobson & Truax (1991) and reviewed in the Method section. These results are presented in Table 6.

Insert Table 6 about here

The test-retest reliability coefficients ranged from .65 to .80. The RCI was ± 10.51 for the first time interval, Baseline to Visit 1, ± 15.47 for Baseline to Visit 2, ± 11.87 for Baseline to Visit 3, ± 14.20 for Baseline to Visit 4, and ± 12.80 for Baseline to Last Visit. CERAD Total Change Scores exceeding the CI represent a statistically reliable change that would occur only 10% of the time by chance.

On the basis of the RCI, AD subjects were classified into one of three subgroups for each of the annual visits: A *gain* group – those exceeding the RCI in a positive direction; a *stable* group – those falling within the RCI boundaries; and a *decline* group – those exceeding the RCI in a negative direction. Table 7 shows the percentages of AD subjects in these three subgroups showing statistically significant change with respect to the RCI.

Insert Table 7 about here

Hypothesis 2 was not fully supported, as the majority of AD subjects did not exhibit a degree of cognitive change that exceeded the confidence intervals established by the RCI from baseline to the second annual visit. However, by Visit 3, 65.2% of the AD group exhibited a decline in Change Scores that exceeded the RCI. Furthermore, when examining the change in Total Score from baseline to each individual's last visit, the majority of AD subjects (55.1%) exceeded the cut-offs established by the RCI. A chi square test of goodness-of-fit was conducted to compare the expected and observed distribution of subjects into the three subgroups (see Table 7).

The distribution of AD subjects into the three subgroups was significantly different ($p < .001$) from the distribution of 90% stable, 5% decline, and 5% gain that would be expected from a standard population with a 90% confidence

interval. Furthermore, the distribution of NC subjects in the same three subgroups from baseline to each visit, with the exception of baseline to Visit 2, was also a non-normal distribution. A two-sided *z*-test of proportions indicated that the difference for the AD group was due to significantly greater percentages of AD subjects with a decline in performance ($p < .001$), whereas the difference for the NC group was due to significantly greater percentages of NC subjects with a gain in performance ($p \leq .03$).

Aim Two

The Total Scores of the AD sample on the MMSE, CDR Sum of Boxes, and BDRS at baseline, last visit, and the annualized Total Change Scores are presented in Table 8. Correlation coefficients were computed among the annualized change scores of these four measures and are displayed in Table 9.

Insert Tables 8 and 9 about here

The results of these analyses showed that the CERAD correlated significantly with all other measures ($p < .001$). Hypothesis 3 was supported, as the annualized CERAD Change Score had the highest correlation with the MMSE, $r(652) = .66$, $p < .001$, followed by the CDR Sum of Boxes, $r(181) = -.42$, $p < .001$, and finally the BDRS, $r(97) = -.38$, $p < .001$. It should be noted that because CDR box scores were not initially included as part of the CERAD

battery, the available sample size for the CDR Sum of Boxes comparison was 183 subjects, versus the 655 subjects available for the CERAD, MMSE, and BDRS. As anticipated, the annualized CERAD Change Score was at least moderately correlated with the annualized scores of the established measures, with the exception of the BDRS, which fell slightly short of the expected range of ± 0.4 to ± 0.6. In general, results revealed a strong positive correlation between CERAD and MMSE change scores and a moderate negative correlation between CERAD and CDR and BDRS change scores.

Aim Three

The differences in annualized Change Scores when subjects were divided based on demographic variables are presented in Figure 3.

Insert Figure 3 about here

Overall, with the exception of education level, each demographic variable was significantly related to the annualized change in CERAD performance, as measured by independent samples *t* tests. Race (black versus white) and gender (male versus female) were each significantly different at the $p < .001$ level. Age at baseline (early versus late) was also significant ($p = .007$), although education (high versus low) was not ($p = .084$).

Race

Hypothesis 4 stated that African-American AD subjects would exhibit a greater annualized change in Total Score performance than Caucasian AD subjects. While there was a significant difference in annualized change scores, $t(652) = -4.00$, $p = <.001$; CI: -4.54 to -1.55, the results were counter to the research hypothesis. African-American subjects ($M = -4.60$, $SD = 7.03$) on average exhibited *less* annualized change in CERAD Total Score performance than Caucasian subjects ($M = -7.64$, $SD = 6.82$).

An examination of demographic information in terms of race for the AD sample at baseline and last visit is presented in Table 10.

Insert Table 10 about here

The Caucasian and African-American samples differed significantly in terms of gender and education. There was a larger proportion of women in the Caucasian sample compared to the African-American group, and Caucasian subjects were better educated ($M = 12.84$, $SD = 3.45$) than their African-American counterparts ($M = 10.93$, $SD = 4.41$). In order to evaluate whether Caucasians and African-Americans differed in terms of mean annualized Change Scores when the effects of gender and education were controlled, a one-way analysis of covariance (ANCOVA) was conducted, which confirmed the initial findings that Caucasian

subjects (*M* = -7.56) exhibited greater annual change than did African-American subjects (*M* = -5.06); $F(1, 650) = 10.44$, $MSE = 484.71$, $p = .001$, partial $\eta^2 = .016$.

Gender

Hypothesis 5 stated that female AD subjects would exhibit a greater annualized change in CERAD Total Score performance than male AD subjects. Contrary to the hypothesis, female subjects' performance declined *less* (*M* = -6.44, *SD* = 7.04) than the males (*M* = -8.22, *SD* = 6.07), $t(653) = -3.27$, $p = <.001$; CI: -2.84 to -0.71.

An examination of demographic information in terms of gender for the AD sample at baseline and last visit is presented in Table 11.

Insert Table 11 about here

Male and female subjects differed significantly in terms of race, age at baseline, and education. There was a larger proportion of Caucasian subjects in the male group, and females were significantly older and less educated (*M* = 73.33, *SD* = 7.73; *M* = 11.98, *SD* = 3.45) than males (*M* = 70.51, *SD* = 7.75; *M* = 13.32, *SD* = 3.81). In order to evaluate whether males and females differed on mean annualized Change Scores when the effects of race, age, and education were controlled, an ANCOVA was conducted, which confirmed the initial findings that

male subjects (M = -7.87) exhibited more annual change than did female subjects (M = -6.70); $F(1, 648) = 4.37$, $MSE = 201.92$, $p = .037$, partial $\eta^2 = .007$.

Education

Hypothesis 6 stated that AD subjects with lower education (≤12 years) would exhibit a greater annualized change in Total Score than those with high education (>12 years). The low education group declined an average of -6.80 *(6.91)* points per year and the high education group declined an average of -7.75 *(6.95)* per year, though there was no significant difference between groups [$t(652) = 1.73$, p=.084].

An examination of demographic information in terms of education for the AD sample at baseline and last visit is presented in Table 12.

Insert Table 12 about here

It was determined that the low and high education groups differed significantly in terms of race, gender, and age at baseline. There was a larger proportion of Caucasian subjects in the high education group (>12 years) compared to the low education group (≤ 12 years). There was also a larger proportion of females in the low education group, and the low education group was significantly older (M = 72.75, SD = 7.69) than the high education sample (M = 71.24, SD = 8.04). In order to evaluate whether high and low education groups differed significantly on

mean annualized Change Scores when the effects of race, gender, and age at baseline were controlled, an ANCOVA was conducted and confirmed the initial findings that the education groups did not differ significantly in terms of mean annual change (high = -7.75, low = -6.80); $F(1, 648) = .456$, $MSE = 21.11$, $p = .500$, partial $\eta^2 = .001$.

Age at Baseline

Hypothesis 7 stated that AD subjects with a younger age at baseline (≤65) would exhibit a greater annualized change in CERAD Total Score than those AD subjects with a later age at entry (>65). Subjects with a younger age at baseline exhibited greater annualized change in Total Score performance ($M = -8.72$, $SD = 6.44$) than subjects who were older at baseline ($M = -6.85$, $SD = 7.00$); $t(652) = -2.72$, $p = .007$; CI: -3.22 to -0.52.

An examination of demographic information in terms of age at entry for the AD sample at baseline and last visit is presented in Table 13.

Insert Table 13 about here

It was determined that early and late age groups differed significantly in terms of gender and education. There was a larger proportion of females in the older age group (60%) compared to the younger age group (57%), and the younger age group was significantly more educated ($M = 13.40$, $SD = 3.25$) than the older age

group (M = 12.37, SD = 3.73). In order to evaluate whether younger and older age groups differed significantly on mean annualized CERAD Change Scores when the effects of gender and education were controlled, an ANCOVA was conducted. The ANCOVA confirmed the initial finding that subjects with a younger age at baseline exhibited more annual change (M = -8.43) than did those with older age at baseline (M = -6.91); $F(1, 649) = 4.84$, $MSE = 226.80$, $p = .028$, partial $\eta^2 = .007$.

Contribution of Demographic Variables

Multiple regression analyses were conducted to determine which demographic variables contributed the most to annualized change in the CERAD performance of AD subjects. These data are reported in Table 14.

Insert Table 14 about here

The criterion variable was the annualized CERAD Change Score from baseline to last visit. Given that each of the demographic variables examined in this study have been shown in prior research to be related to the progression of AD, the multivariate analysis included race, age at baseline, gender, and education. Based on the findings of the Hypotheses 4 and 5, the predictors were specifically Caucasian ethnicity and male gender. Age and education were entered as continuous variables. The stepwise backward method was selected in order to

minimize suppressor effects. All predictors were placed in the model and then the contribution of each was calculated by examining the significance value of the *t*-test for each predictor. If a predictor did not make a statistically significant contribution to the predictive value of the model, then it was removed from the model and the contribution of the remaining predictors was then reassessed.

The linear combination of the demographic variables (race, age, gender, and education) was significantly related to the annualized Change Score, R^2=.05, adjusted R^2 = .04, $F(4, 648) = 8.07$, $p < .001$; however, the education level variable did not predict annualized change significantly over and above the other demographic variables, R^2 change = -0.003, F change (1, 648) = 1.95, p=.164. Therefore, it was removed from the final model, which used a hierarchical approach and had a significant regression equation, R^2 = .05, adjusted R^2= .04, $F(3, 649) = 10.10$, $p < .001$. The regression formula was: Predicted Change Score = -2.64*Caucasian + 0.09* Age of Onset - 1.27*Male – 10.78*constant. These results suggest that subjects who are Caucasian, male, and have an earlier age of onset may exhibit a greater annual decline in CERAD performance; however, the model explained only 5% of the variance.

Exploratory Analysis

An additional aspect of this study was to explore the impact of dementia severity on AD progression. Hypothesis 8 stated that AD subjects with mild dementia (CDR ≤1) would exhibit greater annualized change in Total Score than

those AD subjects with moderate to severe dementia (CDR $\geq$2). This hypothesis was not supported. The mild dementia group exhibited slightly *less* annualized change (M = -7.03, SD = 6.60) than the moderate-to-severe dementia group (M = -7.50, SD = 7.50); however, the difference was not significant [$t(653) = .83$, $p = .41$], indicating that the level of dementia severity with which a subject entered the study did not impact change in CERAD performance.

An examination of demographic information in terms of dementia severity for the AD sample at baseline and last visit is presented in Table 15.

Insert Table 15 about here

It was determined that mild and moderate-to-severe groups differed significantly in terms of education. The mild AD group's CDR-SB score was significantly less (M = 5.66, SD = 1.49, range = 0.5-10) than the CDR-SB score of the moderate-to-severe group (M = 11.46, SD = 2.52, range = 8-18). The two groups differed significantly in regards to the CDR Sum of Boxes, as the mild AD group The mild AD group was significantly more educated (M = 12.89, SD = 3.70) than the moderate-to-severe AD group (M = 12.00, SD = 3.54). In order to evaluate whether the severity groups differed significantly on mean annualized Change Scores when the effects of education were controlled, an ANCOVA was conducted and confirmed the initial findings that there was no difference in terms

of mean annual change by level of severity (CDR $\leq$ 1 = -7.03, CDR $\geq$ 2 = -7.50); $F(1, 653) = 1.34$, $MSE = 1.34$, $p = .247$, partial $\eta^2 = .002$.

DISCUSSION

The CERAD neuropsychological battery provides a reliable and valid means for the assessment of cognition functioning (e.g., Morris, 1989; Welsh-Bohmer, 1997). However, until recently the battery lacked a summary score to provide an overall measure of dementia severity or global cognitive function in AD. This limitation was addressed by the tabulation of the CERAD Total Score (Chandler et al., 2005). With the availability of this composite score, the current study sought to examine the utility of the Total Score in terms of AD progression.

Aim One

Aim 1 sought to establish a meaningful degree of cognitive change by comparing the annual Change Scores of NC subjects to those of AD subjects. It was hypothesized that the AD group would exhibit greater annual change in Total Score performance than the NC group from baseline to the first annual visit. This hypothesis was supported, as AD subjects showed significantly greater annual change in Total Scores than NC subjects, which would be expected given the nature of the illness.

It was further hypothesized that the majority of the AD group would exhibit a degree of change that fell outside the 90% CI established by the RCI. To represent clinically meaningful change, the Change Score within any timeframe should exceed at least ± 10.51 points. This threshold provided by the RCI increases the likelihood that an individual's change in performance reflects

an actual change in cognitive abilities rather than extraneous factors such as testing imprecision. The mean annualized rate of AD progression as measured by the CERAD battery was determined to be -7.2 points, and it was not observed until Visit 3 that the majority of AD subjects exhibited a clinically significant decline in their performance as determined by the calculated RCIs (see Table 7).

The distribution of AD subjects into "gain", "stable", and "decline" subgroups was compared with the distribution that would be expected in unaffected individuals (see Table 7). Although the degree of change did not exceed the established RCI until Visit 3, this approach showed that at each time interval the distribution of the AD group was significantly different than that of a normal population due to significantly larger than expected percentages of decliners. Interestingly, the distribution of the NC group into the gain, stable, and decline subgroups was also significantly different than would be anticipated at some time intervals. This was driven by a significantly higher percentage of NC subjects with improved CERAD performance, which indicates that test-retest effects should be considered when administering the battery.

That a decline beyond the RCI threshold was not observed in the majority of AD subjects until later than anticipated could be due to an accelerating process of AD, such that the disease progresses more slowly in earlier stages than in subsequent years (Storhandt, Grant, Miller, & Morris, 2002). However, this idea was not supported by the findings of the current study, which indicated that the

level of dementia severity did not significantly influence AD progression. In addition, although the hallmark of AD is cognitive impairment, it may not be unusual for an individual to have a stable or slightly improved score between annual assessments due to variability in disease course or perhaps adjustments of medications (van Belle, Uhlmann, Hughes, & Larson, 1990).

Furthermore, the finding that the majority of AD subjects did not decline significantly until Visit 3 could be related to the attenuating factor of selective attrition. Lower-functioning individuals with AD may have dropped out earlier in the study than less-affected individuals, thus biasing the remaining sample with a disproportionate contribution from comparatively higher-functioning individuals. In fact, comparison of the average baseline CERAD performance of subjects with no follow-up data to those with at least one follow-up assessment revealed that those individuals who dropped out were significantly more impaired (M = 30.78, SD = 14.07) than those who remained in the study (M = 39.46, SD = 12.97). Similarly, subjects with at least one follow-up assessment had significantly better baseline MMSE performance (M = 18.50, SD = 5.10) than drop-out subjects (M = 13.71, SD = 6.72). For the 64 AD subjects who had follow-up data at each time interval, the mean annualized decline was -5.4 (*3.69*), which is 1.8 points less than the mean of the larger sample (N = 655). These results provide support for the idea that subjects who were experiencing less pronounced dementia remained in

the study, which suggests that the progression rates found by the current study may not apply to individuals with more severe dementia.

In addition, the Reliable Change Index approach to defining change is relatively conservative (Chelune et al., 1997). Thus, it may have provided overly wide confidence intervals that required AD subjects to exhibit a very large degree of change in order to be categorized as "meaningful." Therefore, Change Scores that do not exceed the specified confidence interval may not necessarily be insignificant but should be interpreted within the context of other sources of data (Chelune et al., 1997).

Aim Two

The second aim of this study explored how the annual CERAD Change Score compared to the annual rate of change in the MMSE, CDR Sum of Boxes, and BDRS, which are all established measures for the assessment of dementia and the measurement of cognitive decline (Berg, 1988; Blessed et al., 1968; Folstein et al., 1975). As shown in Table 8, the mean annualized rate of change exhibited by the AD sample was -7.20 *(6.93)* on the CERAD, -3.43 *(3.33)* on the MMSE, 2.28 *(2.13)* on the CDR, and 1.89 *(1.54)* on the BDRS. The mean MMSE rate of change is in line with prior research findings that MMSE scores decline between 2 and 5 points per year (e.g., Clark et al., 1999; Van Belle, Uhlmann, & Hughes, 1990). In addition, the mean BDRS rate of change is similar to the average decline of 1.3 points per year reported by Holmes & Lovestone (2003).

The CERAD Change Score should correlate with these measures if it is a viable tool for examining the progression of AD; therefore, it was hypothesized that the annualized CERAD Change Score would moderately correlate with the annualized change in the other cognitive measures. This hypothesis was mostly supported, as the CERAD Change Score was significantly correlated with each measure within the moderate range (.4 - .6), with the exception of the BDRS, which fell slightly short of that range (r = -.38) but was still significant (see Table 9).

Both the CERAD and the MMSE assess global cognitive functioning; however, the CDR and BDRS are clinician-rated scales that focus on functional capacity and activities of daily living (ADLs) related to dementia. Furthermore, the CERAD and the MMSE are based on information gathered from direct assessment of the patient, whereas the CDR is partially based and the BDRS fully based on caregiver interview. Given these differences, it was hypothesized that the CERAD Total Change Score would have the highest correlation with the MMSE, followed by the CDR Sum of Boxes, followed by the BDRS. This hypothesis was supported, which provides support for the concurrent validity of the CERAD battery in measuring progression. The results also provide comparative data for annualized change in the CERAD Total Score and other common measures in the assessment of dementia.

However, the strength of the mean annualized total score correlations were somewhat lower than anticipated and fluctuated when examined by level of baseline dementia severity. The correlation between CERAD and the MMSE decreased from r = .67 at CDR Stage.5 to r = .60 at CDR Stage 3 but remained at least moderately strong. The correlation between CERAD and the CDR Sum of Boxes decreased from r = -.54 at CDR Stage .5 to r = -.35 at CDR Stage 3. More dramatically, the correlation between CERAD and the BDRS dropped from r = -.42 at CDR Stage .5 to r = -.13 at CDR Stage 3. Similarly, the correlation of the MMSE and BDRS exhibited a decline in strength from r = -.36 to r = -.18; however, the correlation between the MMSE and CDR remained stable (r = .30). Interestingly, the correlation between the CDR and BDRS declined but remained highly correlated with r = .91 at CDR Stage .5 and r = .73 at CDR Stage 3. This suggests that the measurement of AD progression in this population for individuals with more severely impaired functioning may be more problematic, and likely contributed to the lower than expected correlation coefficients. In addition, since the CERAD and MMSE assess global cognitive functioning versus the CDR and BDRS' focus on functional capacity and ADLs, the scores should not necessarily be expected to covary as the disease progresses.

Aim Three

The third aim of this study focused on the impact of demographic variables on AD progression as measured by the annualized CERAD Total

Change Score. The influence of demographic variables was considered in the current study for two reasons: 1) Literature on the CERAD neuropsychological battery (Chandler et al., 2005; Fillenbaum, 2001; Henderson & Buckwalter, 1994; Welsh, Butters et al., 1994) has shown that performances on the CERAD were affected by such variables as gender, age, and education, and 2) in the CERAD registry, the NC sample was significantly younger, more educated, and had a higher percentage of females than the AD sample.

Race

The influence of race on AD progression is controversial, with studies noting different effects on the rate of progression (Barnes et al., 2005; Fillenbaum et al., 2001; Whyte et al., 2005). In this study, Caucasian subjects exhibited more annual decline in CERAD Total Score than African-American subjects, even after controlling for gender and education. In keeping with prior research that suggests African-Americans may perform more poorly at baseline (Barnes et al., 2005; Welsh-Bohmer et al., 1997), African-Americans were significantly more impaired on baseline MMSE and baseline CERAD Total Scores than their Caucasian counterparts (see Table 10). However, by the last visit assessment, those differences had disappeared. This indicates that although the Caucasian group started out with less impairment, they experienced accelerated decline and essentially "caught up" with the African-American group. This is consistent with the idea that AD may be a "race leveler," such that pre-existing differences and

advantages between groups are outweighed by the dementing process (Welsh et al., 1995).

Although CERAD made a concerted effort to enroll minority subjects, the available sample size of African-Americans was still relatively small (N = 95). The African-American sample had significantly lower baseline Total Scores than the Caucasian sample (34.54 versus 40.16), so it is possible that Caucasian subjects had further to decline whereas the African-American subjects were already in more advanced stages of dementia. Thus, floor effects may have artificially reduced the estimate of cognitive decline in African-Americans. It is not uncommon for African-American individuals to seek treatment for cognitive concerns at a later stage in comparison to Caucasians (Fillenbaum et al., 1998). In fact, comparison of baseline CDR Stage by race showed that 10.5% of African-Americans were already severely demented (CDR = 3) versus 2.5% of Caucasians. The rate of AD progression could be better assessed in a sample of equally impaired Caucasian and African-American individuals at baseline.

Gender

The nature of gender variations in cognitive performance is as yet unclear in the literature. It was anticipated that women would have poorer baseline performance on the CERAD Total Score based on the finding that AD women obtained lower scores than AD men on CERAD subtests involving verbal skills and long-term memory, including the Modified BNT, Verbal Fluency, and Word

List Savings Score and Delayed Recall (Henderson & Buckwalter, 1994). Prior findings suggested that despite men's increased risk of mortality, women experience more severe impairment and faster decline (Lapane et al., 2001).

However, the current study found that men exhibited a greater amount of decline than women, even after controlling for race, age at baseline, and education (see Table 11). Females in the sample had significantly lower Total Scores at baseline than the males, so it is possible that male subjects had further to decline, whereas the female subjects were already in more advanced stages of dementia. In a sample of equally impaired male and female individuals at baseline, the rate of AD progression could be better assessed.

Education

In the current study there was no significant difference in annualized change between high and low education groups. Although the high education group (>12 years) had significantly higher baseline CERAD and MMSE scores than the low education group (≤ 12 years), by the last visit there were no significant differences (see Table 12). While lack of education may be a risk factor for AD (Corrie-Bloom, 2004; Karrasch & Laine, 2003; Stewart et al., 2001), it did not appear that education played a major role in the progression of AD in this sample. It may be that individuals with higher education are more prone to recognize cognitive changes and seek medical attention at an earlier point than their lower-educated counterparts, which could account for the

differences in performance at baseline, rather than assuming that education provides a protective factor.

Age at Baseline

A number of studies have shown that individuals with early-onset AD have a faster rate of decline than those with later onset (e.g. Jacobs et al., 1994; Koss et al., 2006). Due to the insidious nature of dementia, a precise time of onset is difficult to establish, so the current study used age at baseline as a substitute for age of onset as has been done in prior research (Koss et al., 1996). It was hypothesized that AD subjects with a younger age at baseline (≤ 65 years) would exhibit a greater annualized change in CERAD Total Score than AD subjects with an older age at baseline (> 65 years). Although both samples performed equally on the MMSE and CERAD at baseline, the present findings confirm prior indications of accelerated cognitive decline in individuals with early-onset AD (see Table 13).

Contribution of Demographic Variables

In order to determine which demographic variables contributed most to annualized change in CERAD performance, regression analysis was conducted. Each of the four variables examined in this study, age at baseline, race, gender, and education, were selected as predictors. Each contributed significantly to the predictive value of the regression model with the exception of education level. This was not surprising given that there was no significant difference found

between the high and low education groups. Only race, age at baseline, and gender were included in the second model. The model was significantly better at predicting annualized change than using the mean as a "best guess," and indicated that male Caucasians with a younger age at baseline may exhibit greater annual decline in CERAD performance.

However, the overall predictive value of the regression model was relatively weak, as it explained approximately 5% of the variance (see Table 14). This suggests that other factors contribute to AD progression as measured by the CERAD Total Change Score. For example, prior literature has identified marital status as a possible factor, indicating that being unmarried is a risk factor for AD and may play a role in the dementing process (Helmer et al., 1999; Heyman et al., 1997). These results also imply that perhaps individual characteristics are not that important in terms of disease progression. Patterns of change over time on cognitive tests reflect an underlying biological process of progressive neuronal dysfunction in AD that may not be significantly influenced by demographic variables. Overall, although demographic variables may not play a large role in AD progression, they are a factor that should be considered when devising and employing prediction models.

Exploratory Analysis

Results indicated that the level of dementia severity with which subjects entered the study did not significantly impact changes in annualized CERAD

performance. This is surprising given the substantial amount of research identifying dementia severity as a key factor in AD progression (e.g. Storhandt et al., 2002). Although the mild severity group (CDR Stage ≤ 1) had significantly higher CERAD and MMSE total scores than the moderate to severe group (CDR Stage ≥ 2) at baseline and last visit, there was no significant difference in the rate of decline (see Table15). This suggests that AD progresses at similar rates regardless of initial severity based on group data. However, there may have been individual variability that was not represented in this analysis. Furthermore, this finding could be attenuated by the restricted range of CDR scores at baseline (.5 – 3.0), and the fact that the majority of subjects (59%) were classified at CDR Stage 1, limiting the number of subjects in other stages that were available for comparison. Given that the AD sample was fairly impaired at baseline, there may not have been much room for further decline, resulting in floor effects.

Limitations of the Current Study

This study was subject to the inherent limitations of longitudinal designs, namely, selective attrition likely affected sample composition at follow-up assessment visits, resulting in successively smaller *n*s and inconsistency in subjects returning at each visit. In addition, while the study involved the use of longitudinal data, follow-up data for the entire sample were not available because not every subject was examined at each time interval. This study was limited to Caucasian and African-American subjects, so the CERAD Total Score's utility as

a measure of cognitive assessment and/or progression among other minority populations is unknown. As with all studies utilizing the CERAD registry, an additional limitation to this study is the relatively high level of education in both AD and NC groups (M = 13.3 years of education). It is possible that individuals with less than 8 years of education manifest differently in terms of progression than the results reported here.

Furthermore, it should be noted that the CERAD Total Score is made up of subtests that do not assess all possible areas of cognitive impairment in AD, specifically executive functioning and nonverbal memory are lacking. It is possible that Change Scores may not fully estimate the cognitive decline in individuals with prominent deficits in executive functioning or nonverbal memory.

The RCI method of determining clinically significant change in scores utilized in this study is well researched. However, certain aspects of this method have been criticized. As noted above, overly wide confidence intervals may result in missed detection of individuals whose cognitive functioning did change meaningfully over the follow-up period (Temkin et al., 1997). Although the confidence interval provided by RCIs is useful, once the degree of change extends beyond the critical range it does not provide information regarding the relative magnitude of the change (Hermann et al., 1996). Given that in a sample of AD affected individuals each annual assessment would be expected to reveal

significant change in performance, it may be that the RCI method is too conservative for this population and not the best approach for assessing AD progression. Suggested improvements on the RCI method include the use of practice effect corrections to increase prediction accuracy (Chelune et al., 1993) or the use of regression-based norms to account for both practice effects and examine the magnitude of change (Sawrie, Chelune, Naugle, & Lüders, 1996).

Future Directions

The current study established the CERAD Total Score as a measure of AD progression comparable to such established measures as the MMSE, CDR, and BDRS. Additional research geared toward exploring the utility of the CERAD Total Score in the staging of AD would be valuable. The establishment of cut-off scores for mild, moderate, and severe dementia, along with normative data, would allow the CERAD battery to be used as a dementia staging tool and be useful for comparing CERAD performance to performance on other cognitive measures. In addition, the CERAD battery could be further employed in drug studies with AD populations by examining the change in CERAD Total Score in order to assess the effectiveness of pharmacological or other interventions.

Conclusions

AD is a chronic neurodegenerative disorder and the most common cause of dementia in the United States, making the reliable measurement and prediction of cognitive decline of critical concern. The CERAD neuropsychological battery

is established as a reliable and valid means of assessing cognitive dysfunction in AD, and was recently improved upon by the addition of a Total Score. CERAD Total Change Scores demonstrated concurrent validity with the MMSE, CDR, and BDRS, providing comparative data for annualized change. The examination of demographic variables provides further information regarding the influence of individual characteristics on levels of cognitive functioning. This study further extended the utility of the CERAD neuropsychological battery by establishing it as a valid measure of progression by deriving annualized Total Change Scores and corresponding confidence intervals with which to define clinically meaningful change.

APPENDIX A

CERAD Neuropsychological Battery

Verbal Fluency (Adapted from Isaacs & Kennie, 1973)

This is a measure of verbal fluency within a semantic category. The original test (Isaacs & Kennie, 1973) consisted of four sets (colors, animals, fruits, and towns). Only the animal naming set was included in the CERAD neuropsychological battery. First a practice trial is given in which the subject is asked to name examples of articles of clothing. If the subject successfully names two examples (additional instruction can be given at this stage), the scored set is administered in which he or she is asked to name as many animals as possible in one minute. The total score is calculated by summing the number of words the subject produces each fifteen seconds. Perseverations or losses of set do not receive credit. Perseverations are repetitions of the same word, and losses of set are productions of words that are not animals. One-month test-retest correlations of $r = .67$ in an elderly control sample and $r = .73$ in a mild AD sample have been reported (Morris et al., 1989). The control sample had a mean performance of 18.0 ($SD = 4.8$) and subjects with mild dementia obtained a mean score of 8.8 ($SD = 3.9$) (Morris et al., 1989). The Verbal Fluency test has been established as a useful tool for differentiating mild AD from normal aging and staging the level of dementia (Welsh et al., 1992).

Modified Boston Naming Test (BNT; Adapted from Kaplan et al., 1978)

The Modified BNT is a 15-iem measure of confrontation naming ability that contains five high, medium, and low frequency items from the original BNT. The examiner presents the subject with a series of black and white drawings and asks him or her to provide the name of the object pictured. Ten seconds are allowed for each picture. The subject can be prompted if he or she provides an ambiguous answer (e.g., "is there another name for that?" if the word "boat" was provided for "canoe"). Each correct response is given one point. One-month test-retest correlations of r = .56 in an elderly control sample and r = .89 in a mild AD sample have been reported (Morris et al., 1989). The control sample had a mean baseline performance of 14.6 (SD = 0.6) and subjects with mild dementia obtained a mean score of 11.8 (SD = 2.7). The Modified BNT has been established as a useful tool for differentiating mild AD from normal aging when used in conjunction with a delayed recall measure (Welsh et al., 1992).

Constructional Praxis (Adapted from Rosen et al., 1984)

Constructional Praxis is a test of visuoconstructional ability using four simple geometric figures from the original work of Rosen et al. (1984). The subject is asked to copy figures of increasing complexity (circle, diamond, overlapping rectangles, and a cube). The training guide for CERAD administration provides scoring criteria with a maximum of 11 points. One-month test-retest correlations of r = .54 in an elderly control sample and r = .78 in

a mild AD sample have been reported (Morris et al., 1989). The control sample had a mean performance of 10.1 (SD = 1.2) and subjects with mild dementia obtained a mean score of 7.8 (SD = 2.3). Constructional Praxis has been established as a useful tool in the differential staging of mild and moderate AD when used in conjunction with Word List Recall (Welsh et al., 1992).

Word List Tasks (Adapted from Atkinson et al., 1971)

Learning

Word List Memory or Word List Learning is a measure of verbal learning. Ten nouns are simultaneously read aloud by the examiner and shown from a booklet one at a time for two seconds to ensure that the subject understands each word. The subject is then allowed 90 seconds to tell the examiner as many words as he or she can remember. The same 10 words are presented in a different order for three trials. Recall is assessed after each trial. Words said by the subject that were not on the word list are recorded as intrusion errors. There is a maximum score of 30 points for this task. One-month test-retest correlations of $r = .62$ in an elderly control sample and $r = .68$ in a mild AD sample have been reported (Morris et al., 1989). The control sample had a mean performance of 21.1. (SD = 3.7) and subjects with mild AD obtained a mean score of 8.8 (SD = 4.1).

Recall

Word List Recall is a test of delayed verbal recall. Constructional Praxis is used as a distracter item between Word List Learning and Recall. Subjects are

asked to provide as many items as they can recall from the Word List Learning task. One point is earned for each of the 10 words recalled, and intrusion errors are recorded. One-month test-retest correlations of r = .64 in an elderly control sample and r = .60 in a mild AD sample have been reported (Morris et al., 1989). The control sample had a mean performance of 7.2 (*SD* = 1.8) and subjects with mild AD obtained a mean score of 0.9 (*SD* = 1.4).

Recognition

The Word List Recognition task assesses how many words from the learning trials the subject can discriminate from a list of 20 words. One point is given for each item the subject correctly identifies as being on the list (Total Yes Correct) or not on the list (Total No Correct), resulting in a maximum of 10 Total Yes Correct and 10 Total No Correct. Morris et al. (1989) added together the Total Yes and No Correct and subtracted 10, resulting in a number from 0-10. These authors used this method in a study comparing elderly control subjects to subjects with mild AD. One-month test-retest correlations of r = .36 in the control sample and r = .52 in the mild AD sample and mean performances of 9.6 (SD = 0.8) in controls and 4.8 (SD = 2.7) in subjects with mild AD were reported. Measures of memory, particularly delayed recall, have been established as useful tools to distinguish AD from normal aging (Peterson et al., 1999; Welsh et al., 1991). However, the floor effects of these measures may limit their utility in the staging of dementia severity (Welsh et al., 1992).

TABLES

Table 1

Characteristics of the CERAD Registry Sample (Continuous Variables)

	AD (N = 655) *M (SD)*	NC (N = 383) *M (SD)*	Statistic	*p*
Age at Baseline	72.1 *(7.9)*	68.9 *(7.9)*	$t(1036) = 6.30$	<.001
Education Level	12.6 *(3.7)*	13.9 *(3.1)*	$t(1036) = -6.13$	<.001
Baseline CERAD	39.4 *(13.0)*	79.8 *(9.4*	$t(1036) = -53.21$	<.001
Baseline MMSE	18.48 *(5.1)*	28.81 *(1.6)*	$t(1036) = -38.80$	<.001
Baseline BDRS[a]	4.3 *(2.4)*	-	-	-

[a] The BDRS was not administered to NCs.

Table 2

Characteristics of the CERAD Registry Sample (Categorical Variables)

	AD (N = 655)	NC (N = 383)
	N *(%)*	N *(%)*
Age at Baseline[^]		
Younger (≤65)	125 *(19)*	111 *(29)*
Older (>65)	530 *(81)*	272 *(71)*
Education		
Low (≤12)	381 *(58)*	163 *(43)*
High (>12)	274 *(42)*	220 *(57)*
Gender		
M	281 *(43)*	129 *(34)*
F[*]	374 *(57)*	254 *(66)*
Race [* ^]		
Caucasian	560 *(86)*	360 *(94)*
Black	95 *(14)*	23 *(6)*
CDR Stage[a]		
.5	25 *(3.8)*	-
1	386 *(58.9)*	-
2	220 *(33.6)*	-
3	24 *(3.7)*	-

[^] Significant difference within AD and NC groups at $p < .001$ using χ^2.

[*] Significant difference between AD and NC groups at $p < .001$ using χ^2.

[a] All NCs were CDR Stage 0.

Table 3

CERAD Total Score Tabulation Method

Subtest	Maximum Points Possible
Verbal Fluency	24[a]
Modified BNT*	15
Word List Learning	30
Constructional Praxis	11
Word List Recall	10
Word List Recognition Discriminability	10[b]
Total Score	100

Note. [a] Verbal Fluency does not have a ceiling when administered using standard instructions. For calculation purposes, a cap of 24 was placed on Verbal Fluency, which represents one standard deviation above the normal aging population mean per Welsh, Butters et al. (1994).

[b] Recognition Discriminability was calculated by subtracting the number of false positives from the number of true positives (range = 0 to 10, no negative values are allowed).

* BNT = Boston Naming Test.

Table 4

Mean CERAD Total Score by Group and Visit

	AD M *(SD)* N	NC M *(SD)* N	Statistic	p value
Baseline	39.4 *(13.0)* 655	79.8 *(9.4)* 383	*t* (1036) = -53.21	< .001
Visit 1	33.2 *(15.5)* 594	81.8 *(10.1)* 342	*t* (934) = -51.10	< .001
Visit 2	28.8 *(16.2)* 375	82.0 *(11.3)* 259	*t* (632) = -45.54	< .001
Visit 3	25.4 *(17.8)* 204	83.3 *(10.4)* 214	*t* (416) = -40.77	< .001
Visit 4	24.5 *(17.5)* 98	83.8 *(11.0)* 178	*t* (274) = -34.52	< .001

Table 5

Mean CERAD Total Change Score by Group and Visit

	AD M *(SD)* N	NC M *(SD)* N	Statistic	p value
Visit 1 - Baseline	-6.5 *(8.9)* 594	2.2 *(6.3)* 342	*t* (934) = -15.82	< .001
Visit 2 - Visit 1	-6.9 *(7.9)* 321	-0.4 *(7.9)* 238	*t* (557) = -9.89	< .001
Visit 3 - Visit 2	-6.5 *(8.0)* 162	1.0 *(8.0)* 169	*t* (329) = -8.41	< .001
Visit 4 - Visit 3	-6.1 *(6.9)* 76	0.4 *(5.8)* 142	*t* (216) = -7.45	< .001
Annualized CERAD Total Change Score [a]	-7.2 *(6.9)* 655	1.0 *(3.4)* 383	*t* (1036) = -21.78	< .001

[a] [(Last Visit – Baseline)/(# Years in Between)].

Table 6

Test-Retest Reliability Coefficients and Reliable Change Index Scores Based on Standard Errors of Measurement for the CERAD NC Subjects

	N	Test-Retest Reliability	Standard Error of Measurement	Standard Error of Difference	RC (90%) Confidence Interval
Baseline to Visit 1	342	.80	4.53	6.41	± 10.51
Baseline to Visit 2	259	.65	6.67	9.43	± 15.47
Baseline to Visit 3	214	.76	5.12	7.24	± 11.87
Baseline to Visit 4	178	.69	6.12	8.66	± 14.20
Baseline to Last Visit	383	.74	5.52	7.80	± 12.80

Note. Test-retest reliability coefficients based on the correlation between the mean CERAD Total Score at baseline, (M = 79.8, SD = 9.4), and each subsequent visit.

Table 7

Chi Square Goodness of Fit Test Showing Observed Distributions of AD and NC Subjects into Gain, Stable, and Decline Subgroups Based Upon the RCIs

	AD				NC			
	Gain N *(%)*	Stable N *(%)*	Decline N *(%)*	χ^2	Gain N *(%)*	Stable N *(%)*	Decline N *(%)*	χ^2
V1-V0[a]	13 *(2)*	411 *(69)*	170 *(29)*	700.73**	26 *(8)*	308 *(90)*	8 *(2)*	9.47*
V2-V0	4 *(1)*	229 *(61)*	142 *(38)*	856.65**	8 *(3)*	244 *(94)*	7 *(3)*	5.14
V3-V0	5 *(2)*	66 *(33)*	133 *(65)*	1556.39**	20 *(9)*	190 *(89)*	4 *(2)*	12.31*
V4-V0	0 *(0)*	33 *(34)*	65 *(66)*	776.59**	16 *(9)*	156 *(88)*	6 *(3)*	6.72*
LV - V0	10 *(2)*	305 *(43)*	340 *(55)*	3035.63**	29 *(7)*	340 *(89)*	14 *(4)*	6.52*

[a] V0 denotes Baseline Visit, V1 through V4 denotes the first annual visit through the fourth annual visit, LV denotes Last Visit

** χ^2 is significant at $p \leq .0001$.

* χ^2 is significant at $p \leq .03$.

Table 8

Total Scores of AD Sample on Summary Measures

	Total Score at Baseline M *(SD)*	Total Score at Last Visit M *(SD)*	Annualized Total Change Score M *(SD)*
CERAD	39.35 *(13.01)*	25.41 *(16.37)*	-7.20 *(6.93)*
MMSE	18.48 *(5.07)*	11.18 (*7.46)*	-3.43 *(3.33)*
CDR	7.43 *(3.26)*	12.71 *(4.98)*	2.28 *(2.13)*
BDRS	4.30 *(2.36)*	9.82 *(4.32)*	1.89 *(1.54)*

Note. $N = 655$ for all measures except the CDR ($N = 183$).

MMSE: Mini Mental State Examination

CDR: Clinical Dementia Rating Scale Sum of Boxes

BDRS: Blessed Dementia Rating Scale

Table 9

Pearson Correlation Coefficients Among CERAD, MMSE, CDR, BDRS Annualized Change Scores in AD Sample

	MMSE	CDR[1]	BDRS
CERAD	*r* = .66*	*r* = -.42*	*r* = -.38*
MMSE		*r* = -.42*	*r* = -.41*
CDR[1]			*r* = .60*

[1] Clinical Dementia Rating Scale Sum of Boxes

*p < .001

Table 10

Demographic Information for AD Subjects by Race

	Caucasian M *(SD)* (N = 559)	African-American M *(SD)* (N = 95)	Statistic	*p*
Gender				
Male (%)	46%	25%	$\chi^2(1) = 14.22$	<.001
Female (%)	54%	75%		
Education	12.84*(3.45)*	10.93 *(4.41)*	$t(114) = 4.02$	<.001
Age at Baseline	71.92 *(7.99)*	73.28 *(6.99)*	$t(652) = -1.56$	.120
Baseline MMSE	18.69 *(4.89)*	17.26 *(5.94)*	$t(117) = 2.21$	.029
Last Visit MMSE	10.90 *(7.43)*	12.84 *(7.48)*	$t(652) = -2.35$	.019
Baseline CERAD	40.16 *(12.74)*	34.54 *(13.57)*	$t(652) = 3.94$	<.001
Last Visit CERAD	25.26 *(16.18)*	26.33 *(17.53)*	$t(652) = -.589$	.556
MMSE Annualized Total Change Score	-3.63 *(3.39)*	-2.94 *(2.64)*	$t(652) = -3.78$	<.001
CERAD Annualized Total Change Score	-7.64 *(6.82)*	-4.60 (*7.03*)	$t(652) = -4.00$	<.001

Table 11

Demographic Information for AD Subjects by Gender

	Male M *(SD)* (N = 281)	Female M *(SD)* (N = 374)	Statistic	*p*
Race				
Caucasian (%)	91%	81%	$\chi^2(1) = 14.22$	<.001
African-American (%)	9%	19%		
Education	13.32 *(3.81)*	11.98 *(3.45)*	$t(570) = 4.64$	<.001
Age at Baseline	70.51 *(7.75)*	73.33 *(7.73)*	$t(652) = -4.61$	<.001
Baseline MMSE	18.82 *(4.80)*	18.22 *(5.26)*	$t(652) = 1.49$	.137
Last Visit MMSE	10.92 *(7.23)*	11.39 *(7.63)*	$t(653) = -.793$	.428
Baseline CERAD	41.69 *(12.30)*	37.58 *(13.26)*	$t(653) = 4.05$	<.001
Last Visit CERAD	25.83 *(16.02)*	25.09 *(16.65)*	$t(653) = .572$	.568
MMSE Annualized Total Change Score	-3.76 *(3.29)*	-3.18 *(3.33)*	$t(653) = -2.23$	.026
CERAD Annualized Total Change Score	-8.22 *(6.70)*	-6.44 *(7.04)*	$t(653) = -3.27$	<.001

Table 12

Demographic Information for AD Subjects by Education Level

	Low (≤12) M *(SD)* (N = 381)	High (>12) M *(SD)* (N = 273)	Statistic	*p*
Gender				
Male (%)	36%	53%	$\chi^2(1) = 20.89$	<.001
Female (%)	64%	47%		
Race				
Caucasian (%)	82%	90%	$\chi^2(1) = 9.53$	.002
African-American (%)	18%	10%		
Age at Baseline	72.75 *(7.69)*	71.24 *(8.04)*	$t(652) = 2.44$	.015
Baseline MMSE	17.71 *(5.26)*	19.60 *(4.54)*	$t(628) = -4.92$	<.001
Last Visit MMSE	10.95 *(7.20)*	11.53 *(7.81)*	$t(652) = -.971$	.332
Baseline CERAD	37.32 *(13.10)*	42.24 *(12.32)*	$t(652) = -4.86$	<.001
Last Visit CERAD	24.48 *(16.15)*	26.77 *(16.62)*	$t(652) = 1.77$	.077
MMSE Annualized Total Change Score	-3.22 *(3.25)*	-3.75 *(3.42)*	$t(652) = 1.93$	.054
CERAD Annualized Total Change Score	-6.80 *(6.91)*	-7.75 *(6.95)*	$t(652) = 1.73$	.084

Table 13

Demographic Information for AD Subjects by Age at Baseline

	Younger (≤ 65) M *(SD)* (N = 124)	Older (> 65) M *(SD)* (N = 529)	Statistic	*p*
Gender				
Male (%)	54%	40%	$\chi^2(1) = 7.77$	.005
Female (%)	46%	60%		
Race				
Caucasian (%)	90%	85%	$\chi^2(1) = 1.90$	.168
African-American (%)	10%	15%		
Education	13.40 *(3.25)*	12.37 *(3.73)*	$t(651) = 2.85$	.005
Baseline MMSE	17.88 *(5.12)*	18.61 *(5.06)*	$t(651) = -1.45$	.147
Last Visit MMSE	8.71 *(6.87)*	11.76 *(7.49)*	$t(197) = -4.37$	<.001
Baseline CERAD	39.74 *(12.92)*	39.27 *(13.04)*	$t(651) = .365$	.715
Last Visit CERAD	21.70 *(15.94)*	26.28 *(16.38)*	$t(651) = -2.81$	.005
MMSE Annualized Total Change Score	-4.07 *(3.40)*	-3.27 *(3.29)*	$t(651) = -2.44$	.015
CERAD Annualized Total Change Score	-8.72 *(6.44)*	-6.85 *(7.01)*	$t(651) = -2.72$	.007

Table 14

Summary of Multiple Regression Analysis for Variables Affecting Annualized CERAD Total Change Score of AD Sample (N = 653)

Variable	*B*	*SE B*	β	*p*	Model R^2	Incremental R^2 Change
Model 1					.05	N/A
Race (White)	-2.47	0.78	-.13	.002		
Age at Baseline	0.08	0.04	.09	.020		
Gender (Male)	-1.16	0.56	-.08	.037		
Education	-0.11	0.08	-.06	.164		
Model 2					.05	-.003
Race (White)	-2.64	0.77	-.13	.001		
Age at Baseline	0.09	0.03	.10	.010		
Gender (Male)	-1.27	0.55	-.09	.022		

Table 15

Demographic Information for AD Subjects by Dementia Severity

	Mild[a] M (SD) (N = 411)	Mod - Severe [b] M (SD) (N = 244)	Statistic	*p*
Race				
Caucasian (%)	85%	86%	$\chi^2(1) = .102$	.750
African-American (%)	15%	14%		
Gender				
Male (%)	45%	39%	$\chi^2(1) = 2.50$	.114
Female (%)	55%	61%		
Education	12.89 *(3.70)*	12.00 *(3.54)*	$t(652) = 3.04$	.002
Age	71.79 *(7.78)*	72.68 *(7.97)*	$t(652) = -1.40$	.162
Baseline CDR-SB[c]	5.66 *(1.49)*	11.46 *(2.52)*	$t(181) = -19.39$	<.001
Last Visit CDR-SB	13.18 (5.74)	17.10 (5.40)	$t(588) = -8.19$	<.001
Baseline MMSE	20.39 *(3.99)*	15.24 *(5.07)*	$t(652) = 14.40$	<.001
Last Visit MMSE	13.19 *(7.49)*	7.80 *(6.07)*	$t(653) = 9.55$	<.001
Baseline CERAD	43.98 *(11.01)*	31.54 *(12.38)*	$t(653) = 13.34$	<.001
Last Visit CERAD	29.74 *(16.06)*	18.11 *(14.18)*	$t(653) = 9.35$	<.001
MMSE Annualized Total Change Score	-3.28 *(3.48)*	-3.67 *(3.03)*	$t(653) = 1.46$	.146
CERAD Annualized Total Change Score	-7.03 *(6.60)*	-7.50 *(7.50)*	$t(653) = .834$	.405

[a] CDR Stage ≤1

[b] CDR Stage ≥2

[c] CDR-SB = CDR Sum of Boxes score.

FIGURES

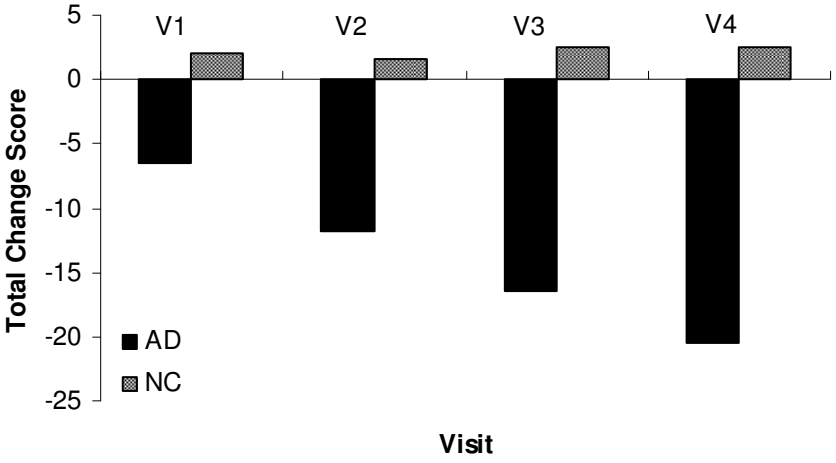

Figure 1. Mean CERAD Total Change Score from baseline to each follow-up visit for AD and NC groups. (V1 through V4 denotes the first annual visit through the fourth annual visit.). AD $N = 655$, NC $N = 383$.

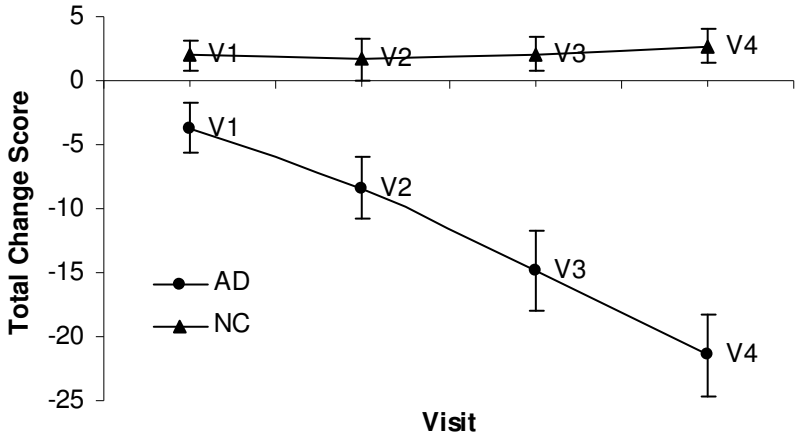

Figure 2. Mean CERAD Total Change Score from baseline to each follow-up visit for AD and NC subjects with follow-up data at each time interval. (V1 through V4 denotes the first annual visit through the fourth annual visit.). AD N = 64, NC N = 118.

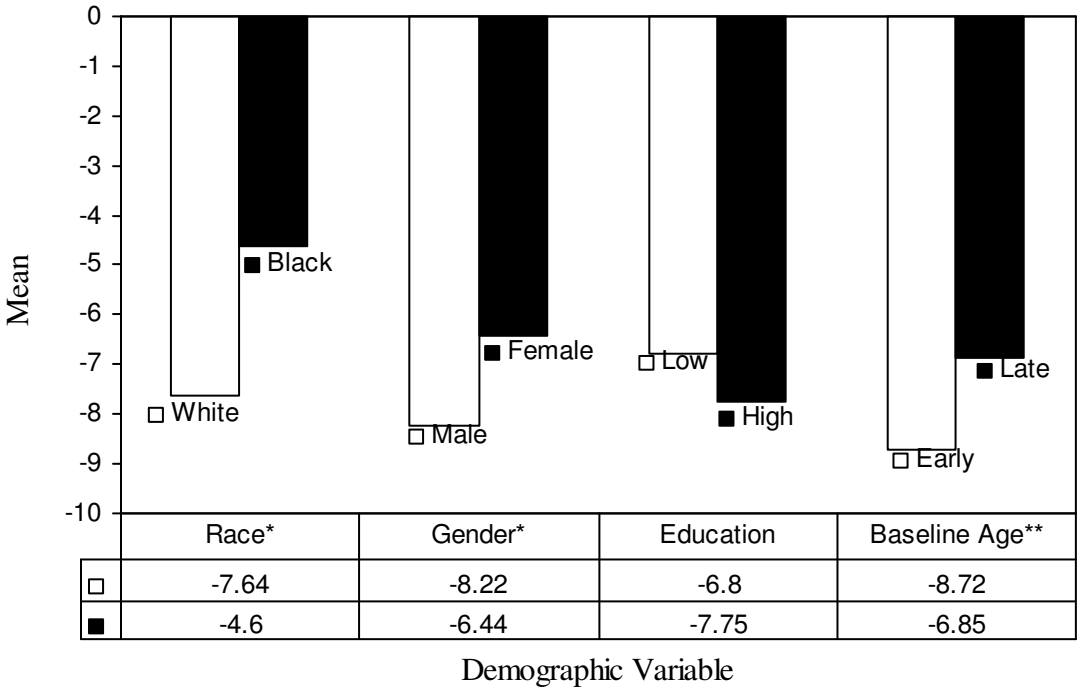

Figure 3. Mean annualized CERAD Total Change Score by demographic variables.

* Difference is significant at $p < .001$

** Difference is significant at $p = .007$

References

Adak, S., Illouz, K., Gorman, W., Tandon, R., Zimmerman, E., & Kaye, J. (2004). Predicting the rate of cognitive decline in aging and early Alzheimer disease. *Neurology, 63,* 108 – 114.

Anastasi, A. & Urbina, S. (1997). *Psychological Testing. (7th ed.)* Upper Saddle River: Simon & Schuster.

Anstey, K. & Hofer, S. (2004). Longitudinal designs, methods, and analysis in psychiatric research. *Australian and New Zealand Journal of Psychiatry, 38,* 93 - 104.

Atkinson, R. & Shiffrin, R. (1971). The control of short-term memory. *Scientific American, 221,* 82 – 90.

Barnes. L., Wilson, R., Li, Y., Aggarwal, N., Gilley, D., & McCann, J., et al. (2005). Racial differences in the progression of cognitive decline in Alzheimer's disease. *American Journal of Geriatric Psychiatry, 13,* 959 - 967.

Becker, J., Huff, F., Nebes, R., Holland, A., & Boller, F. (1988). Neuropsychological function in Alzheimer's disease. Pattern of impairment and rates of progression. *Archives of Neurology, 45,* 263 – 268.

Berg, L. (1988). Clinical Dementia Rating (CDR). *Psychopharmacology Bulletin, 24,* 637 – 639.

Blessed, G., Tomlinson, B., & Roth, M. (1968). The association between quantitative measures of dementia and of senile change in the cerebral gray matter of elderly subjects. *British Journal of Psychiatry, 114,* 797 – 811.

Bowler, J., Munroz, D., Merskey, H., & Hachinski, V. (1998). Factors affecting the age of onset and rate of progression of Alzheimer's disease. *Journal of Neurology and Neurosurgery Psychiatry, 65,* 184 – 190.

Brocco, L., Gellato, R., Grigoletto, F., & Lippi, A., (1994). Factors affecting course and survival in Alzheimer's disease: A 9-year longitudinal study. *Archives of Neurology, 51,* 1213 - 1219.

Brookmeyer, R., Currada, M., Curriero, F., & Kawas, C. (2002). Survival following a diagnosis of Alzheimer's disease. *Archives of Neurology, 59,* 1764 - 1767.

Buckholtz, N. (1997). Background and purpose of the Consortium to Establish a Registry for Alzheimer's Disease. *Neurology, 49,* S1.

Burke, W., Miller, J., Rubin, E., Morris, J., Coben, L, & Duchek, J., et al. (1988). Reliability of the Washington University Clinical Dementia Rating. *Archives of Neurology, 45,* 31 – 32.

Burns, A., Jacoby, R., & Levy, R. (1991). Progression of cognitive impairment in Alzheimer's disease. *Journal of American Geriatric Society, 21,* 363 – 370.

Bush, C., Kozak, J., & Elmslie, T. (1997). Screening for cognitive impairment in the elderly. *Canadian Family Physician, 43,* 1763 – 1768.

Chandler, M., Lacritz, L., Hynan, L., Allen, G., Deschner, M., Weiner, M., & Cullum, M. (2005). A total score for the CERAD neuropsychological battery. *Neurology, 66,* 102 - 106.

Chelune, G., Naugle, R., Luders, H., Sedlack, J., & Awad, I. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology, 7,* 41 – 52.

Clark, C., Sheppard, L., Fillenbaum, G., Galasko, D., Morris, J., & Koss, E., et al. (1999). Variability in annual Mini-Mental State Examination Score in patients with probable Alzheimer's disease. *Archives of Neurology, 56,* 857 - 862.

Collie, A., Shafiq-Antonacci, R., Maruff, P., Tyler, P., & Currie, J. (1999). Norms and the effects of demographic variables on a neuropsychological battery for use in healthy aging Australian populations. *Australian and New Zealand Journal of Psychiatry, 33,* 568 – 575.

Consortium to Establish a Registry for Alzheimer's Disease. (1996). Manual of operations: Consortium to establish a registry for Alzheimer's disease (CERAD). Unpublished CD-ROM Documentation. Durham, NC: Duke University Medical Center.

Corey-Bloom. (2004). Alzheimer's disease. *Continuum: Lifelong Learning in*

Cullum, M. & Rosenberg, R. (1998). Memory loss – when is it Alzheimer's disease? *The Journal of the American Medical Association, 279,* 1689 - 1690.

Delis, D., Kramer, J., Kaplan, E., & Ober, B. (1987). *California Verbal Learning Test: Adult version manual.* San Antonio: The Psychological Corporation.

Demadura, T., Delis, D., Jacobson, M., & Salmon, D. (2001). Do subgroups of patients with Alzheimer's disease exhibit asymmetric deficits on memory tests? *Journal of Clinical and Experimental Neuropsychology, 23,* 164 – 171.

Desai, A. & Grossberg, G. (2005). Diagnosis and treatment of Alzheimer's Disease. *Neurology, 64,* 34 - 39.

Division of Neurology at Duke University Medical Center (2003). CERAD An Overview: The Consortium to Establish a Registry for Alzheimer's Disease. http://cerad.mc.duke.edu/Default.htm [On-line.].

Doody, R., Massman, P., Dunn, J. (2001). A method for estimating progression rates in Alzheimer's disease. *Archives of Neurology, 58,* 449-454.

Doraiswamy, P., Bieber, F., Kaiser, L., Krishnan, K., Reuning-Scherer, J., et al. (1997). The Alzheimer's disease assessment scale: Patterns and predictors of baseline cognitive performance in multicenter Alzheimer's disease trials. *Neurology, 48,* 1511-1517.

Duke, L. & Kaszniak, A. (2000). Executive control functions in degenerative dementias: a comparative review. *Neuropsychology Review, 10,* 75 – 99.

Edland, S., Beekly, D., Barnhart, R., & van Belle, G. (1997). Design and implementation of a longitudinal multicenter database. *Neurology, 49,* S17 – S19.

Feldman, H. & Woodward, M. (2005). The staging and assessment of moderate to severe Alzheimer's disease. *Neurology, 65,* S10-S17.

Fhsani, M., Bedlich, C., McNicoll, L., & Fly, E. (2003). Underrecognition of preexisting cognitive impairment by physicians in older ICU patients. *CHEST Journal, 124,* 2267 – 2274.

Fillenbaum, G., Peterson, B., & Morris, J. (1996). Estimating the validity of the Clinical Dementia Rating Scale: The CERAD experience. *Aging (Milano), 8,* 379 - 385.

Fillenbaum, G., Peterson, B., Welsh-Bohmer, K., Kukull, W., & Heyman, A. (1998). Progression of Alzheimer's disease in black and white patients: The CERAD experience, Part XVI. *Neurology, 51,* 154 – 158.

Fillenbaum, G. & Woodbury, M. (1998). Typology of Alzheimer's disease: Findings from CERAD data. *Aging and Mental Health, 2,* 105 – 127.

Fisher, N., Rourke, B., & Bieliauskas, L. (1999). Neuropsychological subgroups of patients with Alzheimer's disease: An examination of the first 10 years

of CERAD data. *Journal of Clinical and Experimental Neuropsychology, 21,* 488 – 518.

Foldi, N., Majerovitz, S., Sheikh, K., & Rodriguez, E. (1999). The Test for Severe Impairment: Validity with the Dementia Rating Scale and utility as a longitudinal measure. *The Clinical Neuropsychologist, 13,* 22 – 29.

Folstein, M., Folstein, S., & McHugh, P. (1975). "Mini-Mental State:" A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189 – 198.

Galasko D., Edland, S., Morris, J., Clark, C., Mohs, R., & Koss, E. (1995). The Consortium to Establish a Registry for Alzheimer's disease (CERAD): Part XI. Clinical milestones in patients with Alzheimer's disease followed over 3 years. *Neurology, 45,* 1451 – 1455.

Guilford, J. P. (1956). Fundamental statistics in psychology and education. New York: McGraw Hill.

Helmer, C., Damon, D., Letenneur, L., Fabrigoule, C., & Barberger-Gateau, P., et al. (1999). Marital status and risk of Alzheimer's disease: A French population-based cohort study, *Neurology, 53,* 1953 - 1958.

Henderson, V. & Buckwalter, J. (1994). Cognitive deficits of men and women with Alzheimer's disease. *Neurology, 44,* 90 - 96.

Hermann, B., Seidenburg, M., Schoenfeld, J., Peterson, J., Leveroni, C., & Wyler, A. (1996). Empirical techniques for determining the reliability, magnitude,

and pattern of neuropsychological change after epilepsy surgery. *Epilepsia, 37,* 942 – 950.

Heyman, A. & Fillenbaum, G. (1997). Overview: clinical sites, case material, and special studies. *Neurology, 49,* S2 - S6.

Heyman, A., Peterson, B., Fillenbaum, G., & Pieper, C. (1997). Predictors of time to institutionalization of patients with Alzheimer's disease. The CERAD experience, part XVII. *Neurology, 48,* 1304 – 1309.

Hinkle, D., Wiersma, W., Jurs, S. (2003). In Applied Statistics for the Behavioral Sciences, *Correlation: a measure of relationship* (pp.108-109). Boston, NY: Houghton Mifflin.

Holmes, C. & Lovestone, S. (2003). Long-term cognitive and functional decline in late onset Alzheimer's disease: therapeutic implications. *Age and Ageing, 32,* 200 – 204.

Hughes, C., Berg, L., Danziger, W., Coben, L., & Martin, R. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry, 140,* 566 – 572.

Isaacs, B. & Kennie, A. (1973). The Set test as an aid to the detection of dementia in old people. *British Journal of Psychiatry, 123,* 467 – 470.

Jacobs, D., Sano, M., & Marder, K. (1994). Age of onset of Alzheimer's disease: Relation to pattern of cognitive dysfunction and rate of decline. *Neurology, 44,* 1215 – 1220.

Jacobs, D., Sano, M., Marder, K., Bell, K., & Stern, Y. (1995). Neuropsychological detection and characterization of preclinical Alzheimer's disease. *Neurology, 45,* 957 - 962.

Jacobson, N. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12 – 19.

Johnson, S., Hughes, S., Bullock, R., & Hindmarch, I. (2003). Prediction of the rate of decline in cognitive function in Alzheimer's disease: A model based on simple demographic data and widely used rating scales. *Dementia and Geriatric Cognitive Disorders, 16,* 276 – 282.

Juva, K., Makela, M., & Erkinjuntti, T. (1997). Functional assessment scales in detecting dementia. *Age and Aging, 26*, 393 – 400.

Kaltreider, L., Cicerello, A., Lacritz, L., Honig, L., Rosenberg, R., & Cullum, M. (2000). Comparison of the CERAD and CVLT list-learning tasks in Alzheimer's disease. *The Clinical Neuropsychologist, 14,* 269 - 274.

Kaplan, E., Goodglass, H., & Weintraub, S. (1978). The Boston Naming Test. Boston, MA: Veterans Administration Medical Center.

Karrasch, M. & Laine, M. (2003). Age, education, and test performance on the Finnish CERAD. *Acta Neurologica Scandanavia, 108,* 97 – 101.

Katzman R., Brown T., & Fuld P. (1983). Validation of a short Orientation-Memory-Concentration Test of cognitive impairment. *American Journal of Psychiatry, 140,* 734 – 739.

Kokmen, E., Ozsarfati, Y., Beard, C., O'Brien, P., & Rocca, W. (1996). The impact of referral bias on clinical and epidemiological studies of Alzheimer's disease. *Journal of Clinical Epidemiology, 49,* 79 – 83.

Koss, E., Edland, S., Fillenbaum, G., Mohs, R., Clark, C., Galasko, D., et al. (1996). Clinical and neuropsychological differences between patients with earlier and later onset of Alzheimer's disease: A CERAD analysis, part XII. *Neurology, 46,* 136 – 141.

Lapane, K., Gambassi, G., Landi, F., Sgadari, A., Mor, V., & Bernabei. (2001). Gender differences in predictors of mortality in nursing home residents with AD. *Neurology, 56,* 650 - 654.

Larson, E., Shadlen, M., Wang, L., McCormick, W., Bowen, J., et al. (2004). Survival after initial diagnosis of Alzheimer's disease. *Annals of Internal Medicine, 140,* 501 – 509.

Locascio, J., Growdon, J., & Corkin, S. (1995). Cognitive test performance in detecting, staging, and tracking Alzheimer's disease. *Archives of Neurology, 52,* 1087 - 1099.

Lopez, O. Wisniewski, S., Hamilton, R., Becker, J., Kaufer, D., et al. (2000). Predictors of progression in patients with AD and Lewy bodies. *Neurology, 54,* 1774 - 1779.

Mattis, S., (1988). *Dementia Rating Scale: Professional manual.* Odessa, FL: Psychological Assessment Resources.

McKhann, G., Drachmann, D., Folstein, M., Katzmann, R., Price, D., & Stadlan, E. (1984). 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, 34,* 939 – 944.

Morris J. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology, 43,* 2412 – 2414.

Morris, J., Edland, S., Clark, C., Galasko, D., Koss, E. Mohs, R., et al. (1993). The consortium to establish a registry for Alzheimer's disease (CERAD). Part IV. Rates of cognitive change in the longitudinal assessment of probable Alzheimer's disease. *Neurology, 43,* 2457 – 2465.

Morris, J., Heyman, A., Mohs, R., Hughes, J., van Belle, G., Fillenbaum, G., et al. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology, 39,* 1159 – 1165.

Neumann, P., Araki, S., Arcelus, A., Longo, A., Papadopoulos, G., Kosik, K., et al. (2001). Measuring Alzheimer's disease progression with transition probabilities: Estimates from CERAD. *Neurology, 57,* 957 - 964.

Olin, J. & Zelinski, E. (1991). The 12-month stability of the Mini Mental State Examination. *Psychological Assesssment, 3,* 427 – 432.

Pisani, M., Inouye, S., McNicoll, L., & Redlich, C. (2003). Screening for pre-existing cognitive impairment in older intensive care unit patients: Use of proxy assessment. *Journal of the American Geriatric Society, 51,* 689 - 693.

Relkin, N. (2000). Screening and early detection of dementia. *American Journal of Managed Care, 6,* 1111 – 1124.

Reitan, R. & Wolfson, D. (1993). *The Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation*. (2nd ed.) Tucson, AZ: Neuropsychology Press.

Ripich, D., Carpenter, B., & Ziol, E., (1997). Comparison of African-American and white persons with Alzheimer's disease on language measures. *Neurology, 48,* 781 - 783.

Rojas, D. & Bennett, T. (1995). Single versus composite score discriminative validity with the Halstead-Reitan Battery and the Stroop Test in mild brain injury. *Archives of Clinical Neuropyschology, 10,* 101 – 110.

Rosen, W. Mohs, R., & Davis, K. (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry, 141,* 1356-1364.

Salmon, D., Thal, L., Butters, N., & Heindel, W. (1990). Longitudinal evaluation of dementia of the Alzheimer type: A comparison of 3 standardized mental status examinations. *Neurology, 40,* 1225 - 1230.

Satzger, W., Hampel, H., Padberg, F., Burger, K., Nolde, T., Ingrassia, G. et al. (2001). Practical application of the CERAD test battery as a neuropsychological dementia screening test. *Nervenarzt, 72,* 196 – 203.

Sclan, S. & Kanowski, S. (2001). Alzheimer's disease: Stage-related interventions. *Lippincott's Case Management, 6,* 48 - 63.

Spreen, O. & Strauss, E. (1998). Mini-Mental State Examination (MMSE). In *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd ed., pp. 65-74). New York: Oxford University Press.

Stewart, R., Richards, M., Carol B., & Mann, A. (2001). Cognitive function in UK community-dwelling African Caribbean elders: normative data for a test battery. *International Journal of Geriatric Psychiatry, 16,* 518 – 527.

Stern, Y., Steven, A., Tang, M., & Tsai, W. (1999). Rate of memory decline in AD is related to education and occupation: Cognitive reserve? *Neurology, 53,* 1942 - 1947.

Stern, Y., Tang, M., Denaro, J., & Mayeaux, R. (1995). Increased risk of mortality in Alzheimer's disease patients with more advanced educational and occupational attainment. *Annals of Neurology, 37,* 590 - 595.

Storandt, M., Grant, E., Miller, P., & Morris, J. (2002). Rates of progression in mild cognitive impairment and early Alzheimer's disease. *Neurology, 59,* 1034 - 1041.

Tombaugh, T. & McIntyre, N. (1992). The Mini-Mental State Examination: A comprehensive review. *JAGS, 40,* 922 – 935.

Tractenberg, R., Schafer, K., Morris, J. (2001). Interobserver disagreements in Clinical Dementia Rating assessment: Interpretation and implications for training. *Alzheimer's Disease and Associated Disorders, 15,* 155-161.

Van Belle, G., Uhlmann, R., & Hughes, J. (1990). Reliability of estimates of changes in mental status test performance in senile dementia of the Alzheimer type. *Journal of Clinical Epidemiology, 43,* 589 – 595.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale: Administration and Scoring Manual.* (3rd ed.) San Antonio: The Psychological Corporation.

Welsh, K., Butters, N., Hughes, J., Mohs, R., & Heyman, A. (1991). Detection of abnormal memory decline in mild cases of Alzheimer's disease using CERAD neuropsychological measures. *Archives of Neurology, 48,* 278 - 281.

Welsh, K., Butters, N., Hughes, J., Mohs, R., & Heyman, A. (1992). Detection and staging of dementia in Alzheimer's disease. Use of the neuropsychological measures developed for the Consortium to Establish a Registry for Alzheimer's Disease. *Archives of Neurology, 49,* 448 - 452.

Welsh, K., Butters, N., Mohs, R., Beekly, D., Edland, S., Fillenbaum, G. et al. (1994). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology, 44,* 609 – 614.

Welsh-Bohmer, K. & Mohs, R. (1997). Neuropsychological assessment of Alzheimer's disease. *Neurology, 49,* S11 - S13.

Whyte, S., Cullum, M., Hynan, L., Lacritz, L., Rosenberg, R., & Weiner, M. (2005). Performance of elderly Native Americans and Caucasians on the CERAD neuropsychological battery. *Alzheimer Disease and Associated Disorders, 19,* 74 - 78.

Wiederholt, W., Cahn, D., Butters, N., Salmon, D., Kritz-Silverstein, D., & Barrett-Connor, E. (1993). Effects of age, gender, and education on selected neuropsychological tests in an elderly community cohort. *Journal of the American Geriatric Society, 41,* 639 – 647.

Woodbury, M. & Fillenbaum, G. (1996). Psychometric characteristics of the Mini-Mental State Examination in patients with Alzheimer's disease - A

grade of membership analysis of CERAD data: Part II. *International Journal of Geriatric Psychiatry, 11,* 543 – 553.

Vita

Heidi Christine Rossetti was born in Minnesota on April 18, 1982, the daughter of Mr. John Richardson and Mrs. Judy Richardson. She and her family moved to Vienna, Austria in 1989 where she attended Vienna Christian School until 1995 when they moved to Texas. After completing high school in Allen, Texas in 2000, she entered the University of Texas at Dallas, where she graduated magna cum laude and received her Bachelor of Arts degree in the field of Psychology in December 2003. In August 2004, she entered graduate school at the University of Texas Southwestern Medical Center School at Dallas (UTSW) in pursuit of her Master of Science degree in Rehabilitation Counseling Psychology. In August 2006, she entered the Clinical Psychology program at UTSW and is expected to graduate with her doctoral degree in 2010.

Permanent Address: 3924 Windford Drive
Plano, TX 75025