Certified Normal: Alzheimer's Disease Biomarkers and Normative Estimates of Cognitive Functioning

Jason Hassenstab, PhD[1,2,9], **Rachel Chasse, BA**[1,2], **Perri Grabow, MA**[1,11], **Tammie L.S. Benzinger, MD, PhD**[1,3,10], **Anne M. Fagan, PhD**[1,2,4], **Chengjie Xiong, PhD**[1,5], **Mateusz Jasielec, MS**[5], **Elizabeth Grant, PhD**[1,5], and **John C. Morris, MD**[1,2,6,7,8]

[1]Charles F. and Joanne Knight Alzheimer's Disease Research Center, Washington University School of Medicine, St. Louis, MO. USA

[2]Department of Neurology, Washington University School of Medicine, St. Louis, MO. USA

[3]Department of Radiology, Washington University School of Medicine, St. Louis, MO. USA

[4]Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, MO. USA

[5]Department of Biostatistics, Washington University School of Medicine, St. Louis, MO. USA

[6]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO. USA

[7]Department of Physical Therapy, Washington University School of Medicine, St. Louis, MO. USA

[8]Department of Occupational Therapy, Washington University School of Medicine, St. Louis, MO. USA

[9]Department of Psychology, Washington University in St. Louis, St. Louis, MO. USA

[10]Department of Neurological Surgery, University of Missouri-St. Louis, St. Louis, MO, USA

[11]Department of Psychology, University of Missouri-St. Louis, St. Louis, MO, USA

Abstract

Normative samples drawn from older populations may unintentionally include individuals with preclinical Alzheimer's disease (AD) pathology, resulting in reduced means, increased variability, and overestimation of age-effects on cognitive performance. 264 cognitively normal (CDR=0) older adults were classified as biomarker-negative ("Robust Normal," n=177) or biomarker-positive ("Preclinical Alzheimer's Disease" (PCAD), n=87) based on amyloid imaging, cerebrospinal fluid biomarkers, and hippocampal volumes. PCAD participants performed worse than Robust Normals on nearly all cognitive measures. Removing PCAD participants from the normative sample yielded higher means and less variability on episodic memory, visuospatial

CORRESPONDENCE: Jason Hassenstab, PhD, 4488 Forest Park Ave., Suite 130, St. Louis, MO 63108. Office: 314-747-4032, Fax: 314-286-2443, ; Email: hassenstabj@wustl.edu

Author Manuscript

ability, and executive functioning measures. These results were more pronounced in participants aged 75 and older. Notably, removing PCAD participants from the sample significantly reduced age effects across all cognitive domains. Applying norms from the Robust Normal sample to a separate cohort did not improve CDR classification when using standard deviation cutoff scores. Overall, removing individuals with biomarker evidence of preclinical AD improves normative sample quality and substantially reduces age-effects on cognitive performance, but provides no substantive benefit for diagnostic classifications.

Keywords

Alzheimer's disease; biomarkers; cognition; normative data; memory; preclinical disease

1. INTRODUCTION

As the era of Alzheimer's disease (AD) secondary prevention trials begins, there is increased interest in detecting the earliest cognitive changes in the course of the disease. A decline in cognitive functioning is unquestionably the most face valid indicator of AD progression, but accurately capturing the earliest declines is dependent upon several factors, including the psychometric characteristics of the cognitive tests used and the quality of the normative sample used as a reference group. Cognitive measures must have adequate reliability and validity and should be sensitive and specific to healthy versus diseased states. At an individual level, capturing changes in cognitive performance is best accomplished by comparing current test performance with prior assessments. When this data is not available, a useful approach is to compare test performance to that of healthy persons of a similar demographic profile. Detecting cognitive decline in this situation is highly dependent upon the normative sample used as a basis for comparison. Thus, reference groups used to provide normative values should be composed of individuals who are comparable to the individuals being tested in terms of ages, education levels, and premorbid functioning. Participants in normative samples should also be carefully screened for health conditions that may impact cognitive performance. Yet, despite efforts to produce healthy normative samples, a substantial portion of included individuals may have subtle declines in cognitive performance due to undetected underlying disease (Sliwinski et al., 1996; Storandt and Morris, 2010). Most neurodegenerative diseases have a "preclinical" phase in which the disease process is underway in the brain, but the degree of pathology is not sufficient to produce overt clinical symptoms (Foltynie, 2003; Price and Morris, 1999). In older populations, as much as 30–40% of individuals over the age of 65 may be in a preclinical stage of AD (Price et al., 2009).

The pathological hallmarks of AD can be quantified *in vivo* using cerebrospinal fluid (CSF) assays and neuroimaging methods such as positron emission tomography (PET) and structural and functional magnetic resonance imaging (MRI). CSF biomarkers including amyloid-β_{1-42} ($A\beta_{1-42}$), total tau (t-tau) and phosphorylated tau$_{181}$ (p-tau$_{181}$), and neuroimaging markers including [^{11}C] Pittsburgh compound B (PIB) PET and hippocampal volume, have been used to describe the course of the preclinical stages of AD (Bateman et al., 2012; Knopman et al., 2012; Vos et al., 2013). Results from these studies have

consistently shown that AD pathology may begin well before overt symptoms of dementia are present. β-amyloidosis begins at least 10–20 years prior to clinical diagnosis, followed by tau aggregation into neurofibrillary tangles resulting in neuronal injury (Bateman et al., 2012; Jack et al., 2013; Sperling et al., 2011; Villemagne et al., 2013). The final proposed preclinical stage of AD is characterized by "subtle cognitive decline", however, this has yet to be fully operationalized (Sperling et al., 2011). Studies that have modeled cognitive trajectories in the preclinical stage of AD have shown cognitive declines beginning within approximately 7–10 years of clinical diagnosis (Grober et al., 2008; Roe et al., 2013; Saxton et al., 2004), with a pronounced acceleration 3–5 years prior to diagnosis (Howieson et al., 2008; Johnson et al., 2009). Because the preclinical stage of AD is defined by the absence of clinically significant cognitive and functional impairment, conventional normative samples do not exclude individuals with preclinical AD, and therefore may be less sensitive to detecting subtle impairments in cognitive functioning (Sliwinski et al., 1996; Storandt and Morris, 2010).

Sliwinski and colleagues (Sliwinski et al., 1996) demonstrated that failing to screen for preclinical AD cases has three primary influences on normative data. First, individuals with preclinical disease tend to have slightly worse performance, resulting in normative data with reduced mean scores. Second, variability in cognitive performance is more pronounced in preclinical AD, resulting in normative data with larger standard deviations and skewed distributions. Finally, since there is a strong correlation between dementia risk and age, the influence of age on cognitive performance would be magnified, especially in older age ranges where the prevalence of preclinical AD is much higher. Several investigators have since attempted to improve upon conventional normative data by using methods to identify individuals in the preclinical stage of AD and exclude them from normative samples. The majority of existing studies have applied a "longitudinal" exclusion criterion wherein individuals that have been followed for several years and progressed to dementia are selectively removed from normative samples, ostensibly resulting in a more robust normative dataset. The longitudinal method has consistently produced normative datasets with higher mean scores, reduced variability, and less influence of age on cognitive test performance (De Santi et al., 2008; Pedraza et al., 2010; Ritchie et al., 2007; Sliwinski et al., 1996; Storandt and Morris, 2010). However, the clinical utility of the longitudinal method has been unclear. Some studies have reported improved diagnostic classification accuracy using robust normative data (De Santi et al., 2008; Holtzer et al., 2008), whereas others have not (Ritchie et al., 2007; Storandt and Morris, 2010). One possible reason why these studies have demonstrated mixed results could be that the longitudinal method relies solely on clinical diagnosis of dementia, and therefore excludes only persons in the later preclinical stages or those with more a more rapid course of progression. Thus, there is a risk that a large percentage of individuals who did not progress to a clinically symptomatic stage of dementia during the study follow-up period may indeed have AD pathology that is affecting their cognitive performance.

The purpose of this study was to determine if the identification and subsequent removal of individuals who are cognitively normal, but exhibit biomarker evidence of preclinical AD pathology, would improve normative cognitive data and show better correspondence with the Clinical Dementia Rating (CDR) (Morris, 1993). To this end, we selected cognitively normal

participants from ongoing studies of normal aging and dementia who had completed biomarker studies. Participants were classified according to our previously published cutoff values for AD biomarkers as belonging to a biomarker-positive group with Preclinical Alzheimer's Disease pathology (PCAD) or a biomarker-negative group of "Robust Normals". We hypothesized that removing participants with PCAD would increase mean scores, decrease variability, normalize distributions, and reduce age effects compared to a conventional normative sample that included participants with PCAD. We also hypothesized that application of the Robust Norms to a separate cohort of longitudinally followed participants would improve correspondence with CDR classification accuracy when standard deviation cutoff scores were used.

2. METHODS

2.1 Participants

Participants were older adult volunteers enrolled in ongoing studies of aging and dementia at the Knight Alzheimer's Disease Research Center (KADRC) at the Washington University School of Medicine. Inclusion/Exclusion criteria and assessment methodology have been detailed in previous publications (Berg et al., 1998; Coats and Morris, 2005). KADRC participants were living independently in the community at study entry and underwent annual clinical assessment unless prevented by death, illness, refusal, or relocation from the greater St. Louis area. Each participant and their collateral source, a close friend or family member, were interviewed with standard instruments with respect to cognitive and functional abilities (Morris et al., 2006), and that information was used by experienced physicians and nurse-clinicians to determine the Clinical Dementia Rating (CDR (Morris, 1993). The Human Research Protection Office at Washington University School of Medicine approved the KADRC studies, including the Healthy Aging and Senile Dementia Study (P01AG003991), the Alzheimer's Disease Research Center study (P50AG05681), and the Antecedent Biomarkers for AD: the Adult Children Study (P01AG026276). Written informed consent was obtained from all participants at enrollment.

2.1.1 Normative Samples—Participants for the normative samples (Figure 1) were selected from the larger KADRC cohort based on the following criteria: Completion of baseline cognitive assessment; Clinical Dementia Rating (CDR) score of 0 at all available visits; at least one annual follow-up assessment; at least 65 years of age at baseline; completion of lumbar puncture (for CSF biomarkers) within 18 months of baseline or completion of PIB-PET within 18 months of baseline; and structural MRI within 18 months of baseline.

2.1.2 Longitudinal Sample—Participants for the Longitudinal sample (Figure 1) were selected from the larger KADRC cohort based on the following criteria: CDR 0 at study entry and completed at least two follow-up clinical and cognitive assessments; at least 65 years of age at baseline. Participants in the longitudinal sample were further classified as "Stable" if they never progressed to a CDR >0 during the study follow-up period or as a "Progressor" if they converted to a CDR >0 with a diagnosis of symptomatic AD and

remained at that level for at least two consecutive annual assessments and through their last available follow-up visit.

2.2 Procedures

All participants underwent annual clinical evaluation and cognitive assessment, which included CDR and a comprehensive neuropsychological battery. The CDR is a global dementia staging tool that assesses the presence or absence of dementia and, when present, its severity. The global CDR stages are 0, indicating cognitive normality, and 0.5, 1, 2, and 3, indicating very mild dementia, and mild, moderate, and severe dementia, respectively (Berg et al., 1992). CDR score and diagnosis were assigned at each visit by trained clinicians. In addition, participants complete blood tests to determine apolipoprotein (APOE) genotyping, as described in earlier reports (Pizzie et al., 2013). Participants were classified as ε4+ (44,34,24) or ε4− (33,23,22).

2.2.1 Cerebrospinal Fluid (CSF) Biomarkers—CSF sample collection has been described in detail (Fagan et al., 2007; Vos et al., 2013). Briefly, CSF samples (20–25 mL) were collected at study entry by experienced neurologists after an overnight fast. Samples were analyzed for total tau (t-tau), phosphorylated tau (p-tau$_{181}$), and amyloid-β_{1-42} (Aβ_{1-42}) by ELISA (INNOTEST; Fujirebio (formerly Innogenetics), Ghent, Belgium). CSF biomarkers were dichotomized (normal or abnormal) using cutoffs derived in our previous studies (Vos et al., 2013). Cutoffs for each biomarker were determined by a receiver operating characteristic curve that differentiated participants who were CDR 0 at baseline from those who were CDR 0.5 with a diagnosis of symptomatic AD. Cutoffs for abnormal values were less than 459 pg/mL for Aβ_{1-42}, greater than 339 pg/mL for t-tau, and greater than 67 pg/mL for p-tau$_{181}$.

2.2.2 Neuroimaging Biomarkers—Detailed information on the PET PIB imaging and analysis procedures have been reported (Mintun et al., 2006). In brief, approximately 12mCi of [^{11}C] PIB was administered intravenously during a 60-minute dynamic PET scan in 3-dimensional mode (septa retracted; 24×5-second frames; 9×20-second frames; 10×1-minute frame). PIB image analysis was performed for specific regions of interest (ROIs) as detailed previously (Mintun et al., 2006). The ROIs were applied to the PET data, yielding high-resolution regional time-activity curves. Time-activity curves were analyzed for PIB-specific binding by Logan graphical analysis using the cerebellum ROI data as a reference region. The Logan analysis yields a tracer distribution volume ratio (DVratio), which was then converted to an estimate of the binding potential (BP) for each ROI: BP X DVratio X 1.61 The BP expresses the regional PIB binding values in a manner directly proportional to the number of binding sites. The BP values from the prefrontal cortex, gyrus rectus, lateral temporal, and precuneus ROIs were averaged in each subject to calculate a mean cortical binding potential (MCBP), as these regions have been shown to have high PIB uptake among participants with AD. PIB-PET MCBP values above 0.18 were considered indicative of significant amyloid deposition consistent with underlying AD pathology. This cutoff was chosen based on prior work in this population (Mintun et al., 2006), where we found that nearly all individuals with a clinical diagnosis of AD had MCBP values greater than 0.18, and all young cognitively normal individuals had MCBP values less than 0.18.

In addition to PET imaging, all participants underwent anatomic MRI. Data were acquired on a Siemens TIM Trio 3T scanner. T_1-weighted images were acquired using a MPRAGE sequence with: repetition time (TR)=2400 ms, echo time (TE) = 3.16 ms, flip angle (FA) = 8°, field of view (FOV) = 256 mm, in plane resolution 176 × 256, slice thickness = 1mm acquired in sagittal orientation. Images had a 1 mm isotropic resolution. Regional volumes were obtained using the FreeSurfer imaging analysis suite, version 5.1 (Desikan et al., 2006; Fischl et al., 2002). During processing, each voxel is assigned a neuroanatomical label based on probabilistic information derived from a manually labeled training set, which included healthy young and older adults. Previous work indicates that this technique generates volumes with a high correspondence to manually generated volumes (Desikan et al., 2006; Fischl et al., 2002). We had no *a priori* hypotheses about laterality of hippocampal volumes, so volumes of left and right hippocampus were summed. An analysis of covariance approach (Buckner et al., 2004; Jack et al., 1989) was used to adjust hippocampal volumes for total intracranial volume (ICV): Adjusted hippocampal volume = raw volume – (b^* [ICV – mean ICV]), where b is the slope of the regression of the hippocampal volume on ICV. Participants with an adjusted hippocampal volume >/= 1.5 SDs below the group mean were considered to have significant neurodegeneration consistent with preclinical AD.

2.2.3 Neuropsychological Assessment—All participants completed a battery of standard neuropsychological measures within approximately 2 weeks of their clinical evaluation. KADRC's cognitive batteries have been described in detail in prior publications (Johnson et al., 2008; Pizzie et al., 2014). The core of the battery consists of measures drawn from the Uniform Data Set (UDS), which are collected at all Alzheimer's Disease Centers across the United States (Johnson et al., 2008; Morris et al., 2006; Weintraub et al., 2009). Tests included in the current study consisted of UDS measures as well as other well-known measures of neuropsychological function. We list the individual measures here by theoretical domain with the understanding that certain measures may be represented by more than one domain. Episodic memory measures included the Logical Memory Immediate and Delayed recall from the Wechsler Memory Scale-Revised (WMS-R; (Wechsler, 1987)), WMS Associate Learning, Free and Cued Selective Reminding Test (FCSRT; (Grober et al., 1988)), and the Benton Visual Retention Test (Benton, 1963). Executive functions and visuospatial measures include the Trailmaking Test Parts A & B (Armitage, 1946), Digit Span forwards and backwards from WMS-R, the Digit Symbol Coding subtest from the Wechsler Adult Intelligence Scale-Revised (WAIS-R; (Wechsler, 1981)), Mental Control from the WMS, and the Block Design subtest from WAIS. Language and semantic knowledge measures included Category Fluency for Animals and Vegetables (Goodglass and Kaplan, 1983), Word Fluency for letters /s/ and /p/ (T. G. Thurstone and L. L. Thurstone, 1948)), the 30 odd items from the Boston Naming Test (Goodglass and Kaplan, 1983; Weintraub et al., 2009), and the Information subtest from the WAIS.

2.2.4 Normative Sample Classification—Participants were stratified into two groups based on fluid and neuroimaging biomarker evidence of AD pathology. Participants with biomarker levels inconsistent with AD pathology were classified as Robust Normals, while participants with biomarkers levels exceeding empirically defined thresholds were classified as having Preclinical Alzheimer's Disease pathology (PCAD). These groups were further

stratified by age into 65–74 years and 75 years and older based on the distribution of ages in the sample. The PCAD classification was based on proposed criteria used to stage preclinical AD (Sperling et al., 2011) that have subsequently been empirically validated in autosomal dominant AD and in sporadic AD (Bateman et al., 2012; Knopman et al., 2012; Vos et al., 2013). Cutoffs for CSF biomarkers (Vos et al., 2013), PIB PET (Mintun et al., 2006), and hippocampal volume were used in the following classification algorithm (also see Table 1): For a PCAD classification, participants must have evidence of either β-amyloidosis alone, or β-amyloidosis and neurodegeneration. Evidence of β-amyloidosis was defined as CSF $A\beta_{1-42}$ values <459 pg/mL and/or PIB PET MCBP > 0.18. Evidence of neuronal injury was defined as CSF t-tau > 339 pg/mL, and/or $p\text{-tau}_{181}$ >67 pg/mL, and/or adjusted hippocampal volume >−1.5 SDs below the group mean. As the focus of this manuscript was on the impact of preclinical Alzheimer's disease on normative cognitive data, individuals with evidence of neurodegeneration alone, often referred to as suspected non-AD pathology (SNAP; (Jack et al., 2012; Vos et al., 2013), were not included in the PCAD group.

2.2.5 Statistical Analyses—All analyses and graphics were completed using the R statistical computing environment (www.r-project.org). Group differences in continuous demographic variables were examined with Student's t-tests. Categorical demographic variables were compared using the Chi-squared statistic. Comparison of individual test scores between the combined, PCAD, and robust normal groups were conducted in each age group (65–74 and 75 and older) using linear regression models with terms for gender and education. Levene's test for homogeneity of variance was used to compare variances for each measure. Age effects were examined in all participants in the normative samples with linear regression models. Diagnostic sensitivity and specificity in the longitudinal test sample were calculated with the area under the curve (AUC) from Receiver Operating Characteristic (ROC) analyses. Classification accuracy (true positives + true negatives) of cutoff points based on z-scores of −1.0 and −1.5 was examined for each measure individually. Comparisons of classification accuracy between data standardized with Robust normal and Combined samples used the McNemar test for paired binomial proportions (McNemar, 1947). The McNemar test criterion uses the following equation (De Santi et al., 2008):

$$Z_0 = \frac{C-R}{\sqrt{C+R}}$$

Where C = number of cases classified as positive by Combined norms but negatively by Robust Norms and R = number of cases classified as positive using Robust norms but negative using Combined norms. Significance levels are obtained by comparing the square of Z_0 to a Chi-squared distribution with df = 1 (Agresti, 2002; De Santi et al., 2008).

3. Results

3.1 Demographics

3.1.1 Normative Groups—Participants in the normative groups were mostly female, Caucasian, ranged in age from 65–86 years of age, and most had completed a college degree

(see Table 2). As expected, the PCAD group was slightly older than the Robust Normals group ($t = 1.94$, p = 0.054) and had a higher prevalence of APOE ε4+ ($\chi^2 = 22.15$, p < .001). No group differences were observed for years of education, MMSE, or distribution of gender and race. By definition, biomarkers were significantly different between Robust Normals and PCAD groups.

3.1.2 Longitudinal Sample—Participants in the Longitudinal sample were 65–87 years old at baseline, most had completed a college degree, and were mostly Caucasian and female. Participants completed at an average of 9.5 ± 5.5 visits over an average of 13.6 ± 7.1 follow-up years. The Stable and Progressor groups were similar in age at baseline, education, MMSE score at baseline, and the distribution of gender. There were nonsignificant trends for higher prevalence of *APOE* ε4+ ($\chi^2 = 2.76$, p = 0.099) and a higher percentage of Caucasians in the Progressors ($\chi^2 = 3.63$, p = 0.057).

3.2 Cognitive Performance in the Normative Samples

Means and standard deviations of the Robust Normal, PCAD, and Combined groups are provided in Tables 3–4. Mean scores for Robust Normal and Combined groups are similar to previous reports describing the UDS neuropsychological measures (Hayden et al., 2011; Weintraub et al., 2009). According to Sliwinski and colleagues (1996), removing individuals with preclinical AD from normative samples should 1) increase mean cognitive scores; 2) decrease variability; and 3) reduce the effects of age on cognitive performance. Below, we present comparisons of mean scores and variability estimates separately for each age group. We then present the effects of age on cognitive performance across all participants in the normative samples.

3.2.1 Ages 65–74—Out of a total of 180 participants between the ages of 65.0 and 74.9 years, 127 (70.5%) were classified as Robust Normal and 53 (29.5%) met criteria for PCAD (Table 3). Regression models, including terms for gender and education, revealed similar performance across all groups in this age range. No significant differences emerged on episodic memory or language/semantic measures. Performance was also similar on most tests of executive/visuospatial functioning, with the exception of WAIS Block Design, where the PCAD group performed significantly worse than Robust Normals. Variability estimates were compared with Levene's test for homogeneity of variances. There were no significant differences in variability between PCAD and Robust Normals on any cognitive measure. However, compared to Robust Normals, there were trends for increased variability in the PCAD group on Logical Memory Delayed ($F_{(1,130)} = 3.13$, p = 0.079) and FCSRT free recall ($F_{(1,155)} = 3.74$, p = 0.055). There were no significant differences in variability between Robust Normals and the Combined Sample.

3.2.2 Ages 75 and older—Out of a total of 84 participants aged 75.0 years or older, 50 (59.5%) were classified as Robust Normal and 34 (40.5%) met criteria for PCAD (Table 4). Participants in the PCAD group performed worse than Robust Normals on nearly all measures of episodic memory including the immediate and delayed recall conditions of the WMS-R Logical Memory, the total recall condition of the FCSRT, and the delayed recall condition of the BVRT. In addition, worse performance was seen on several measures of

executive/visuospatial functions including the Trailmaking Test Parts A & B, WMS-R Digit Symbol Coding, and WAIS Block Design. PCAD participants also had worse performance than Robust Normals on category fluency and the Boston Naming Test. Compared to the Combined sample, the Robust Normals performed significantly better on only the BVRT Delayed, and had non-significant trends for better performance on Logical Memory Immediate and Delayed. Results from Levene's test revealed higher variability estimates in the PCAD group than the Robust Normals on the FCSRT total recall ($F_{(1,111)}$ = 8.47, p = 0.004) and BVRT delayed recall ($F_{(1,82)}$ = 4.22, p = 0.043). Removing PCAD from the sample resulted in a trend towards reduced variability in the Robust Normals compared to the Combined sample ($F_{(1,102)}$ = 3.16, p = 0.078) on WAIS Block Design.

3.2.3—Age effects were examined by regressing age on cognitive performance (Table 5 and Figure 2). Almost all measures of cognitive function exhibited profound age effects in the PCAD group, with the most consistent effects observed on episodic memory measures. Variance attributable to age ranged from 12.5% on FCSRT total recall to 29.8% on the Benton Visual Retention Test delayed recall. Age effects in the PCAD group on Executive/ Visuospatial measures and Language/Semantic measures were less consistent, but a pattern emerged where age effects were most pronounced on speeded tests. Variance estimates ranged from 5.2% on Trails A up to 25.6% on the WMS-R Digit Symbol Coding test. As expected, removing the PCAD individuals from the Combined sample attenuated age effects on nearly all individual tests.

3.2.4 Progression Rates in the Normative Sample—We examined the progression rates to CDR >0 in the normative sample using follow-up data when available. Over an average of 5.3 years of follow-up, 16/177 (9%) of the Robust Normal group progressed to a CDR >0 and 23/87 (26.4%) of the PCAD group progressed to a CDR >0, suggesting that the biomarker classification of PCAD is strongly related to clinical progression. Similar results were found in our prior publications in this cohort (Roe et al., 2013; Vos et al., 2013).

3.3 Application of Robust Normative Data to a Longitudinal Sample

Normative data from the Robust Normals and Combined groups were applied to the Longitudinal sample to determine if removing PCAD individuals would result in improved CDR diagnostic sensitivity and specificity. Norms were applied at the time of clinical diagnosis (as indicated by a CDR > 0 with a diagnosis of AD) for Progressors and at most recent assessment for Stable participants. Z-score cutoffs were set at −1.0 and −1.5, in keeping with common neuropsychological methods for defining cognitive impairment (Holtzer et al., 2008; Jak et al., 2009; Petersen et al., 1999; Storandt and Morris, 2010; Winblad et al., 2004). At the time of CDR diagnosis (for Progressors) and most recent assessment (for Stable participants), there were no significant differences in age or other demographic characteristics (Table 2). The areas under the curve (AUCs; Table 6) from the receiver operating characteristic analyses ranged from poor (BVRT Delayed AUC = 0.63) to good (FCSRT Total Recall AUC = 0.89–0.90) for episodic memory measures, regardless of whether combined or robust norms were used. AUCs for other measures were poor overall and were not substantially improved using Robust Normal data. Cutoff scores of −1.0 and −1.5 standard deviations below the mean produced unacceptable detection (sensitivity) rates

on all measures (Table 7). On tests of episodic memory, where cutoff scores are commonly used to aid clinical decision-making, detection rates were also poor, ranging from 0.26 to 0.71, although specificity was substantially higher, ranging from 0.58 to 0.97. Using norms from Robust Normals did not significantly improve sensitivity or specificity compared to the Combined sample according to the McNemar test statistic.

The FCSRT author has published a cut-score of 24 for the free recall condition and 46 for the total recall score for classification of normal versus those with very mild dementia, based on a receiver operating characteristic (ROC) curve analysis (Grober et al., 2010). Use of this cut-score in our sample resulted in a sensitivity of 0.87 and a specificity of 0.79. Using our convention of −1 and −1.5 standard deviations, cut scores on FCSRT free recall for the Robust Normal group would fall at scores of 27 and 25 for ages 65–74.9 and at scores of 22 and 19 for ages 75 and older. Sensitivity and specificity using robust norms cut scores (Table 7) were lower than those observed in Grober et al., again suggesting that standard deviation cut scores are inferior to other methods.

4. Discussion

The goal of the study was to determine how inclusion of cognitively normal individuals with biomarker and neuroimaging evidence of preclinical AD pathology influences normative cognitive data in terms of mean scores, variability estimates, age effects, and diagnostic classifications. We hypothesized that removing participants that exceeded empirically validated thresholds for amyloid neuroimaging, cerebrospinal fluid AD biomarkers, and hippocampal volume would improve the psychometric characteristics of normative data and enhance sensitivity and specificity of clinical diagnosis. Overall, we found that failing to remove participants with preclinical AD pathology (PCAD) had multiple deleterious impacts on normative data and that the most pronounced effects were seen in individuals aged 75 years and older, when AD risk is substantially increased. Perhaps the most notable finding was that inclusion of participants with PCAD greatly exaggerated age-related declines in cognitive functioning across multiple domains. Contrary to expectations, diagnostic accuracy using z-score cutoffs remained unacceptably poor, and was not improved by using normative data from the Robust Normals, thus reinforcing that reliance on divergence from group normative values, regardless of the quality of the normative data, too often fails to detect clinically symptomatic individuals.

This study is among the first to use biomarker and imaging data for normative sample inclusion criteria. Most studies have used longitudinal follow-up to exclude individuals that develop dementia during follow-up periods with the goal of removing the influence of preclinical AD or other pathology from normative samples. While this "longitudinal" method has several strengths, it suffers from its reliance on a clinical diagnosis of dementia that must occur during the follow-up period in order to justify removing a participant from the normative sample. It is likely that several individuals in these samples may already have AD pathology, but are not manifesting symptoms sufficient to warrant a clinical diagnosis during the follow-up period, or the follow-up was too short to provide an opportunity to progress to symptomatic AD. As mentioned previously, several recent studies have shown that the preclinical phase of AD can last decades and that an accelerating course of subtle

cognitive decline appears 7–10 years prior to clinical diagnosis (Bateman et al., 2012; Grober et al., 2008; Howieson et al., 2008; Johnson et al., 2009; Roe et al., 2013). In addition, clinical diagnosis, even from National Institute on Aging funded Alzheimer's Disease Centers, is inaccurate in 15% or more of cases, when compared to the gold standard of neuropathologic diagnosis (Beach et al., 2012).

Despite the clear limitations in the use of longitudinal samples to generate robust normative data, our approach of using biomarker and imaging evidence of preclinical AD did not substantially improve mean scores or variability estimates in the 65–74.9 year old group. Among the 75 and older participants, however, there were some modest improvements on measures of episodic memory and decreased variability on visuospatial performance, suggesting that there may be a stronger influence of AD pathology in participants above the age of 75 years. It is well understood that the likelihood of dementia increases with increasing age (Alzheimer's Association, 2014) and indeed the percentage of the sample classified as having AD pathology was slightly higher in the 75 and older normative group than the group aged 65–74.9 (40.5% vs 29.5%, respectively). However, the prevalence rates were based on the presence of any biomarker of AD pathology or hippocampal volume reduction, and did not attempt to classify individuals according to preclinical stages or severity of AD pathology. Considering base rates of AD across the lifespan, it is likely that the older PCAD individuals had more disease burden and were more likely to have more pronounced cognitive declines.

The influence of age on test performance in cognitively normal older adults appears to be strongly related to the presence of AD pathology. Age-related decline in cognitive abilities has been described as a generalized process that impacts all cognitive domains, but is most pronounced on measures that emphasize speed of performance (Salthouse, 2000) and on measures of episodic memory (Deary et al., 2009; Luo and Craik, 2008; Small, 2001). In simple regression models that used age to predict cognitive performance, we of course replicated what multiple prior studies have shown: across nearly all tests of cognitive function, performance worsens with increasing age and a substantial amount of variance in performance can be explained by age. Identification and removal of participants with preclinical AD pathology had profound effects on these estimates, and this was particularly evident on tests of episodic memory and tests with a speed component. For example, age accounted for 12.4% of variance on the Benton Visual Retention Test delayed recall in the combined sample, a highly significant effect ($p < 0.0001$). After removing the PCAD participants, this effect was substantially attenuated, explaining just 3.2% of the variance ($p = 0.045$). Variance due to age on episodic measures was reduced by an average of 5.05%, and was no longer significant on WMS-R Logical Memory Immediate and Delayed recall. Similarly, on speeded tests, including Trails A & B, WMS-R Digit Symbol, WAIS Block Design, Animal and Vegetable Fluency, and Word Fluency, variance due to age was reduced by an average of 3.6% and age effects were no longer significant on WAIS Block Design. The influence of age on cognitive performance in older adults appears to be highly related to the degree and stage of AD pathology (Price et al., 2009; Rodrigue et al., 2012; Sliwinski et al., 2003; Vos et al., 2013). In some studies, common neuropathologies including β-amyloidosis and neurofibrillary tangle pathology can account for nearly all cognitive declines attributable to age (Wilson et al., 2010; Yu et al., 2014a), however, our data and

evidence from other cohorts suggest that residual age-related cognitive decline still occurs after controlling for AD pathology (Oh et al., 2012; Yu et al., 2014b), albeit at a much slower rate.

Diagnostic accuracy using commonly used cutoff points at 1.0 and 1.5 standard deviations below the mean produced notably poor detection rates regardless of the normative dataset used. However, some measures of episodic memory did approach acceptable detection rates and were slightly enhanced with normative data from the Robust Normals. Using a −1.5 standard deviation cutoff score on FCSRT Free Recall, sensitivity was increased from 0.39 using Combined norms to 0.71 using Robust norms, but specificity dropped from 0.88 with Combined norms to 0.82 with Robust norms. Similarly, a −1.5 standard deviation cutoff score on WMS-R Logical Memory Delayed recall improved from a sensitivity of 0.35 using Combined norms to a sensitivity of 0.58 using Robust norms, and specificity remained unchanged at 0.96. The best performing individual measure was the FCSRT Total Recall Score, which had a sensitivity of 0.71 and a specificity of 0.90 at a −1.0 standard deviation cutoff. However, using Robust norms did not improve classification accuracy on this measure. It is important to note that the FCSRT Total Recall represents the sum of the free recall (more difficult) and cued (less difficult) conditions and had a large ceiling effect in our data. This ceiling effect has been reported previously by the test's authors in older cognitively normal populations (Grober et al., 1998), which calls into question the utility of the total recall score in detecting early stage declines. Overall, use of cutoff scores to detect very mild symptomatic AD (including individuals who may be characterized as having mild cognitive impairment) was only minimally enhanced by use of Robust norms. In some cases, use of Robust norms actually made clinical detection slightly worse. Even with the most optimal combinations of sensitivity and specificity, as seen on the FCSRT total recall using Robust norms, a clinical diagnosis based on cutoff scores on cognitive tests was incorrect in nearly 30 percent of cases (Table 7). This finding is also consistent with our prior report using a much smaller autopsy-confirmed cohort (Storandt and Morris, 2010).

While our study has several strengths, including the large sample of individuals with amyloid neuroimaging, CSF biomarkers, structural neuroimaging, and cognitive assessments, our results must be understood in the context of several relevant limitations. First, our sample is composed mainly of highly educated and Caucasian participants who volunteer for clinical, cognitive, neuroimaging, and fluid biomarker studies. Thus, the results may not generalize to those from more diverse racial and ethnic backgrounds or those with lower levels of educational attainment. Second, our cognitive measures are widely used in aging and dementia studies but due to the historical legacy of our dataset are somewhat limited in their sensitivity and in coverage of cognitive domains. Third, the number older participants with complete imaging and biomarker evaluation was limited, leaving a relatively smaller sample size in the 75 years and older PCAD group for analyses. Fourth, our methods for clinical diagnosis rely on clinical assessment and the CDR, and are not influenced by results of our cognitive battery. We believe this is a strength in that we avoid circularity when comparing cognitive function across groups, but this contrasts with methods used by many clinicians, who would diagnose dementia or mild cognitive impairment with objective evidence of decline on at least one domain of cognition. However, we have shown that the CDR corresponds highly with clinical diagnoses, including mild

cognitive impairment (Morris et al., 2001). Finally, our use of categorical distinctions for biomarkers and hippocampal volumes may be suboptimal for defining preclinical AD. The classifications are based on cross-sectional data that does not account for intra-individual changes or individual differences. Confirmation of our results will require additional studies that focus on serially collected neuroimaging and fluid biomarkers.

Acknowledgments

References

Agresti, A. Categorical Data Analysis. 2. Hoboken, NJ: 2002.

Alzheimer's, Association. 2014 Alzheimer's disease facts and figures. Alzheimers Dement. 2014; 10:e47–92. [PubMed: 24818261]

Armitage SG. An analysis of certain psychological tests used for the evaluation of brain injury. Psychological Monographs. 1946; 60:i–48. DOI: 10.1037/h0093567

Bateman RJ, Xiong C, Benzinger TLS, Fagan AM, Goate A, Fox NC, Marcus DS, Cairns NJ, Xie X, Blazey TM, Holtzman DM, Santacruz A, Buckles V, Oliver A, Moulder K, Aisen PS, Ghetti B, Klunk WE, McDade E, Martins RN, Masters CL, Mayeux R, Ringman JM, Rossor MN, Schofield PR, Sperling RA, Salloway S, Morris JC. Dominantly Inherited Alzheimer Network. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. N Engl J Med. 2012; 367:795–804. DOI: 10.1056/NEJMoa1202753 [PubMed: 22784036]

Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. J Neuropathol Exp Neurol. 2012; 71:266–273. DOI: 10.1097/NEN.0b013e31824b211b [PubMed: 22437338]

Benton AL. The revised visual retention test: clinical and experimental applications. Psychological Corporation. 1963

Berg L, McKeel DW Jr, Miller JP, Storandt M, Rubin EH, Morris JC, Baty J, Coats M, Norton J, Goate AM. Clinicopathologic studies in cognitively healthy aging and Alzheimer disease: relation of histologic markers to dementia severity, age, sex, and apolipoprotein E genotype. Archives of Neurology. 1998; 55:326. [PubMed: 9520006]

Berg L, Miller JP, Baty J, Rubin EH, Morris JC, Figiel G. Mild senile dementia of the Alzheimer type. 4. Evaluation of intervention. Annals of Neurology. 1992; 31:242–249. DOI: 10.1002/ana.410310303 [PubMed: 1637132]

Buckner RL, Head D, Parker J, Fotenos AF, Marcus D, Morris JC, Snyder AZ. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. NeuroImage. 2004; 23:724–738. DOI: 10.1016/j.neuroimage.2004.06.018 [PubMed: 15488422]

Coats M, Morris JC. Antecedent biomarkers of Alzheimer's disease: the adult children study. Journal of Geriatric Psychiatry and Neurology. 2005; 18:242–244. DOI: 10.1177/0891988705281881 [PubMed: 16306247]

De Santi S, Pirraglia E, Barr W, Babb J, Williams S, Rogers K, Glodzik L, Brys M, Mosconi L, Reisberg B, Ferris S, de Leon MJ. Robust and conventional neuropsychological norms: diagnosis and prediction of age-related cognitive decline. Neuropsychology. 2008; 22:469–484. DOI: 10.1037/0894-4105.22.4.469 [PubMed: 18590359]

Deary IJ, Corley J, Gow AJ, Harris SE, Houlihan LM, Marioni RE, Penke L, Rafnsson SB, Starr JM. Age-associated cognitive decline. British Medical Bulletin. 2009; 92:135–152. DOI: 10.1093/bmb/ldp033 [PubMed: 19776035]

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage. 2006; 31:968–980. DOI: 10.1016/j.neuroimage.2006.01.021 [PubMed: 16530430]

Fagan AM, Roe CM, Xiong C, Mintun MA, Morris JC, Holtzman DM. Cerebrospinal Fluid tau/β-Amyloid42 Ratio as a Prediction of Cognitive Decline in Nondemented Older Adults. Archives of Neurology. 2007; 64:343–349. DOI: 10.1001/archneur.64.3.noc60123 [PubMed: 17210801]

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002; 33:341–355. [PubMed: 11832223]

Foltynie T. The cognitive ability of an incident cohort of Parkinson's patients in the UK. The CamPaIGN study. Brain. 2003; 127:550–560. DOI: 10.1093/brain/awh067 [PubMed: 14691062]

Goodglass H, Kaplan E. Boston diagnostic aphasia examination booklet. 1983

Grober E, Buschke H, Crystal H, Bang S, Dresner R. Screening for dementia by memory testing. Neurology. 1988; 38:900–903. [PubMed: 3368071]

Grober E, Hall CB, Lipton RB, Zonderman AB, Resnick SM, Kawas C. Memory impairment, executive dysfunction, and intellectual decline in preclinical Alzheimer's disease. J Int Neuropsychol Soc. 2008; 14:266–278. DOI: 10.1017/S1355617708080302 [PubMed: 18282324]

Grober E, Lipton RB, Katz M, Sliwinski M. Demographic influences on free and cued selective reminding performance in older persons. J Clin Exp Neuropsychol. 1998; 20:221–226. DOI: 10.1076/jcen.20.2.221.1177 [PubMed: 9777476]

Grober E, Sanders AE, Hall C, Lipton RB. Free and cued selective reminding identifies very mild dementia in primary care. Alzheimer Dis Assoc Disord. 2010; 24:284–290. DOI: 10.1097/WAD.0b013e3181cfc78b [PubMed: 20683186]

Hayden KM, Reed BR, Manly JJ, Tommet D, Pietrzak RH, Chelune GJ, Yang FM, Revell AJ, Bennett DA, Jones RN. Cognitive decline in the elderly: an analysis of population heterogeneity. Age and Ageing. 2011; 40:684–689. DOI: 10.1093/ageing/afr101 [PubMed: 21890481]

Holtzer R, Goldin Y, Zimmerman M, Katz M, Buschke H, Lipton RB. Robust norms for selected neuropsychological tests in older adults. Archives of Clinical Neuropsychology. 2008; 23:531–541. DOI: 10.1016/j.acn.2008.05.004 [PubMed: 18572380]

Howieson DB, Carlson NE, Moore MM, Wasserman D, Abendroth CD, Payne-Murphy J, Kaye JA. Trajectory of mild cognitive impairment onset. J Int Neuropsychol Soc. 2008; 14:192–198. DOI: 10.1017/S1355617708080375 [PubMed: 18282317]

Jack CR Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, Shaw LM, Vemuri P, Wiste HJ, Weigand SD, Lesnick TG, Pankratz VS, Donohue MC, Trojanowski JQ. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. The Lancet Neurology. 2013; 12:207–216. DOI: 10.1016/S1474-4422(12)70291-0 [PubMed: 23332364]

Jack CR, Knopman DS, Weigand SD, Wiste HJ, Vemuri P, Lowe V, Kantarci K, Gunter JL, Senjem ML, Ivnik RJ, Roberts RO, Rocca WA, Boeve BF, Petersen RC. An operational approach to National Institute on Aging-Alzheimer's Association criteria for preclinical Alzheimer disease. Annals of Neurology. 2012; 71:765–775. DOI: 10.1002/ana.22628 [PubMed: 22488240]

Jack CR, Twomey CK, Zinsmeister AR, Sharbrough FW, Petersen RC, Cascino GD. Anterior temporal lobes and hippocampal formations: normative volumetric measurements from MR images in young adults. Radiology. 1989; 172:549–554. [PubMed: 2748838]

Jak AJ, Bondi MW, Delano-Wood L, Wierenga C, Corey-Bloom J, Salmon DP, Delis DC. Quantification of five neuropsychological approaches to defining mild cognitive impairment. Am J Geriatr Psychiatry. 2009; 17:368–375. DOI: 10.1097/JGP.0b013e31819431d5 [PubMed: 19390294]

Johnson DK, Storandt M, Morris JC, Galvin JE. Longitudinal study of the transition from healthy aging to Alzheimer disease. Archives of Neurology. 2009; 66:1254–1259. DOI: 10.1001/archneurol.2009.158 [PubMed: 19822781]

Johnson DK, Storandt M, Morris JC, Langford ZD, Galvin JE. Cognitive profiles in dementia: Alzheimer disease vs healthy brain aging. Neurology. 2008; 71:1783–1789. DOI: 10.1212/01.wnl. 0000335972.35970.70 [PubMed: 19029518]

Knopman DS, Jack CR, Wiste HJ, Weigand SD, Vemuri P, Lowe V, Kantarci K, Gunter JL, Senjem ML, Ivnik RJ, Roberts RO, Boeve BF, Petersen RC. Short-term clinical outcomes for stages of NIA-AA preclinical Alzheimer disease. Neurology. 2012; 78:1576–1582. DOI: 10.1212/WNL. 0b013e3182563bbe [PubMed: 22551733]

Luo L, Craik FI. Aging and memory: a cognitive approach. The Canadian Journal of Psychiatry/La Revue canadienne de psychiatrie. 2008

McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika. 1947; 12:153–157. [PubMed: 20254758]

Mintun M, Larossa G, Sheline Y, Dence C, Lee S, Mach R, Klunk W, Mathis C, DeKosky S, Morris J. [11C] PIB in a nondemented population Potential antecedent marker of Alzheimer disease. Neurology. 2006; 67:446–452. DOI: 10.1212/01.wnl.0000228230.26044.a4 [PubMed: 16894106]

Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology. 1993; 43:2412–2414. [PubMed: 8232972]

Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH, Berg L. Mild cognitive impairment represents early-stage Alzheimer disease. Archives of Neurology. 2001; 58:397–405. [PubMed: 11255443]

Morris JC, Weintraub S, Chui HC, Cummings J, Decarli C, Ferris S, Foster NL, Galasko D, Graff-Radford N, Peskind ER, Beekly D, Ramos EM, Kukull WA. The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. Alzheimer Dis Assoc Disord. 2006; 20:210–216. DOI: 10.1097/01.wad.0000213865.09806.92 [PubMed: 17132964]

Oh H, Madison C, Haight TJ, Markley C, Jagust WJ. Effects of age and β-amyloid on cognitive changes in normal elderly people. Neurobiology of aging. 2012; 33:2746–2755. DOI: 10.1016/j.neurobiolaging.2012.02.008 [PubMed: 22429886]

Pedraza O, Lucas JA, Smith GE, Petersen RC, Graff-Radford NR, Ivnik RJ. Robust and Expanded Norms for the Dementia Rating Scale. Archives of Clinical Neuropsychology. 2010; 25:347–358. DOI: 10.1093/arclin/acq030 [PubMed: 20427376]

Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Archives of Neurology. 1999; 56:303–308. [PubMed: 10190820]

Pizzie R, Hindman H, Roe CM, Head D, Grant E, Morris JC, Hassenstab JJ. Physical activity and cognitive trajectories in cognitively normal adults: the adult children study. Alzheimer Dis Assoc Disord. 2014; 28:50–57. DOI: 10.1097/WAD.0b013e31829628d4 [PubMed: 23739296]

Pizzie R, Hindman H, Roe CM, Head D, Grant E, Morris JC, Hassenstab JJ. Physical Activity and Cognitive Trajectories in Cognitively Normal Adults: The Adult Children Study. Alzheimer Dis Assoc Disord Publish Ahead of Print. 2013; :1.doi: 10.1097/WAD.0b013e31829628d4

Price JL, McKeel DW, Buckles VD, Roe CM, Xiong C, Grundman M, Hansen LA, Petersen RC, Parisi JE, Dickson DW, Smith CD, Davis DG, Schmitt FA, Markesbery WR, Kaye J, Kurlan R, Hulette C, Kurland BF, Higdon R, Kukull W, Morris JC. Neuropathology of nondemented aging: presumptive evidence for preclinical Alzheimer disease. Neurobiology of aging. 2009; 30:1026–1036. DOI: 10.1016/j.neurobiolaging.2009.04.002 [PubMed: 19376612]

Price JL, Morris JC. Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. Annals of Neurology. 1999; 45:358–368. [PubMed: 10072051]

Ritchie LJ, Frerichs RJ, Tuokko H. Effective Normative Samples For the Detection Of Cognitive Impairment in Older Adults. Clin Neuropsychol. 2007; 21:863–874. DOI: 10.1080/13854040701557239 [PubMed: 17853155]

Rodrigue KM, Kennedy KM, Devous MD, Rieck JR, Hebrank AC, Diaz-Arrastia R, Mathews D, Park DC. β-Amyloid burden in healthy aging: regional distribution and cognitive consequences. Neurology. 2012; 78:387–395. DOI: 10.1212/WNL.0b013e318245d295 [PubMed: 22302550]

Roe CM, Fagan AM, Grant EA, Hassenstab J, Moulder KL, Dreyfus DM, Sutphen CL, Benzinger TL, Mintun MA, Holtzman DM. Amyloid imaging and CSF biomarkers in predicting cognitive

impairment up to 7.5 years later. Neurology. 2013; 80:1784–1791. DOI: 10.1212/WNL. 0b013e3182918ca6 [PubMed: 23576620]

Salthouse TA. Aging and measures of processing speed. Biol Psychol. 2000; 54:35–54. [PubMed: 11035219]

Saxton J, Lopez OL, Ratcliff G, Dulberg C, Fried LP, Carlson MC, Newman AB, Kuller L. Preclinical Alzheimer disease: Neuropsychological test performance 1.5 to 8 years prior to onset. Neurology. 2004; 63:2341–2347. DOI: 10.1212/01.WNL.0000147470.58328.50 [PubMed: 15623697]

Sliwinski M, Lipton RB, Buschke H, Stewart W. The effects of preclinical dementia on estimates of normal cognitive functioning in aging. J Gerontol B Psychol Sci Soc Sci. 1996; 51:P217–25. [PubMed: 8673642]

Sliwinski MJ, Hofer SM, Hall C, Buschke H, Lipton RB. Modeling Memory Decline in Older Adults: The Importance of Preclinical Dementia, Attrition, and Chronological Age. Psychol Aging. 2003; 18:658–671. DOI: 10.1037/0882-7974.18.4.658 [PubMed: 14692855]

Small SA. Age-related memory decline: current concepts and future directions. Archives of Neurology. 2001; 58:360–364. [PubMed: 11255438]

Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR, Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Siemers E, Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorad M, Wagster MV, Phelps CH. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Presented at the Alzheimer"s & dementia: the journal of the Alzheimer"s Association. 2011; :280–292. DOI: 10.1016/j.jalz.2011.03.003

Storandt M, Morris JC. Ascertainment bias in the clinical diagnosis of Alzheimer disease. Archives of Neurology. 2010; 67:1364–1369. DOI: 10.1001/archneurol.2010.272 [PubMed: 21060013]

Thurstone TG, Thurstone LL. SRA primary mental abilities. 1948

Villemagne VL, Burnham S, Bourgeat P, Brown B. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. The Lancet. 2013; doi: 10.1016/S1474-4422(13)70044-9

Vos SJ, Xiong C, Visser PJ, Jasielec MS, HASSENSTAB J, Grant EA, Cairns NJ, Morris JC, Holtzman DM, Fagan AM. Preclinical Alzheimer's disease and its outcome: a longitudinal cohort study. Lancet Neurology. 2013; 12:957–965. DOI: 10.1016/S1474-4422(13)70194-7 [PubMed: 24012374]

Wechsler D. WMS-R: Wechsler Memory Scale-Revised: Manual. 1987

Wechsler D. Wechsler: Wechsler adult intelligence scale-revised - Google Scholar. 1981

Weintraub S, Salmon D, Mercaldo N, Ferris S, Graff-Radford NR, Chui H, Cummings J, Decarli C, Foster NL, Galasko D, Peskind E, Dietrich W, Beekly DL, Kukull WA, Morris JC. The Alzheimer"s Disease Centers" Uniform Data Set (UDS): the neuropsychologic test battery. Alzheimer Dis Assoc Disord. 2009; 23:91–101. DOI: 10.1097/WAD.0b013e318191c7dd [PubMed: 19474567]

Wilson RS, Leurgans SE, Boyle PA, Schneider JA, Bennett DA. Neurodegenerative basis of age-related cognitive decline. Neurology. 2010; 75:1070–1078. DOI: 10.1212/WNL. 0b013e3181f39adc [PubMed: 20844243]

Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO, Nordberg A, Backman L, Albert M, Almkvist O, Arai H, Basun H, Blennow K, de Leon M, DeCarli C, Erkinjuntti T, Giacobini E, Graff C, Hardy J, Jack C, Jorm A, Ritchie K, van Duijn C, Visser P, Petersen RC. Mild cognitive impairment - beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. J Intern Med. 2004; 256:240–246. DOI: 10.1111/j. 1365-2796.2004.01380.x [PubMed: 15324367]

Yu L, Boyle PA, Leurgans S, Schneider JA, Bennett DA. Neurobiology of Aging. NBA. 2014a; 35:819–826. DOI: 10.1016/j.neurobiolaging.2013.10.074

Yu L, Boyle PA, Segawa E, Leurgans S, Schneider JA, Wilson RS, Bennett DA. Residual Decline in Cognition After Adjustment for Common Neuropathologic Conditions. Neuropsychology. 2014b; doi: 10.1037/neu0000159

Highlights

- Normative samples may include individuals with preclinical AD.

- Preclinical AD has deleterious impacts on normative cognitive data.

- Age-related cognitive decline is reduced or attenuated when excluding preclinical AD.

- Use of standard deviation cutoff scores produces poor diagnostic accuracy.
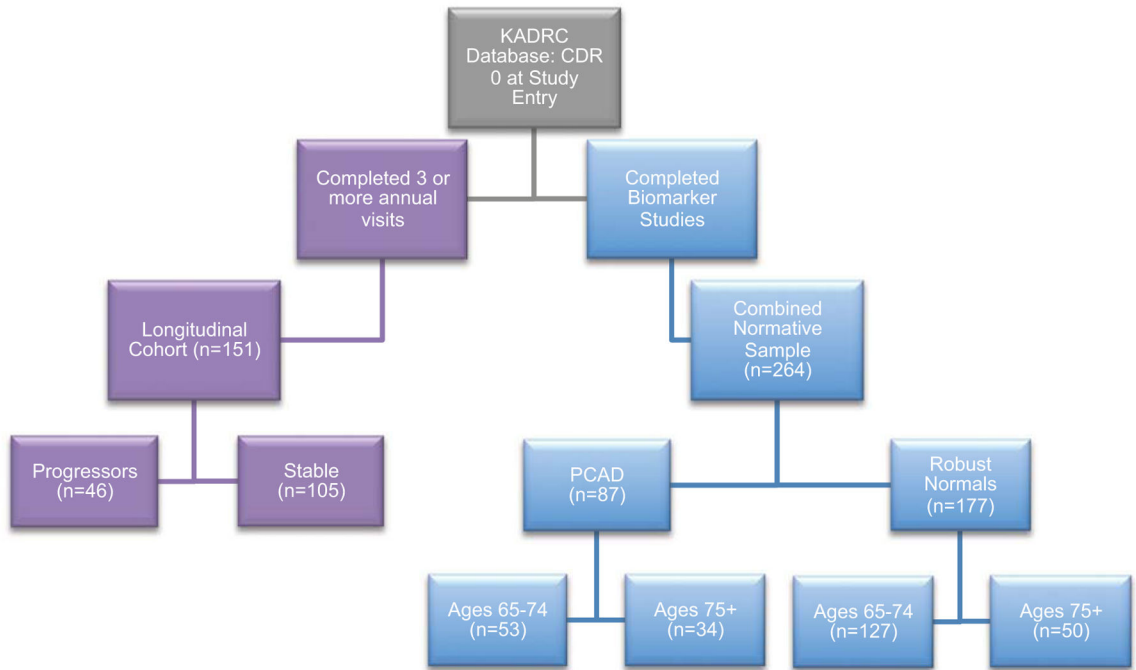
Figure 1.
Participant selection flowchart. KADRC = Knight Alzheimer's Disease Research Center,
PCAD = Preclinical Alzheimer's Disease Pathology.
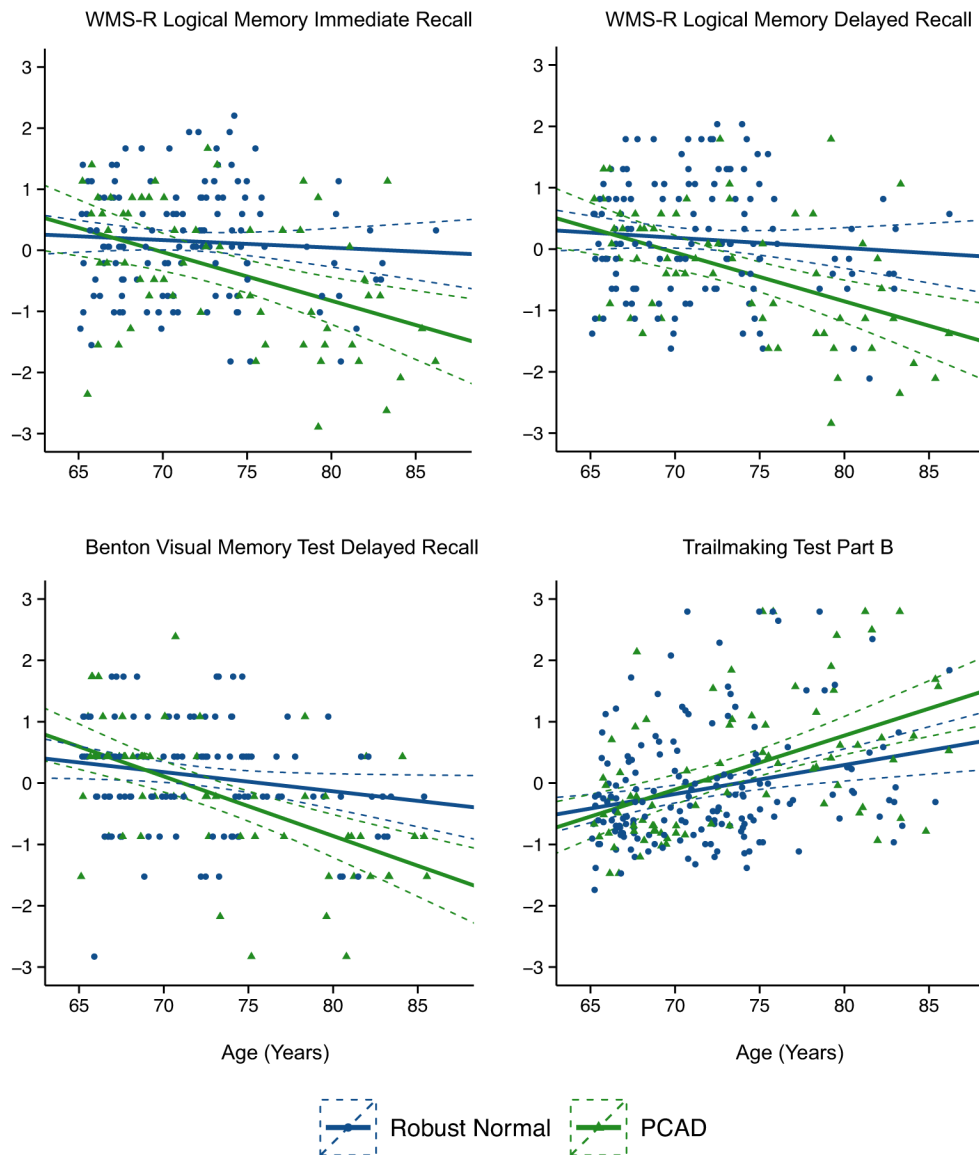
Figure 2.
Age-effects on cognitive performance. X axis represents age in years. Y axis represents z-score transformed cognitive score (for ease of interpretation). Dashed lines represent 95% confidence intervals. Statistics provided for all measures in Table 5. Abbreviations: PCAD = Preclinical Alzheimer's Disease pathology. WMS-R = Wechsler Memory Scale-Revised.

Table 1

Criteria for Suspected Alzheimer's Disease Pathology (SADP)

Must have CSF or Neuroimaging evidence of β-amyloidosis alone or β-amyloidosis and Neurodegeneration

Pathology	CSF Biomarker		Neuroimaging Biomarker
β-amyloidosis	$A\beta_{1-42} < 459$ pg/mL	-or-	PIB PET MCBP > 0.18
Neurodegeneration	t-tau > 339 pg/mL -or- $p\text{-tau}_{181} > 67$ pg/mL	-or-	Hippocampal volume > −1.5 SDs

Table 2

Demographic and Biomarker variables for A) Normative sample and B) Longitudinal test sample.

A. Normative Sample	Robust Normals	SADP	Combined Sample
n	177	87	264
Age (years, SD)[a]	71.62 (5.0)	73.10 (6.2)	72.11 (5.4)
Education (years, SD)	15.32 (2.6)	15.85 (3.0)	15.5 (2.8)
Gender (%Female)	58.19%	40.23%	58.71%
APOE Status (%e4+)[***]	21.60%	50.60%	31.25%
Caucasian (%)	94%	93%	93.70%
MMSE (raw score, SD)	28.89 (1.3)	29.11 (1.1)	29.02 (1.2)
CSF Amyloid-β_{1-42} (pg/mL)[***]	750.93 (215.5)	393.35 (172.0)	631.20 (263.1)
CSF Total Tau (pg/mL)[***]	288.35 (138.5)	401.31 (241.8)	326.17 (187.0)
CSF P-Tau$_{181}$ (pg/mL)[***]	54.63 (24.0)	70.49 (35.9)	59.94 (29.5)
PIB-PET (MCBP, SD)[***]	0.04 (0.1)	0.42 (0.3)	0.18 (0.3)
Hippocampal Volume (mL, SD)*	7240.32 (845.4)	6942.90 (904.7)	7136.38 (876.2)

B. Longitudinal Sample	Stable	Progressor
n	105	46
Age at baseline (years, SD)	74.79 (7.26)	76.22 (7.14)
Age at diagnosis (years, SD)	--	81.51 (6.82)
Education (years, SD)	15.43 (2.96)	14.9 (3.5)
Gender (%Female)	54.29%	56.50%
APOE Status (%e4+)[a]	28.30%	43.48%
Caucasian (%)[a]	87.76%	97.72%
MMSE at baseline (raw score, SD)	28.37 (1.7)	28.88 (1.24)
MMSE at diagnosis (raw score, SD)	--	26.77 (2.19)

[***] $p < 0.001$

[**] $p < 0.01$

[a] $p < 0.10$ for comparisons of Robust Normals vs. SADP and Stable vs. Progressor

Table 3

Cognitive performance at AGES 65–74 in Robust Normals, SADP, and the Combined Sample. Group comparisons adjusted for gender and education.

Cognitive Tests	A. Robust Normals raw score (SD)	B. SADP raw score (SD)	C. Combined Sample raw score (SD)	A vs B	A vs C	B vs C
Episodic Memory						
WMS-R Logical Memory Immediate	13.52 (3.26)	13.34 (3.6)	13.47 (3.34)	$t(3,128) = -0.87$	$t(3,225) = -0.35$	$t(3,163) = -0.64$
WMS-R Logical Memory Delayed	12.66 (3.79)	12.11 (3.14)	12.52 (3.63)	$t(3,128) = -1.48$	$t(3,225) = -0.59$	$t(3,163) = -1.08$
WMS Associate Learning	13.97 (3.63)	14.42 (3.42)	14.11 (3.56)	$t(3,169) = 0.37$	$t(3,290) = 0.15$	$t(3,221) = 0.29$
FCSRT Total Recall	47.94 (0.24)	47.91 (0.37)	47.93 (0.28)	$t(3,153) = -0.74$	$t(3,267) = -0.31$	$t(3,196) = -0.51$
FCSRT Free Recall	31.07 (4.38)	30.29 (5.26)	30.50 (5.03)	$t(3,153) = -0.19$	$t(3,267) = -0.08$	$t(3,196) = -0.15$
BVRT Delayed	6.80 (1.38)	6.68 (1.41)	6.72 (1.39)	$t(3,121) = -0.02$	$t(3,206) = -0.03$	$t(3,161) = -0.04$
Executive/Visuospatial						
Trails A (seconds)	31.94 (9.85)	32.32 (10.11)	32.06 (9.9)	$t(3,176) = 0.34$	$t(3,303) = 0.13$	$t(3,229) = 0.28$
Trails B (seconds)	76.96 (27.83)	79.19 (27.66)	78.56 (27.72)	$t(3,166) = 0.06$	$t(3,288) = 0.03$	$t(3,214) = 0.04$
WMS-R Digit Span Forward	6.79 (1.13)	6.71 (1.07)	6.76 (1.11)	$t(3,170) = -0.57$	$t(3,292) = -0.24$	$t(3,222) = -0.42$
WMS-R Digit Span Backward	4.84 (1.23)	5.12 (1.17)	4.93 (1.22)	$t(3,170) = 1.17$	$t(3,292) = 0.49$	$t(3,222) = 0.84$
WMS-R Digit Symbol Coding	51.5 (9.46)	49.34 (10.45)	49.99 (9.79)	$t(3,169) = -0.80$	$t(3,290) = -0.37$	$t(3,221) = -0.51$
WMS Mental Control	7.71 (1.6)	7.62 (1.82)	7.68 (1.66)	$t(3,170) = -0.84$	$t(3,292) = -0.34$	$t(3,222) = -0.63$
WAIS Block Design	35.15 (7.7)	31.89 (7.01)	32.85 (7.35)	$t(3,168) = -2.66$**	$t(3,289) = -1.09$	$t(3,219) = -1.92$[a]
Language/Semantic						
Animal Fluency	21.14 (5.15)	21.82 (4.83)	21.33 (5.06)	$t(3,167) = 0.64$	$t(3,289) = 0.26$	$t(3,216) = 0.48$
Vegetable Fluency	15.05 (4.16)	14.97 (4.50)	15.03 (4.23)	$t(3,128) = -0.75$	$t(3,225) = -0.52$	$t(3,163) = -0.52$
Word Fluency Letters S&P total	31.24 (9.42)	33.98 (9.46)	32.06 (9.48)	$t(3,170) = 1.32$	$t(3,292) = 0.55$	$t(3,222) = 0.96$
Boston Naming Test	55.88 (4.25)	56.58 (4.17)	56.09 (4.23)	$t(3,170) = 0.92$	$t(3,292) = 0.39$	$t(3,222) = 0.66$
WAIS Information	21.58 (3.96)	21.18 (3.95)	21.46 (3.95)	$t(3,169) = -1.39$	$t(3,291) = -0.58$	$t(3,220) = -1.01$

***
$p < 0.001$

**
$p < 0.01$

*
$p < 0.05$,

[a]
$p < 0.10$

Abbreviations: SD = Standard Deviation, WMS-R = Wechsler Memory Scale Revised, WMS = Wechsler Memory Scale, FCSRT = Free and Cued Selective Reminding Test, BVRT = Benton Visual Retention Test, WAIS = Wechsler Adult Intelligence Scale

Table 4

Cognitive performance at AGES 75 and older in Robust Normals, SADP, and the Combined Sample. Group comparisons adjusted for gender and education.

Cognitive Tests	A. Robust Normals	B. SADP	C. Combined Sample	A vs B	A vs C	B vs C
	raw score (SD)	raw score (SD)	raw score (SD)			
Episodic Memory						
WMS-R Logical Memory Immediate	12.86 (3.48)	9.33 (4.06)	11.35 (4.1)	$t(3,59) = -4.04^{***}$	$t(3,95) = -1.96^{a}$	$t(3,86) = -2.33^{*}$
WMS-R Logical Memory Delayed	11.36 (3.85)	7.81 (4.47)	9.84 (4.46)	$t(3,59) = -3.69^{***}$	$t(3,95) = -1.82^{a}$	$t(3,86) = -2.16^{*}$
WMS Associate Learning	12.98 (3.89)	11.42 (3.01)	12.36 (3.63)	$t(3,79) = -1.64$	$t(3,129) = -0.79$	$t(3,112) = -1.10$
FCSRT Total Recall	47.86 (0.47)	47.55 (0.69)	47.73 (0.58)	$t(3,67) = -2.19^{*}$	$t(3,109) = -1.14$	$t(3,96) = -1.33$
FCSRT Free Recall	28.36 (6.48)	26.24 (5.34)	27.49 (6.09)	$t(3,67) = -1.44$	$t(3,109) = -0.66$	$t(3,96) = -1.02$
BVRT Delayed	6.15 (1.17)	4.67 (1.63)	5.59 (1.53)	$t(3,60) = -4.40^{***}$	$t(3,100) = -2.00^{*}$	$t(3,84) = -2.58^{*}$
Executive/Visuospatial						
Trails A (seconds)	35.12 (11.39)	40.24 (14.17)	37.19 (12.76)	$t(3,80) = 2.17^{*}$	$t(3,130) = 1.07$	$t(3,114) = 1.39$
Trails B (seconds)	93.04 (37.16)	115.47 (36.99)	102.12 (38.50)	$t(3,80) = 2.87^{*}$	$t(3,130) = 1.41$	$t(3,114) = 1.79^{a}$
WMS-R Digit Span Forward	6.40 (1.32)	6.32 (1.27)	6.37 (1.30)	$t(3,80) = -0.10$	$t(3,130) = -0.02$	$t(3,114) = -0.11$
WMS-R Digit Span Backward	4.62 (1.12)	4.62 (1.30)	4.62 (1.19)	$t(3,80) = 0.06$	$t(3,130) = 0.10$	$t(3,114) = -0.06$
WMS-R Digit Symbol Coding	48.00 (10.22)	40.24 (8.42)	44.82 (10.22)	$t(3,79) = -3.36^{**}$	$t(3,128) = -1.58$	$t(3,113) = -2.15^{a}$
WMS Mental Control	7.18 (1.83)	7.56 (1.65)	7.33 (1.76)	$t(3,80) = 0.91$	$t(3,130) = 0.50$	$t(3,114) = 0.58$
WAIS Block Design	30.48 (8.97)	26.79 (7.21)	28.99 (8.46)	$t(3,80) = -2.65^{**}$	$t(3,130) = -1.29$	$t(3,114) = -1.69$
Language/Semantic						
Animal Fluency	18.31 (4.71)	16.91 (3.87)	17.73 (4.41)	$t(3,79) = -0.10$	$t(3,130) = -0.02$	$t(3,114) = -0.11$
Vegetable Fluency	14.14 (4.08)	11.96 (4.03)	13.21 (4.17)	$t(3,59) = -2.03^{*}$	$t(3,95) = -1.03$	$t(3,86) = -1.27$
Word Fluency Letters S&P total	29.04 (9.99)	29.27 (9.77)	29.13 (9.84)	$t(3,79) = -0.05$	$t(3,129) = -0.02$	$t(3,112) = -0.02$
Boston Naming Test	53.88 (5.21)	51.59 (6.38)	52.95 (5.79)	$t(3,80) = -2.49^{*}$	$t(3,130) = -1.20$	$t(3,114) = -1.61$
WAIS Information	20.96 (4.67)	20.35 (4.96)	20.71 (4.77)	$t(3,80) = -1.27$	$t(3,130) = -0.61$	$t(3,114) = -0.85$

***
$p < 0.001$

**
$p < 0.01$

*
$p < 0.05$,

a
$p < 0.10$

Table 5

Variance in cognitive performance attributable to age.

Cognitive Tests	Robust Normals	SADP	Combined Sample
Episodic Memory	R^2	R^2	R^2
WMS-R Logical Memory Immediate	0.004	0.181[***]	0.064[***]
WMS-R Logical Memory Delayed	0.007	0.214[***]	0.072[***]
WMS Associate Learning	0.024[*]	0.185[***]	0.067[***]
FCSRT Total Recall	0.029[*]	0.125[**]	0.078[***]
FCSRT Free Recall	0.039[*]	0.193[***]	0.082[***]
BVRT Delayed	0.032[*]	0.298[***]	0.124[***]
Executive/Visuospatial			
Trails A (seconds)	0.033[*]	0.052[*]	0.045[***]
Trails B (seconds)	0.065[***]	0.240[***]	0.132[***]
WMS-R Digit Span Forward	0.008	0.022	0.014
WMS-R Digit Span Backward	0.016	0.017	0.014
WMS-R Digit Symbol Coding	0.078[***]	0.256[***]	0.142[***]
WMS Mental Control	0.036[*]	0.001	0.010
WAIS Block Design	0.015	0.195[***]	0.06[***]
Language/Semantic			
Animal Fluency	0.068[***]	0.170[***]	0.101[***]
Vegetable Fluency	0.017	0.072[*]	0.041[**]
Word Fluency Letters S&P total	0.016	0.061[*]	0.026[**]
Boston Naming Test	0.029[*]	0.088[**]	0.053[***]
WAIS Information	0.005	0.001	0.003

[***] $p < 0.001$

[**] $p < 0.01$

[*] $p < 0.05$

Table 6

Results of ROC analyses in the Longitudinal Sample (n=151) using either Combined Norms or Robust Norms

	AUC (95% CI)	
Cognitive Tests	**Combined Norms**	**Robust Norms**
Episodic Memory		
WMS-R Logical Memory Immediate	0.83 (0.74–0.92)	0.87 (0.79–0.95)
WMS-R Logical Memory Delayed	0.85 (0.77–0.93)	0.88 (0.80–0.95)
WMS Associate Learning	0.75 (0.66–0.83)	0.76 (0.66–0.85)
FCSRT Total Recall	0.90 (0.84–0.96)	0.89 (0.83–0.95)
FCSRT Free Recall	0.72 (0.56–0.87)	0.72 (0.57–0.88)
BVRT Delayed	0.63 (0.47–0.79)	0.63 (0.47–0.79)
Executive/Visuospatial		
Trails A (seconds)	0.40 (0.30–0.50)	0.39 (0.29–0.49)
Trails B (seconds)	0.40 (0.30–0.51)	0.39 (0.28–0.49)
WMS-R Digit Span Forward	0.52 (0.41–0.63)	0.52 (0.41–0.63)
WMS-R Digit Span Backward	0.51 (0.40–0.62)	0.51 (0.40–0.62)
WMS-R Digit Symbol Coding	0.66 (0.55–0.76)	0.66 (0.56–0.76)
WMS Mental Control	0.52 (0.41–0.64)	0.52 (0.41–0.64)
WAIS Block Design	0.55 (0.45–0.66)	0.53 (0.43–0.63)
Language/Semantic		
Animal Fluency	0.76 (0.67–0.85)	0.78 (0.69–0.86)
Vegetable Fluency	0.82 (0.73–0.90)	0.84 (0.76–0.92)
Word Fluency Letters S&P total	0.57 (0.46–0.69)	0.58 (0.46–0.69)
Boston Naming Test	0.68 (0.57–0.79)	0.32 (0.21–0.42)
WAIS Information	0.68 (0.58–0.78)	0.68 (0.58–0.78)

Abbreviations: ROC = receiver operating characteristic; AUC = area under the curve

CI: confidence interval

Table 7

Application of normative data to a separate longitudinal sample (n=151). Sensitivity and specificity were calculated at time of clinical diagnosis, based on Clinical Dementia Rating (CDR). Test scores normed using Robust Normal and Combined Samples. Classification accuracy was compared using the McNemar test.

Cognitive Tests	Normative Sample	−1 SD			−1.5 SD		
		Sens	Spec	Z^0	Sens	Spec	Z^0
Episodic Memory							
WMS-R Logical Memory Immediate	Robust Normals	0.65	0.88		0.54	0.95	
	Combined	0.54	0.93	0.07	0.35	0.96	−0.22
WMS-R Logical Memory Delayed	Robust Normals	0.62	0.90		0.58	0.96	
	Combined	0.58	0.93	0.15	0.35	0.96	−0.37
WMS Associate Learning	Robust Normals	0.45	0.87		0.21	0.96	
	Combined	0.39	0.86	−0.14	0.13	0.97	−0.14
FCSRT Total Recall	Robust Normals	0.71	0.88		0.58	0.93	
	Combined	0.71	0.90	0.14	0.58	0.94	0.07
FCSRT Free Recall	Robust Normals	0.71	0.82		0.71	0.82	
	Combined	0.71	0.82	0.00	0.39	0.88	−0.37
BVRT Delayed	Robust Normals	0.61	0.58		0.39	0.78	
	Combined	0.39	0.78	0.34	0.26	0.93	0.33
Executive/Visuospatial							
Trails A	Robust Normals	0.39	0.71		0.26	0.84	
	Combined	0.34	0.74	0.07	0.21	0.88	0.14
Trails B	Robust Normals	0.41	0.75		0.32	0.81	
	Combined	0.32	0.77	−0.07	0.32	0.84	0.21
WMS-R Digit Span Forward	Robust Normals	0.11	0.86		0.05	0.90	
	Combined	0.11	0.86	0.00	0.05	0.90	0.00
WMS-R Digit Span Backward	Robust Normals	0.21	0.86		0.05	0.99	
	Combined	0.21	0.86	0.00	0.05	0.92	−0.49
WMS-R Digit Symbol Coding	Robust Normals	0.59	0.68		0.35	0.80	
	Combined	0.43	0.73	0.00	0.19	0.89	0.21
WMS Mental Control	Robust Normals	0.13	0.76	0.00	0.08	0.91	0.00

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Cognitive Tests	Normative Sample	−1 SD			−1.5 SD		
		Sens	Spec	Z⁰	Sens	Spec	Z⁰
	Combined	0.13	0.76		0.08	0.91	
WAIS Block Design	Robust Normals	0.32	0.76	0.15	0.11	0.87	0.00
	Combined	0.32	0.78		0.08	0.88	
Language/Semantic							
Animal Fluency	Robust Normals	0.53	0.86	−0.14	0.42	0.91	0.00
	Combined	0.53	0.84		0.42	0.91	
Vegetable Fluency	Robust Normals	0.73	0.79	−0.08	0.46	0.91	−0.23
	Combined	0.58	0.83		0.15	0.98	
Word Fluency Letters S&P total	Robust Normals	0.35	0.86	−0.28	0.14	0.94	0.00
	Combined	0.35	0.82		0.14	0.94	
Boston Naming Test	Robust Normals	0.11	0.81	1.42	0.05	0.95	0.41
	Combined	0.45	0.88		0.29	0.92	
WAIS Information	Robust Normals	0.34	0.86	−0.21	0.21	0.90	0.00
	Combined	0.26	0.86		0.21	0.90	

Abbreviations: SD = Standard Deviation, Sens = Sensitivity, Spec = Specificity, Z⁰ = McNemar test statistic