Published in final edited form as:

Neuropsychology. 2008 July ; 22(4): 469–484. doi:10.1037/0894-4105.22.4.469.

Robust and Conventional Neuropsychological Norms: Diagnosis and Prediction of Age-Related Cognitive Decline

Susan De Santi,

Department of Psychiatry, New York University School of Medicine

Elizabeth Pirraglia,

Department of Psychiatry, New York University School of Medicine

William Barr,

Departments of Neurology and Psychiatry, New York University School of Medicine

James Babb,

Department of Radiology, New York University School of Medicine

Schantel Williams,

Department of Psychiatry, New York University School of Medicine

Kimberley Rogers,

Department of Psychiatry, New York University School of Medicine

Lidia Glodzik,

Department of Psychiatry, New York University School of Medicine

Mirosław Brys,

Department of Psychiatry, New York University School of Medicine

Lisa Mosconi,

Department of Psychiatry, New York University School of Medicine

Barry Reisberg,

Department of Psychiatry, New York University School of Medicine

Steven Ferris, and

Department of Psychiatry, New York University School of Medicine and The Nathan S. Kline Institute for Psychiatric Research

Mony J. de Leon

Department of Psychiatry, New York University School of Medicine and The Nathan S. Kline Institute for Psychiatric Research

Abstract

The aim of the study was to compare the performance of Robust and Conventional neuropsychological norms in predicting clinical decline among healthy adults and in mild cognitive impairment (MCI). The authors developed Robust baseline cross sectional and longitudinal change norms from 113 healthy participants retaining a normal diagnosis for at least 4 years. Baseline

Correspondence concerning this article should be addressed to Susan De Santi, Department of Psychiatry, New York University School of Medicine, 550 First Avenue (MHL-408), New York, NY 10016. E-mail: susan.desanti@med.nyu.edu.
Kimberley Rogers is now at the Department of Psychology, The Graduate Center of the City University of New York and the Department of Psychiatry, Mount Sinai School of Medicine.

Supplemental materials: <http://dx.doi.org/10.1037/0894-4105.22.4.469.supp>

Conventional norms were separately created for 256 similar healthy participants without follow-up. Conventional and Robust norms were tested in an independent cohort of longitudinally studied healthy ($n = 223$), MCI ($n = 136$), and Alzheimer's disease (AD, $n = 162$) participants; 84 healthy participants declined to MCI or AD (NL→DEC), and 44 MCI declined to AD (MCI→AD). Compared to Conventional norms, baseline Robust norms correctly identified a higher proportion of NL→DEC with impairment in delayed memory and attention-language domains. Both norms predicted decline from MCI→AD. Change norms for delayed memory and attention-language significantly incremented baseline classification accuracies. These findings indicate that Robust norms improve identification of healthy individuals who will decline and may be useful for selecting at-risk participants for research studies and early interventions.

Keywords

cognition; longitudinal; normative data; decline from normal; mild cognitive impairment

Improved understanding of normal cognitive functioning in an aging population is essential for recognizing the early cognitive changes associated with degenerative diseases such as Alzheimer's disease (AD). Recognizing abnormal cognitive functioning at earlier time points will help identify individuals at risk of developing dementia (Cargin, Maruff, Collie, & Masters, 2006; Tierney, Yao, Kiss, & McDowell, 2005) and those who might benefit from receiving early treatment. Consequently, the use of valid normative information is essential for detecting very early abnormal cognitive impairment that predicts subsequent decline to a clinical diagnosis of AD or its prodrome, mild cognitive impairment (MCI).

Comparing an individual's cognitive performance with that of a reference or normative group studied at a single time point is the conventional method to identify abnormal cognitive performance (Kendall, Marrs-Garcia, Nath, & Sheldrick, 1999; Kendall & Sheldrick, 2000). However, existing norms for older persons have been called into question because the assumption of normality was not well established (Fastenau, Denburg, & Hufford, 1999; Levine, Miller, Becker, Selnes, & Cohen, 2004; Norman, Evans, Miller, & Heaton, 2000; Sheridan et al., 2006) or because the reference groups included individuals at preclinical stages of dementia (Sliwinski, Lipton, Buschke, & Stewart, 1996). The inclusion in a normative sample of individuals who later decline to AD, lower the mean test scores, amplify the variance, and exaggerate the apparent effect of age on cognition (Ritchie, Frerichs, & Tuokko, 2007; Sliwinski et al., 1996). Such Conventional norms may underestimate the recognition of pathology (Galvin et al., 2005; Manly et al., 2005; Sliwinski et al., 1996). Robust norms, developed on normal individuals studied longitudinally to exclude individuals with future cognitive impairment, were more sensitive in detecting dementia (Sliwinski et al., 1996). Although this novel examination of normal cognition identified a potential problem with Conventional norms, the study did not experimentally test the utility of Robust norms compared to Conventional norms or their relative performance in predicting decline. More recently, Robust norms were developed in individuals with a high IQ score who remained longitudinally high cognitive performers (i.e., within 1.5 *SD* of published norms for all tests administered; Rentz et al., 2004), but no comparison with Conventional norms was included in the study. In another recent study Conventional and Robust norms were compared for diagnostic utility, rather than predictive value (Ritchie et al., 2007). To our knowledge, few studies predicting future clinical decline to MCI or AD have tested the performance of normative data (Chen et al., 2001; Marquis et al., 2002; Powell et al., 2006). Most prediction studies typically examine specific tests (Fleisher et al., 2007; Masur, Sliwinski, Lipton, Blau, & Crystal, 1994; Tabert et al., 2006; Tierney et al., 2005) rather than norms as predictors of decline.

Cognitive impairment has been identified in clinically normal individuals who go on to MCI (Blacker et al., 2007; Glodzik-Sobanska et al., 2007; Tierney et al., 1996) or AD (Masur et al., 1994; Meguro et al., 2001; Rubin et al., 1998;). Manly et al. (2005) incorporated a Robust normative group for operationalizing an MCI diagnosis, and identified the prevalence of MCI in an ethnically diverse urban population (Manly et al., 2005). Using Robust norms, they showed that MCI was more prevalent in those older than 75, and in those with less education (<9 years). Cognitive norms in a normal aging population may be confounded by the inclusion of cognitively abnormal individuals who would later go on to develop MCI or AD. The construction of a normative group, after excluding individuals who subsequently developed MCI (Rentz et al., 2004) or AD (Sliwinski et al., 1996), would provide a more homogeneous, clinically stable normal aging reference group to which cognitively impaired clinical groups can be compared. Although the above studies suggest that Robust norms have an advantage beyond Conventional norms, our goal was to experimentally test the utility of this concept in individuals studied longitudinally.

The primary aims of this study were to determine if Robust norms are more accurate than Conventional norms in identifying and predicting cognitive impairment and to determine if change in cognitive performance over time increments the effectiveness of baseline norms in the prediction of future decline. The first hypothesis was that Robust norms from participants with cognitive stability will provide a more accurate measure of normal cognition and distinguish cognitive impairment at baseline in participants who subsequently decline to MCI or AD. The second hypothesis was that change in memory performance over time would improve the utility of baseline performance in predicting decline. Our results support the concept that Robust norms are better than Conventional norms for the identification of current and future cognitive impairment at baseline, and such norms are more effective at later time points when longitudinal change is included.

Method

Participants

We examined data from 890 community dwelling volunteers and patients, retrospectively studied, who were drawn from a larger pool of individuals ($N = 3,987$) participating in brain aging studies at the NYU Alzheimer's Disease Center (ADC) and the affiliated NYU Center for Brain Health (CBH) since 1978. Participants were not actively recruited for this study but were selected from the preexisting database, provided that they fit the inclusion and exclusion criteria. Participants with MCI or AD were required to have at least three evaluations. The majority of participants in the preexisting database had at least 12 years of education (88%), were of middle to upper socioeconomic status, and were predominately White (89%). Participants received an extensive diagnostic evaluation that included medical, neurological, psychiatric, neuropsychological examinations, and brain computerized tomographic (CT) or magnetic resonance (MR) imaging. Informed consent, approved by the NYU School of Medicine IRB, was obtained from all participants at each evaluation.

Study Inclusion Criteria

Participants whose data were included in the current analyses were between the ages of 40 and 92 with at least 12 years of education. Participants who were not native English speakers ($n = 56$) were included if they achieved a scaled score of at least 11 on a WAIS Vocabulary subtest, which is considered an average score in a general population (Wechsler, 1955). Participants were included if their Global Deterioration Scale (GDS; Reisberg, Ferris, de Leon, & Crook, 1982) score was ≤ 5 .

Study Exclusion Criteria

Individuals with a history of stroke or significant evidence of cerebro-vascular disease (modified Hachinski score ≥ 4 ; Hachinski et al., 1975; Rosen, Terry, Fuld, Katzman, & Peck, 1980) were excluded, as were individuals with evidence of any neurological or psychiatric disorder other than MCI or AD, (e.g., normal pressure hydrocephalus or depression as defined by a Hamilton Depression Scale; Hamilton, 1960, score > 9).

Diagnostic Procedures

A semistructured clinical interview using the Brief Cognitive Rating Scale (BCRS; Reisberg & Ferris, 1988) assessed the magnitude of cognitive impairment in concentration, recent and past memory, orientation, and functioning/self care and provided a numerical score for each domain. Information obtained from the BCRS was used to determine the GDS. The GDS (Reisberg et al., 1982), a seven-point rating scale uses validated descriptors to assess the global cognitive and level of functional capacity as follows: normal (NL; GDS = 1 or 2 characterized as cognitively and functionally normal; differentiated by the absence vs. presence of subjective memory complaints, respectively), MCI (GDS = 3), mild to moderate AD (GDS = 4 or 5), and severe AD (GDS ≥ 6 ; Reisberg et al., 1993). The GDS score was assigned independent of, and prior to, the detailed neuropsychological tests in this study. Furthermore, the clinicians performing the GDS assignment were in all cases blind to the findings from the neuropsychological testing procedures.

All diagnoses were made at a consensus meeting. The diagnosis of AD followed NINCDS-ADRDA (McKhann et al., 1984) and *Diagnostic and Statistical Manual of Mental Disorders-IV* (American Psychiatric Association, 1994) criteria. The NL and MCI diagnoses were based on the GDS and relevant medical data obtained at the evaluation and were not directly defined using psychometric test performance (Convit et al., 1997; de Leon et al., 2001, 1993; De Santi et al., 2001; Flicker, Ferris, & Reisberg, 1991; Kluger, Ferris, Golomb, Mittelman, & Reisberg, 1999; Pritchep et al., 2006). Criteria for the diagnosis of MCI were: memory complaint documented by the patient and collateral source, normal general cognition, normal activities of daily living, no dementia (American Psychiatric Association, 1994), mild cognitive deficits (typically memory) and a GDS score of 3. Participants were included in either the normative or test groups as described below, with all groups being mutually exclusive.

Normative Samples

The normative sample was composed of 369 diagnostically normal individuals. These individuals were divided into two mutually exclusive groups: the “Conventional Normative Group” included 256 individuals with only one diagnostic evaluation and the “Robust Normative Group” included 113 individuals who received at least three diagnostic evaluations and retained the clinical diagnosis of normal at each evaluation. The average total observation period for the Robust Normative group was 8.3 ± 4.0 years (range: 3.6 to 21.5 years) over an average of 5 ± 2 total evaluations. Using the Robust Normative group, both cross-sectional baseline and longitudinal change norms were developed. Separate cross-sectional baseline norms were developed using the Conventional Normative group. No longitudinal data was available for the Conventional Normative group as they were only studied once. Participants from this group either did not return or had not yet completed their follow-up evaluation. (See Table 1 for demographic information for these groups, Figure 1 for each group’s diagnosis at three time points, and supplemental Appendix A for group description).

Test Sample

The Conventional and Robust norms were tested in an independent cohort of longitudinally studied individuals called the “test sample” that was mutually exclusive of the two normative

samples. All participants in the test sample had at least three evaluations except the healthy test group (HT) who had two evaluations. The test sample included 223 NL individuals, 136 MCI participants, and 162 AD participants. Of the 223 NL participants, 139 participants were diagnosed as normal at follow-up and composed the HT group. These HT participants were demographically comparable to the Conventional and Robust Normative groups and had one follow-up at the time of this report. Therefore, the time that they remained stable was shorter than the Robust group. The HT group was included within the study design as a control group for the test sample and for comparison with the normative groups (see Table 1). For the HT participants, the interval between baseline and follow-up evaluation was 3.3 ± 2.3 years (range: 1.4 to 15.3 years). The remaining 84 individuals with a normal diagnosis at baseline clinically declined by their final follow-up evaluation (NL→DEC) and received a final diagnosis of MCI ($n = 48$) or AD ($n = 36$). These participants were diagnosed as normal for 6.4 ± 4.4 years (range: 1.3 to 19.7 years) before they were first diagnosed with cognitive impairment (MCI, $n = 75$ or AD, $n = 9$). To examine a possible relationship between cognitive performance and the time to decline, this NL→DEC group was subsequently subdivided into early decliners (decline in less than 5 years after the baseline evaluation; NL_EDEC, $n = 41$) and late decliners (decline 5 or more years after the baseline evaluation; NL_LDEC, $n = 43$). To have a maximal number of participants in both the early and late subgroups, the NL→DEC group was split using the median decline time of 5 years (see Table 2 for demographic information for NL→DEC subgroups).

Of the 136 MCI participants, 92 remained MCI at all follow-up evaluations (stable MCI) and were studied for 3.4 ± 2.2 years (range: 1.1 to 14.5 years). The remaining 44 MCI patients declined to AD at any subsequent evaluation (MCI→AD). These 44 participants remained MCI for 3.1 ± 2.0 years (range: 1.3 to 11.4 years) before receiving a diagnosis of AD.

The AD group consisted of 162 patients who retained the AD diagnosis (AD→AD) and who were studied for 5.1 ± 3.4 years (range: 1.1 to 18.4 years; see Table 3 for demographic information for all test groups, Figure 1 for each group's diagnosis at three time points, and Appendix A for group description).

Cognitive Tests

The cognitive test battery administered in this study included the Guild Memory Test (Gilbert, 1970; Gilbert, Levee, & Catalano, 1968) to assess several components of memory function including: paragraph immediate and delayed recall, immediate and delayed recall of verbal paired associates, and immediate recall of visual paired associates. The test battery also included several subtests from the Wechsler Adult Intelligence Scale (Wechsler, 1955); specifically, Digits Forward and Backward, the Digit Symbol Substitution Test (DSST), and the Vocabulary Test. The WAIS-R (Wechsler, 1981) version of these subtests was administered to participants in this study. In addition, we examined performance on the Object Naming Test (Flicker, Ferris, & Reisberg, 1993), the delayed trial from the NYU Shopping List Test (Flicker et al., 1993), and Perceptual Speed, a number cancellation test (PSPC; Moran & Mefferd, 1959). (See Tables 4 and 5 for baseline cognitive data for all groups). The six subtests from the Guild Memory Test have been shown to correlate with the memory quotient from the Wechsler Memory Scale (r s ranging from .55 to .67, p s < .001; Crook, Gilbert, & Ferris, 1980). Given that the repetition of a memory test may result in a learning effect, the reliability for the Guild Memory Test was not examined directly. However, the two paragraphs that comprise the paragraph recall subtests are significantly correlated ($r = .62$; $p < .001$). Moreover, the Guild Memory Test has good test–retest reliability (r s ranging from .47 to .93; p s < .01; Reisberg et al., 1989). The paragraph subtest of the Guild Memory Test, commonly known as the NYU Paragraph Recall Test, has been studied extensively (Convit et al., 1997; de Leon et

al., 2001; de Leon et al., 1997; de Leon et al., 1993; Flicker et al., 1991; Kluger et al., 1999), and used in clinical drug trials (Petersen et al., 2005; Salloway et al., 2004).

Statistical Analyses

We examined group differences in continuous demographic variables (e.g., age) using analysis of variance (ANOVA) with post hoc Tukey's tests. Effect sizes for *t* tests and ANOVAs were measured with Cohen's *d*. Using this measure, an effect size of .20 is considered a small effect, .50 is considered a medium effect, and .80 is considered a large effect. Group differences for categorical demographic variables and cognitive domains were compared using exact Pearson chi-squared analyses with Bonferroni correction for multiple comparisons. Effects sizes for chi-squared analyses were measured using Cramér's *V*. With this measure a value of .10 is considered a small effect, .30 is considered a moderate effect, and .50 is considered a large effect. The McNemar test for paired binomial proportions was used to test differences in sensitivities for normative comparisons within diagnostic groups. Corrections for multiple comparisons of diagnostic group were not needed for this test because analyses were performed within group. Effects sizes for the McNemar tests were measured using the phi coefficient (ϕ). With this measure a value of .10 is considered a small effect, .30 is considered a moderate effect, and .50 is considered a large effect. Logistic regression analyses were used to test the additive value of norms for longitudinal change to norms for baseline performance. Effect sizes for logistic regression analyses were reported in terms of odds ratios. Statistical significance was defined as *p* values $\leq .05$ except in the Pearson chi-squared analyses where statistical significance was defined as *p* values $\leq .01$ to correct for multiple comparisons. SPSS (Version 12.0) was used for data analyses.

Cross-Sectional Norm Development

Norms for both the Conventional and Robust groups were developed separately based on the baseline cross-sectional test scores. Using linear regression analyses, the influence of age, years of education, and gender on cognitive scores was assessed and found to be associated with the cognitive variables. Age was associated with lower test scores on all 11 cognitive tests in the Conventional Normative group and with 7 of the 11 tests in the Robust Normative group (*ts* for β range from 3.4 to 6.4, *ps* $< .05$). In addition, education and gender were correlated with test performance, although not as strongly as age. Therefore, we included these three demographic variables in creating an estimated equation for cognitive performance on each test, following procedures similar to Heaton et al. (Heaton, Avitable, Grant, & Matthews, 1999). Significant two-way interactions between the variables (e.g., age–education; age–gender; gender–education) were added into the estimated equations. For each test, the intercept plus the beta weights of the demographic variables and the significant interactions were used to calculate predicted scores. We generated a *z* score for each test variable as follows:

$$z = (\text{raw score} - \text{predicted score})/SD,$$

where *SD* is the root mean squared error from the estimated equation.

Longitudinal Change Norms

Longitudinal change norms, calculated from the Robust Normative group, were based on the change in cognitive performance between baseline and the first follow-up visit. Regression models were generated accounting for age, years of education, gender, time between test administrations, baseline score, and any significant two-way interactions of the demographic variables. The resulting estimated equations were used to generate predicted scores for each test. We calculated *z* scores using the equation above.

Definition of Cognitive Domains

Using an ROC curve we identified the cutpoint of the z score with a 90% specificity on each test for both the Conventional and Robust norms. Both sets of norms were assessed in the test sample. Participants in the test sample were then coded as impaired on a cognitive test if their score was below the cutpoint. We examined the sensitivity for each test group using both Conventional and Robust norms. After correcting for multiple comparisons, most tests could distinguish individuals with overt cognitive impairment (i.e., MCI→AD and AD→AD) compared to the HT group. However, only 6 of the 11 tests identified stable MCI patients and none of the tests for either set of norms identified NL→DEC compared to HT participants. Therefore, the test by test approach was not sensitive in detecting healthy participants who would show future cognitive decline.

To improve on the sensitivity to detect impairment, our second approach was to combine tests based on cognitive domains. This approach has been applied by others using various methodologies to examine cognitive change associated with the onset of dementia and normal aging (Galvin et al., 2005; Palmer, Boone, Lesser, & Wohl, 1998; Rubin et al., 1998). To find subtle but genuine impairment in cognition the criteria of defining impairment on at least two tests within a domain provided a more rigorous method than impairment on individual tests within a domain. This reduced the identification of individuals as impaired who simply performed poorly on one particular test. The cognitive tests were categorized based on face validity into four separate domains: (a) working memory, (b) immediate memory, (c) delayed memory, and (d) attention-language. Although it is unconventional to combine attention and language tests into the same domain, only one language test was administered to the majority of the study cohort. Given that no one test was examined in isolation by our criteria and the Object Naming Test was more highly correlated with DSST and PSPC than with tests from other domains, we felt the combination of attention and language was acceptable. The working memory domain consisted of WAIS-R Digits Forward and Backward. The immediate recall of paragraphs, verbal paired associates, and visual paired associates comprised the immediate memory domain. The delayed recall of paragraphs, verbal paired associates, and word list comprised the delayed memory domain. DSST, PSPC, and the Object Naming Test comprised the attention-language domain.

Unlike a test by test approach, identifying cut-off levels for any combination of two or three tests within a domain did not lend itself to using the ROC curve method due to the complexity of dealing with numerous tests. The following procedure was used to define cutoff scores at baseline within a domain. The standard deviation level (e.g., $< -1 SD$, $< -1.5 SD$, $< -2 SD$) was selected for each test such that the combination of any two tests within a domain would classify 5 to 10% of the HT participants as impaired (i.e., 90% to 95% specificity) using Conventional norms. We chose this specificity range to reflect a conservative estimation of normal participants who may decline in the future. The z score cutoff selected for the Robust norms was the level that provided the same specificity as the respective Conventional norms (see Table 6).

Comparing Sensitivity Between Conventional and Robust Norms at Baseline

By design, the specificity for the HT group using either Conventional or Robust norms was identical within each domain. This provided a fair comparison for the sensitivities between the two norm methods. The McNemar test (McNemar, 1947) was applied within each test group to evaluate the performance of the two normative procedures with respect to sensitivity.

The McNemar Test criterion is given by the equation:

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

where C = number of cases correctly classified by Conventional norms only and R = number of cases correctly classified using Robust norms only. The significance level is obtained by comparing the square of Z_0 to a chi-squared distribution with $df = 1$ (Agresti, 2002). The effect size for the McNemar test was calculated using the phi coefficient:

$$\phi = (Z_0^2/n)^{1/2},$$

where Z_0^2 is equivalent to chi-squared in the standard calculation of the phi coefficient.

Definition of Longitudinal Neuropsychological Change Within a Clinically Stable Normal Group

The Robust Norm group was used to define normal change in cognitive performance between the baseline and first follow-up evaluation. The optimal cut-off level for normal change was defined as the level that achieved specificity between 90% and 95% for the Robust Norm group. Excessive cognitive change from baseline to follow-up was defined as exceeding the standard cutoff level (z score) for at least one of the tests within a domain (see Table 6 for cutoff levels). The baseline criterion of impairment on two out of three tests within a domain was too conservative for the longitudinal data as it yielded specificities above 95%.

First Follow-Up

One of the goals of this study was to understand the change within a stable group and to characterize abnormal change. Therefore, at the first follow-up only participants who retained their baseline diagnosis were included in the analyses. For clarity, the declining groups were renamed to include their baseline, first follow-up, and final diagnoses (NL-NL→DEC; $n = 52$; MCI-MCI→AD; $n = 12$; see Appendix A). Within the NL-NL→DEC group, the effect of early (NL-NL_EDEC) and late decline (NL-NL_LDEC) was also examined. Decline within 5 years of the follow-up evaluation was defined as early decline and beyond that time frame decline was defined as late decline. Within the MCI groups, stability versus decline (stable MCI vs. MCI-MCI→AD) was examined. All groups were compared to the HT group.

Results

Demographics

Normative Groups—The Conventional and Robust Norm groups did not differ from each other with respect to baseline age, education, gender, or MMSE scores ($p > .05$; see Table 1).

Test Samples—The HT group was representative of the normative groups described above, and used as the comparison group in all analyses with the other test groups. The HT group did not differ significantly from the two normative groups with respect to age, education, gender, or MMSE scores ($p > .05$; see Table 1), but by study design they had been followed for a shorter period of time compared to the Robust group, $t(250) = 12.39$; $p < .001$, $d = 1.5$.

The HT group was significantly younger than the other test groups, $F(4, 516) = 38.7$; $p < .001$, $d = .94$ to 1.3 . Although the AD→AD group received less education than the HT, NL→DEC, and stable MCI groups, $F(4, 516) = 8.5$; $p < .001$, $d = .68$, the mean education level for the

AD-AD group was nevertheless high at 14.7 years. As expected the AD-AD group had lower MMSE scores compared to all other test groups and the MCI groups had lower MMSE scores compared to the HT group, $F(4, 516) = 165.8; p < .001, d = 1.1$ to 2.3 . There were no gender differences between the groups. There was a significantly higher percentage of Whites in the Stable MCI group (98%) compared to the HT group (89%), $\chi^2(1, N = 230) = 5.9, p < .05, V = .16$. With respect to Apolipoprotein E genotype (ApoE), the NL→DEC group had a significantly lower proportion of participants carrying the E4 allele (19%) compared to HT (39%), $\chi^2(1, N = 145) = 6.8, p < .05, V = .22$, whereas the AD-AD group had a significantly higher proportion of E4 carriers (60%), $\chi^2(1, N = 166) = 7.0, p < .05, V = .21$ (see Table 3).

Conventional Versus Robust Norms at Baseline

Normal Diagnostic Groups

Cognitive performance of NL→DEC compared to HT: Comparing the HT and the NL→DEC group at baseline (when all participants were normal), Robust norms identified a significantly higher proportion of NL→DEC participants with impaired cognition in the delayed memory domain, $\chi^2(1, N = 223) = 8.4, p < .01, V = .19$; and the attention-language domain, $\chi^2(1, N = 223) = 7.1, p < .01, V = .18$. The Conventional norms did not identify significant group differences in any domain. Given that the HT group was significantly younger than the NL→DEC group, we reexamined our data limiting the age of both groups to those above 65 years. Our findings did not change for the delayed memory domain in that a higher proportion of NL→DEC, 65 and older, had impaired delayed memory using Robust norms, $\chi^2(1, N = 147) = 9.3, p \leq .01, V = .25$. There was a trend toward significance in the attention-language domain using Robust norms where a higher proportion of NL→DEC, 65 and older, showed impairment, $\chi^2(1, N = 147) = 5.7, p = .02, V = .20$.

Comparison of norms within NL→DEC: Robust norms had significantly higher sensitivity than the Conventional norms in the delayed memory domain, 20% versus 10%; McNemar: $Z_0 = 3.0; p < .01, \varpi = .29$, and the attention-language domain, 18% versus 7%; McNemar: $Z_0 = 3.0, p < .01, \varpi = .29$, (see Table 7 and Figure 2). We confirmed our results in the sample limited to participants 65 and older group, such that Robust norms outperformed Conventional norms in these domains.

NL→DEC Subgroups

Cognitive Performance of NL_EDEC Compared to HT: The NL_EDEC (decline in less than 5 years after the baseline evaluation, $n = 41$) group was significantly older, $F(2, 220) = 20.6; p \leq .01, d = .80$, and had a lower proportion of ApoE4 carriers, $\chi^2(1, N = 180) = 3.5, p \leq .05, V = .19$, compared to the HT (see Table 2).

A significantly greater percentage of the NL_EDEC group compared to HT were classified as impaired in delayed and immediate memory using Robust norms, $\chi^2(1, N = 180) = 17.1, p < .01, V = .31$; $\chi^2(1, N = 180) = 10.8, p \leq .01, V = .24$, respectively. Only immediate memory showed significant group differences using Conventional norms, $\chi^2(1, N = 180) = 8.4, p \leq .01, V = .22$.

Comparison of norms within NL_EDEC: There was a significant difference in sensitivity between Robust and Conventional norms for delayed memory (32% vs. 15%, respectively) McNemar: $Z_0 = 2.7; p \leq .05, \varpi = .24$, but not for the immediate memory domain (24% and 22%, respectively), (see Table 7 for McNemar results).

Cognitive performance of NL_LDEC compared to HT: As compared to HT the NL_LDEC (decline 5 or more years after the baseline evaluation, $n = 43$) group were significantly older,

$F(2, 220) = 20.6; p < .01, d = 1.1$, and had a lower proportion of ApoE4 carriers, $\chi^2(1, N = 182) = 4.6, p < .05, V = .20$ (see Table 3).

There was no significant difference for any domain between the proportion of NL_LDEC and HT group participants classified as impaired using Conventional norms. Using Robust norms, we found a trend toward significance, $\chi^2(1, N = 182) = 5.7, p = .02, V = .18$, for the proportion of NL_LDEC participants classified as having impaired attention-language compared to HT.

Comparison of norms within NL_LDEC: There were no significant differences in sensitivity between Robust and Conventional norms for attention-language domain (19% vs. 7%, respectively), (see Table 7 for McNemar results).

MCI Diagnostic Groups

Cognitive performance of stable MCI compared to HT: Comparing the stable MCI and the HT groups, both sets of norms identified significantly more stable MCI as showing impairment in delayed memory, Robust norms: $\chi^2(1, N = 231) = 26.6, p < .01, V = .34$; Conventional norms: $\chi^2(1, N = 231) = 15.7, p < .01, V = .26$; and attention-language, Robust norms: $\chi^2(1, N = 231) = 25.3, p < .01, V = .33$; Conventional norms: $\chi^2(1, N = 231) = 13.3, p < .01, V = .24$ domains. Our findings were confirmed when we limited the sample to participants 65 and older. A higher proportion of stable MCI participants had impaired delayed memory, $\chi^2(1, N = 155) = 20.8, p < .001, V = .37$ and attention-language, $\chi^2(1, N = 155) = 16.3, p < .001, V = .33$ domains.

Comparison of norms within stable MCI: The Robust norms had significantly higher sensitivity relative to Conventional norms in the delayed memory (34% vs. 26%; McNemar: $Z_0 = 2.7; p < .05, \omega = .24$), and attention-language (32% vs. 22%; McNemar: $Z_0 = 3.0, p < .05, \omega = .28$) domains (see Table 8 for McNemar results). We confirmed our results in the sample limited to participants 65 and older group, such that Robust norms outperformed Conventional norms in these domains (McNemar $p < .05$).

Cognitive performance of MCI→AD compared to HT: Comparing MCI→AD and HT groups with respect to the percentage of participants classified as impaired, the Robust norms showed a significant group difference in all domains, $\chi^2(1, N = 183)$ range = 10.2 to 76.9, $p \leq .01, V = .24$ to .65, whereas the Conventional norms showed a significant difference in all but the working memory domain $\chi^2(1, N = 183)$ range = 24.4 to 68.7, $p \leq .01, V = .37$ to .61.

Comparison of norms within MCI→AD: The sensitivity for the Robust norms did not significantly differ from that of the Conventional norms for any domain (see Table 8 for McNemar results).

AD→AD Group

Cognitive performance of AD→AD compared to HT: For both norms, the AD→AD group had a significantly higher proportion of individuals classified with impaired cognitive performance in all domains as compared to the HT group, Robust norms: $\chi^2(1, N = 301)$ range = 28.8 to 222.5, $p < .01, V = .31$ to .86; Conventional norms: $\chi^2(1, N = 301)$ range = 20.9 to 222.5, $p < .01, V = .26$ to .86.

Comparison of norms within AD→AD: The sensitivity for the Robust norms did not significantly differ from that of the Conventional norms for any domain (see Table 8 for McNemar results).

In summary, at baseline the NL→DEC and the MCI groups were more impaired on delayed memory and attention-language domains compared to the HT group. The AD group showed

impairment in all domains. See Figure 3 for performance in the delayed memory domain for all groups. Robust norms outperformed Conventional norms in identifying NL→DEC and Stable MCI in the delayed memory and attention-language domains. In NL→DEC, cognitive impairment was more readily identified the closer an individual was to the time of decline.

First Follow-Up

Demographics at Follow-Up—As stated in the Method section, NL and MCI individuals who declined at the first follow-up and progressed to a different diagnostic group were removed from the longitudinal analysis. For example, out of the 84 participants at baseline in the NL→DEC group, 32 declined to MCI or AD at their first follow-up. Therefore, they were removed from all further analyses and the remaining group was renamed NL-NL→DEC, indicating the diagnosis of normal at baseline and at first follow-up, and a future follow-up diagnosis reflecting clinical decline. The same labeling principal was applied to the MCI→AD group, which had 12 participants remaining with an MCI diagnosis at first follow-up (MCI-MCI→AD).

The demographic variables for the individuals remaining in the test groups were reexamined. As compared to the HT group the NL-NL→DEC group ($n = 52$) was significantly older, $F(3, 290) = 34.0$; $p < .001$, $d = .90$, but there were no group differences for MMSE and gender distribution (see Table 9).

As with the baseline results we examined if the point at which decline occurred affected cognitive change. To maximize the size of the subgroups, the median time to decline from first follow-up (5 years) was used to subdivide the group into early (NL-NL_EDEC) and late decliners (NL-NL_LDEC). As compared to the HT group, these two groups were significantly older, $F(2, 188) = 12.9$, $p < .01$, $d = 1.0$ than the HT group. There were no group differences with respect to education, gender, ApoE, and MMSE for the NL-NL_EDEC ($n = 28$), and NL-NL_LDEC ($n = 24$) groups (see Table 9).

Compared to the HT group the Stable MCI and MCI-MCI→AD groups were significantly older, $F(3, 290) = 34.0$, $p < .01$, $d = 1.24$, and had lower MMSE scores, $F(3, 290) = 25.1$, $p < .01$, $d = 1.22$. The Stable MCI group had a higher percentage of Whites (98%) than the HT group (89%), $\chi^2(1, N = 230) = 5.9$, $p < .05$, $V = .16$ (see Table 10).

Combining Baseline and Longitudinal Change Outcomes

Using logistic regression analyses, we examined whether longitudinal change improved the prediction of outcome group beyond that of the cross-sectional baseline performance. Longitudinal change significantly added to baseline performance for delayed memory and attention-language domains in all groups compared to HT (see Table 11 for chi-squared results). The effect sizes were measured in odds ratios (OR) and ranged from 2.6 to 25.2, which indicated large effects. Furthermore, longitudinal change in immediate memory significantly added to baseline performance in stable MCI (OR = 2.5) and MCI-MCI→AD (OR = 4.6) groups compared to HT (see Table 11).

When early and late decline groups were examined, longitudinal change significantly added to baseline prediction for delayed memory (OR = 3.9) and attention-language (OR = 9.1) domains for the NL-NL_EDEC group but not for the NL-NL_LDEC (see Table 11 for chi-squared results). For the NL-NL_LDEC group excessive change was more prevalent than baseline impairment in delayed memory and attention-language domains. For the delayed memory domain within the Stable MCI group, the same proportion of participants had baseline impairment as had excessive change, whereas, a lower proportion showed excessive change than baseline impairment in the attention-language domain. In addition, excessive change

significantly added to baseline impairment for both domains in the prediction of Stable MCI. Similarly, the MCI-MCI→AD group had the same proportion of participants with baseline impairment and excessive change for the delayed memory domain. Within the attention-language domain a higher proportion of these participants had excessive change than baseline impairment.

Discussion

Three main findings emerge from this study. First, Robust norms were more sensitive than Conventional norms in identifying early cognitive impairment in normal individuals who eventually decline to MCI or AD, and in stable MCI. In MCI participants who eventually decline to AD and in AD participants, both Robust and Conventional norms performed equally well at identifying individuals with cognitive impairment. Second, normal individuals who declined within 5 years demonstrated impairment at baseline in immediate and delayed memory domains using Robust norms and in immediate memory using Conventional norms. However, those who declined after 5 years did not show baseline impairment in any domain using either norm method. Third, longitudinal change significantly added to the prediction of decline accuracies of the baseline scores in all groups compared to the HT group in the delayed memory and attention-language domains. Each of these conclusions is discussed in greater detail below.

Robust Norms Are More Sensitive Than Conventional Norms in Predicting Decline in NL Individuals

The present study demonstrates that Robust norms are more sensitive than Conventional norms at identifying normal individuals who will decline in the future. In the delayed memory and attention-language domains, Robust norms show a significant improvement over Conventional norms in predicting decline. Our findings add to the previous studies of Robust norms. To date Robust norms have been provided for memory tests (Marcopulos & McLain, 2003; Rentz et al., 2004; Ritchie et al., 2007; Sliwinski, Buschke, Stewart, Masur, & Lipton, 1997), language tests (Marcopulos & McLain, 2003; Rentz et al., 2004; Zec, Burkett, Markwell, & Larsen, 2007), measures of general cognition (Marcopulos & McLain, 2003), and executive function (Marcopulos & McLain, 2003) and separately for these domains in those with high IQ scores (Rentz et al., 2004). When used, Robust norms have been shown to reduce the variability in older individuals' cognitive scores (Ritchie et al., 2007; Sliwinski et al., 1996), and to be useful in the diagnosis of MCI (Manly et al., 2005; Marcopulos, Gripshover, Broshek, McLain, & Brashear, 1999; Sliwinski et al., 1996). Of the few previous studies that reported data for Robust norms, only one assessed the prediction of future cognitive decline. Marcopulos and McLain generated Robust norms in a rural population and showed that these norms can be used to predict decline. Decline was defined as a drop of 1 *SD* on three or four measures in the group of community dwelling individuals who passed a screening test but not a rigorous diagnostic evaluation. In the current study we show for the first time that Robust norms in comparison to Conventional norms improved prediction of clinically significant cognitive decline, defined as a change in diagnostic category, in individuals diagnosed as normal or with MCI.

At first glance the sensitivity of the Robust norms in the normal decline group seems low at 20% for delayed memory domain and 18% for attention-language domain. However, this is an expected result given that all our participants in this comparison had normal cognitive function at baseline. It has been shown that cognition remains relatively stable until the onset of dementia, but one study showed that at baseline, test completion time was predictive of decline to AD (Galvin et al., 2005). Others have demonstrated that with a diagnosis of dementia Conventional and Robust norms work equally well (Ritchie et al., 2007). Our data supported this finding in that when patients approach a clear-cut AD diagnosis, Conventional or Robust

norms are equally useful. Our study demonstrates the benefit of Robust norms in predicting decline to the prodromal and early stages of AD.

Immediate and Delayed Memory Are More Sensitive in Predicting Decline in NL Individuals Within 5 Years

The present study demonstrates that normal individuals who eventually decline to MCI or AD within 5 years have impaired immediate and delayed memory at baseline. The predictive value of memory performance at the MCI stage is well-documented in the AD literature (Artero, Tierney, Touchon, & Ritchie, 2003; Backman, Small, & Fratiglioni, 2001; Kluger et al., 1999; Tierney et al., 1996; Tierney et al., 2005). The present study adds to these findings by showing that memory impairment can be identified prior to the clinical transition to MCI. These findings support previous observations that longitudinally normal individuals who eventually develop MCI or AD show impaired delayed memory years prior to decline (Elias et al., 2000; Fox, Warrington, Seiffer, Agnew, & Rossor, 1998; Masur et al., 1994; Powell et al., 2006; Rubin et al., 1998). Robust and Conventional norms were not significantly different in terms of identifying those who show normal early decline. We believe that with greater statistical power from a larger sample, Robust norms will outperform Conventional norms for identifying future impairment in the delayed memory domain. Currently our group size is small ($n = 41$), but with 77 participants, 80% power is achieved in detecting differences in sensitivity of this domain.

Normal participants who eventually decline to MCI or AD in 5 or more years had no cognitive impairment at baseline. This finding is in contrast with results from studies that examined cases diagnosed as normal who declined to AD or other dementias, and found that at baseline, naming and DSST, two tests used within the language-attention domain in the current study, were impaired (Rubin et al., 1998). Similar to our findings, others did not find naming to be sensitive to early AD (Testa et al., 2004). In healthy aging on the other hand, language, as measured using naming, showed slight performance decline in persons over the age of 70 (Zec, Markwell, Burkett, & Larsen, 2005).

The finding that individuals with a normal diagnosis who decline within 5 years have cognitive impairment and those who decline at a later time point have no cognitive impairment prompts discussion about diagnostic criteria for normality. To date few studies have concentrated on operationalizing a definition of normality. Edwards, Lindquist, and Yaffe (2004) reported on a normal cohort of participants from an Alzheimer's disease Research Center followed for approximately 2 years. Although a higher proportion of those normal participants that declined to MCI or AD at follow-up had memory symptoms at baseline, this proportion did not differ statistically from those who remained normal (Edwards et al., 2004). Blacker and colleagues (2007) followed 107 normal individuals for approximately 5 years and found that poorer memory performance at baseline predicted decline to MCI in 5 years and that these problems were not reported by the participant or an informant during a clinical interview (Blacker et al., 2007). Clinical interviews are traditionally used to stage individuals and the two most popular instruments are the Clinical Dementia Rating Scale (CDR; Morris et al., 1997) and the GDS (Reisberg et al., 1982) described above. With the exception of the classification of Age Associated Memory Impairment (Crook et al., 1986), psychometric cut-off levels are not used in an operationalized definition of normality in older persons. Moreover, Petersen and colleagues (2001) suggested that although neuropsychological information can be helpful for diagnostic purposes, it cannot be used alone for an MCI or AD diagnosis. Rather, clinical judgment is also required. Petersen et al. reported that many factors (e.g., age, education, ethnic background, etc.) can affect test performance and that distinguishing different subtypes of MCI and dementia can be challenging due to overlapping neuropsychological scores. With this in mind, our findings and those of Blacker et al. (2007) suggest that it might be possible to

operationally define cognitive performance in normal individuals and to separate those likely to decline in the future from those who will remain stable. With a more careful operational definition of normality (i.e., Robust norms), it is possible that psychometric discrimination of cognitively impaired normal and MCI populations may be more accurately defined. Our data suggest that in normal individuals the lack of impairment in any domain cross-sectionally and the lack of change over time may indicate continued cognitive stability for at least 5 more years.

Longitudinal Change Adds to Baseline Prediction of Future Decline

We found that for both normal and MCI participants who decline, there is excessive change in the delayed memory and attention-language domains. This finding is consistent with the observation that older individuals who do not develop dementia have stable cognition, whereas those that do develop dementia fail to show a practice effects on repeated cognitive testing (Galvin et al., 2005).

Clinical methods for predicting which normal individuals will show cognitive decline in the future have not been well studied and require further attention. One major obstacle to identifying those at risk for future cognitive impairment is that there is a low incidence of decline and slow progression of normal older persons to dementia. For example, in a prior longitudinal study of more than 200 participants in our patient population, results over a mean interval on 3.8 years indicated that of 21 GDS = 1 normal participants, none declined to a dementia diagnosis. Of 105 GDS = 2 normal participants, approximately 4% per year declined to dementia (Kluger et al., 1999). However, when decline to MCI or dementia has been studied, GDS Stage 2 normal participants with subjective memory decline have been found to decline at a rate of approximately 7% per year (Prichep et al., 2006; Reisberg & Gauthier, in press). PET imaging studies have been shown to be predictive of decline from NL to MCI and AD (de Leon et al., 2001; Jagust et al., 2006; Mosconi et al., 2007). However, these techniques are invasive and expensive. Electroencephalographic changes have also been found to be useful in predicting cognitive decline in normal GDS Stage 2 participants with subjective impairment (Prichep et al., 2006). However, additional predictive methods are clearly necessary. Although impaired baseline cognitive performance among normal individuals has previously provided information about risk of AD (Masur et al., 1994), the current study shows that it is also informative for predicting future MCI. A cognitive predictor is very much needed to identify individuals at increased risk for MCI and to identify those who may be candidates for early intervention. As MCI is of interest in clinical trials, knowing who is destined to develop MCI is a first step in studying treatment of preclinical MCI.

Methodological Considerations

Robust norms, as conceptualized by Sliwinski et al., (1996) were generated by removing future decliners to AD from the normative group. Such Robust norms more accurately estimate the common effect of age on cognition in the absence of very early disease. We refined the methods used by Sliwinski et al. by retrospectively excluding individuals from the Robust normative group who eventually declined to MCI as well as those who eventually declined to AD. Manly et al. (2005) previously suggested that Robust norms may be useful in studies with MCI because such norms should be more sensitive in detecting MCI. Our data are the first to show that Robust norms are in fact more sensitive than Conventional norms for identifying which normal individuals decline to MCI or AD. MCI is often considered a prodromal stage of AD (Flicker et al., 1991; Morris et al., 2001; Petersen et al., 2001; Petersen et al., 1999; Reisberg, 1986), and these individuals show detectable cognitive change (Manly et al., 2005; Reisberg et al., 1988). Excluding normal participants who subsequently decline to MCI from the normative group resulted in a more conservative estimation (higher cut scores) of normal cognitive performance.

Our Robust normative group and the healthy test group had a higher representation of apoE 4 carriers compared to the NL→DEC group. This was surprising given that apoE 4 is a risk factor for AD. One reason to account for this disparity is that cognitively normal individuals who participate in research studies on AD tend to have a family member who had AD and therefore they have a higher probability of being apoE 4 carriers. The normal decline group in our study, however, did not reflect the normative population. When we separated the declining group into those that declined to MCI versus AD, we found that the NL-MCI group had proportions of E4 carriers that were not significantly different from the stable MCI group although the NL-AD participants were similar to the healthy test group and both normative groups. There was a lower proportion of NL→DEC who declined to AD (39%), and therefore the overall group resembled the stable MCI group. Some have reported that cognitive performance is lower in normal individuals who are apoE 4 carriers (Bondi et al., 1995; Flory, Manuck, Ferrell, Ryan, & Muldoon, 2000; O'Hara et al., 1998) however, Jorm et al. (2007) showed that between the ages of 20 to 65 apoE genotype did not influence cognition. Given that our normative groups had similar proportions of carriers and noncarriers we are confident that our results are not biased. We might in fact have underestimated the cognitive impairment in our NL→DEC group.

Given that the healthy test group was significantly younger than the other test groups, we reexamined our data limiting the age of all test groups to those above 65 years. Our baseline findings did not change. However, as expected, a higher proportion of normal decliners and stable MCI participants had impaired delayed memory and attention-language compared to the healthy test group. Robust norms outperformed Conventional norms in these domains. This was particularly important with respect to comparing the healthy test and NL→DEC groups as older age is a risk factor for decline (Gao, Hendrie, Hall, & Hui, 1998; Kukull et al., 2002; Ritchie & Kildea, 1995).

With the specificity of the Robust and Conventional norms equally constrained, the sensitivity of the Robust norms was significantly higher than for the Conventional norms for the normal decline and stable MCI groups. For longitudinal change over time, the results were also unaffected in that the normal early decline group showed excessive change in delayed memory and attention-language domains compared to the healthy test group. Therefore, we are confident that the age distribution of the healthy test group and the other test groups did not bias our results.

Our Conventional and Robust Normative samples were mutually exclusive of each other and exclusive of the test sample. Therefore, we could more independently assess the utility of the norms. We were able to test the efficiency of using Robust and Conventional norms in identifying cognitive impairment in normal, MCI and AD groups. Few previous studies have compared these two types of norms. In these few studies, the diagnostic groups either were not mutually exclusive (Ritchie et al., 2007; Sliwinski et al., 1996) or the study did not evaluate the utility of these norms in predicting decline from NL (Manly et al., 2005; Ritchie et al., 2007). The literature on Robust norms is growing and with increased availability of data from longitudinal normal cohorts, we suggest that Robust norms be generated for published neuropsychological tests so as to increase the sensitivity of identifying those at high risk for early decline.

Limitations of the Present Study

The current study must be examined in light of its limitations. The participant samples used in this study were not random population samples. All participants were volunteers participating in brain aging studies at a research center. The resulting samples were better educated, healthier, and ethnically less heterogeneous than the general population. Thus, we do not know if Robust norms would be equally useful when applied to the general population and those with less

education. We anticipate, however, that Robust norms would be applicable to the general population as the literature has shown the utility of these norms in a random sample of older participants and in those with less education (Manly et al., 2005; Sliwinski et al., 1996). Further, because we limited the study to normal, MCI, and AD participants, the generalizability of our findings to other dementing disorders is unknown. This study was specifically designed to examine the sensitivity of our cognitive measures to predict decline in longitudinally validated clinical groups so as to establish the utility of these norms for diagnostically homogenous normal, MCI, and AD populations. Future studies are warranted to explore the specificity of Robust norms for predicting progression to AD. Finally, with the exception of the immediate and delayed paragraph recall tests, the version of the measures used in this study are specific to our ADC and may not be commonly used by other centers or clinicians. However, our measures are, of course, available for use by other investigators, and they are similar to other tests of relevant cognitive domains that are commonly used. We acknowledge that many of these measures (e.g., digits forward and backward) have been updated. However, given the ongoing longitudinal research of our center, the older versions remain in use. Nevertheless, Robust norms outperform Conventional norms and we believe this generalization will hold true for diverse versions of these or similar tests.

Conclusions

The results of this study highlight the importance of using a longitudinally stable normal cohort to develop cognitive test norms. Not only are cross-sectional Robust norms more useful than Conventional cross-sectional norms but longitudinal norms provide additional value for predicting future cognitive performance. Some normal individuals who will decline to MCI or AD have baseline impairment in memory function and, depending on the interval between baseline and decline, this impairment may also include other cognitive domains. Together, baseline performance and longitudinal change provide unique information about which normal individuals will decline. Our findings should be useful for augmenting participant selection for research protocols.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported by NIH-NIA Grants AG12101, AG13616, AG08051 and AG022374. We thank the NYUSoM ADC Clinical Core and Datacore for their assistance in this work, especially Alok Vedvyas. Special thanks to Juan Li for her statistical assistance.

References

- Agresti, A. Categorical data analysis. Vol. 2. Hoboken, NJ: Wiley; 2002.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Vol. 4. Washington, DC: Author; 1994.
- Artero S, Tierney MC, Touchon J, Ritchie K. Prediction of transition from cognitive impairment to senile dementia: A prospective, longitudinal study. *Acta Psychiatrica Scandinavica* 2003;107:390–393. [PubMed: 12752036]
- Backman L, Small BJ, Fratiglioni L. Stability of the preclinical episodic memory deficit in Alzheimer's disease. *Brain* 2001;124:96–102. [PubMed: 11133790]
- Blackner D, Lee H, Muzikansky A, Martin EC, Tanzi R, McArdle JJ, et al. Neuropsychological measures in normal individuals that predict subsequent cognitive decline. *Archives of Neurology* 2007;64:862–871. [PubMed: 17562935]

- Bondi MW, Salmon DP, Monsch AU, Galasko D, Butters N, Klauber MR, et al. Episodic memory changes are associated with the APOE-epsilon 4 allele in nondemented older adults. *Neurology* 1995;45:2203–2206. [PubMed: 8848194]
- Cargin JW, Maruff P, Collie A, Masters C. Mild memory impairment in healthy older adults is distinct from normal aging. *Brain and Cognition* 2006;60:146–155. [PubMed: 16446021]
- Chen P, Ratcliff G, Belle SH, Cauley JA, DeKosky S, Ganguli M. Patterns of cognitive decline in presymptomatic Alzheimer disease. *Archives of General Psychiatry* 2001;58:853–858. [PubMed: 11545668]
- Convit A, de Leon MJ, Tarshish C, De Santi S, Tsui W, Rusinek H, et al. Specific hippocampal volume reductions in individuals at risk for Alzheimer's disease. *Neurobiology of Aging* 1997;18:131–138. [PubMed: 9258889]
- Crook T, Bartus RT, Ferris SH, Whitehouse P, Cohen GD, Gershon S. Age-associated memory impairment: Proposed diagnostic criteria and measures of clinical change—Report of a National Institute of Mental Health work group. *Developmental Neuropsychology* 1986;2:261–276.
- Crook T, Gilbert JG, Ferris SH. Operationalizing memory impairment in elderly persons: The Guild Memory Test. *Psychological Reports* 1980;47:1315–1318. [PubMed: 7220760]
- de Leon MJ, Convit A, Wolf OT, Tarshish CY, De Santi S, Rusinek H, et al. Prediction of cognitive decline in normal elderly subjects with 2-[18F]fluoro-2-deoxy-D-glucose/positron-emission tomography (FDG/PET). *Proceedings of the National Academy of Science* 2001;98:10966–10971.
- de Leon MJ, George AE, Golomb J, Tarshish C, Convit A, Kluger A, et al. Frequency of hippocampal formation atrophy in normal aging and Alzheimer's disease. *Neurobiology of Aging* 1997;18:1–11. [PubMed: 8983027]
- de Leon MJ, Golomb J, George AE, Convit A, Tarshish CY, McRae T, et al. The radiologic prediction of Alzheimer's disease: The atrophic hippocampal formation. *American Journal of Neuroradiology* 1993;14:897–906. [PubMed: 8352162]
- De Santi S, de Leon MJ, Rusinek H, Convit A, Tarshish CY, Boppana M, et al. Hippocampal formation glucose metabolism and volume losses in MCI and AD. *Neurobiology of Aging* 2001;22:529–539. [PubMed: 11445252]
- Edwards ER, Lindquist K, Yaffe K. Clinical profile and course of cognitively normal patients evaluated in memory disorders clinics. *Neurology* 2004;62:1639–1642. [PubMed: 15136703]
- Elias MF, Beiser A, Wolf PA, Au R, White RF, D'Agostino RB. The preclinical phase of Alzheimer disease: A 22-year prospective study of the Framingham cohort. *Archives of Neurology* 2000;57:808–813. [PubMed: 10867777]
- Fastenau PS, Denburg NL, Hufford BJ. Adult norms for the Rey-Osterrieth Complex Figure Test and for supplemental recognition and matching trials from the Extended Complex Figure Test. *The Clinical Neuropsychologist* 1999;13:30–47. [PubMed: 10937646]
- Fleisher AS, Sowell BB, Taylor C, Gamst AC, Petersen RC, Thal LJ. Clinical predictors of progression to Alzheimer disease in amnesic mild cognitive impairment. *Neurology* 2007;68:1588–1595. [PubMed: 17287448]
- Flicker C, Ferris SH, Reisberg B. Mild cognitive impairment in the elderly: Predictors of dementia. *Neurology* 1991;41:1006–1009. [PubMed: 2067629]
- Flicker C, Ferris SH, Reisberg B. A two-year longitudinal study of cognitive function in normal aging and Alzheimer's disease. *Journal of Geriatric Psychiatry and Neurology* 1993;6:84–96. [PubMed: 8512635]
- Flory JD, Manuck SB, Ferrell RE, Ryan CM, Muldoon MF. Memory performance and the apolipoprotein E polymorphism in a community sample of middle-aged adults. *American Journal of Medical Genetics* 2000;96:707–711. [PubMed: 11121165]
- Fox NC, Warrington EK, Seiffer AL, Agnew SK, Rossor MN. Presymptomatic cognitive deficits in individuals at risk of familial Alzheimer's disease. A longitudinal prospective study. *Brain* 1998;121:1631–1639. [PubMed: 9762953]
- Galvin JE, Powlishta KK, Wilkins K, McKeel DW Jr, Xiong C, Grant E, et al. Predictors of preclinical Alzheimer disease and dementia: A clinicopathologic study. *Archives of Neurology* 2005;62:758–765. [PubMed: 15883263]

- Gao S, Hendrie HC, Hall KS, Hui S. The relationship between age, sex, and the incidence of dementia and Alzheimer disease: A meta-analysis. *Archives of General Psychiatry* 1998;55:809–815. [PubMed: 9736007]
- Gilbert, JG. Guild Memory Test manual. Newark, NJ: UNICO National Mental Health Research Center; 1970.
- Gilbert JG, Levee RF, Catalano FL. A preliminary report on a new memory scale. *Perceptual and Motor Skills* 1968;27:277–278. [PubMed: 5685703]
- Glodzik-Sobanska L, Reisberg B, De Santi S, Babb JS, Pirraglia E, Rich KE, et al. Subjective memory complaints: Presence, severity and future outcome in normal older subjects. *Dementia and Geriatric Cognitive Disorders* 2007;24:177–184. [PubMed: 17641528]
- Hachinski VC, Iliff LD, Zilhka E, Du Boulay GH, McAllister VL, Marshall J, et al. Cerebral blood flow in dementia. *Archives of Neurology* 1975;32:632–637. [PubMed: 1164215]
- Hamilton M. A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry* 1960;23:56–62.
- Heaton RK, Avitable N, Grant I, Matthews CG. Further crossvalidation of regression-based neuropsychological norms with an update for the Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology* 1999;21:572–582. [PubMed: 10550815]
- Jagust W, Gitcho A, Sun F, Kuczynski B, Mungas D, Haan M. Brain imaging evidence of preclinical Alzheimer's disease in normal aging. *Annals of Neurology* 2006;59:673–681. [PubMed: 16470518]
- Jorm AF, Mather KA, Butterworth P, Anstey KJ, Christensen H, Eastel S. *APOE* genotype and cognitive functioning in a large age-stratified population sample. *Neuropsychology* 2007;21:1–8. [PubMed: 17201525]
- Kendall PC, Marrs-Garcia A, Nath SR, Sheldrick RC. Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology* 1999;67:285–299. [PubMed: 10369049]
- Kendall PC, Sheldrick RC. Normative data for normative comparisons. *Journal of Consulting and Clinical Psychology* 2000;68:767–773. [PubMed: 11068962]
- Kluger A, Ferris SH, Golomb J, Mittelman MS, Reisberg B. Neuropsychological prediction of decline to dementia in nondemented elderly. *Journal of Geriatric Psychiatry and Neurology* 1999;12:168–179. [PubMed: 10616864]
- Kukull WA, Higdon R, Bowen JD, McCormick WC, Teri L, Schellenberg GD, et al. Dementia and Alzheimer disease incidence: A prospective cohort study. *Archives of Neurology* 2002;59:1737–1746. [PubMed: 12433261]
- Levine AJ, Miller EN, Becker JT, Selnes OA, Cohen BA. Normative data for determining significance of test-retest differences on eight common neuropsychological instruments. *The Clinical Neuropsychologist* 2004;18:373–384. [PubMed: 15739809]
- Manly JJ, Bell-McGinty S, Tang MX, Schupf N, Stern Y, Mayeux R. Implementing diagnostic criteria and estimating frequency of mild cognitive impairment in an urban community. *Archives of Neurology* 2005;62:1739–1746. [PubMed: 16286549]
- Marcopulos BA, Gripshover DL, Broshek DK, McLain CA, Brashear HR. Neuropsychological assessment of psychogeriatric patients with limited education. *The Clinical Neuropsychologist* 1999;13:147–156. [PubMed: 10949156]
- Marcopulos BA, McLain CA. Are our norms “normal”? A 4-year follow-up study of a biracial sample of rural elders with low education. *The Clinical Neuropsychologist* 2003;17:19–33. [PubMed: 12854008]
- Marquis S, Moore MM, Howieson DB, Sexton G, Payami H, Kaye JA, et al. Independent predictors of cognitive decline in healthy elderly persons. *Archives of Neurology* 2002;59:601–606. [PubMed: 11939895]
- Masur DM, Sliwinski M, Lipton RB, Blau AD, Crystal HA. Neuropsychological prediction of dementia and the absence of dementia in healthy elderly persons. *Neurology* 1994;44:1427–1432. [PubMed: 8058143]
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department

- of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;34:939–944. [PubMed: 6610841]
- McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947;12:153–157.
- Meguro K, Shimada M, Yamaguchi S, Ishizaki J, Yamadori A, Sekita Y. A 5-year retrospective examination of cognitive screening test stages in normal older adults and patients with Alzheimer's disease: The Tajiri project. *Journal of Gerontology: Psychological Sciences* 2001;56B:P314–P318.
- Moran LJ, Mefferd RB. Repetitive psychometric measures. *Psychological Reports* 1959;5:269–275.
- Morris JC, Ernesto C, Shaefer K, Coats M, Leon S, Sano M, et al. Clinical dementia rating (CDR) training and reliability protocol: The Alzheimer Disease Cooperative Study Unit experience. *Neurology* 1997;48:1508–1510. [PubMed: 9191756]
- Morris JC, Storandt M, Miller P, McKeel DW, Price JL, Rubin EH, et al. Mild cognitive impairment represents early-stage Alzheimer disease. *Archives of Neurology* 2001;58:397–405. [PubMed: 11255443]
- Mosconi L, De Santi S, Li J, Tsui WH, Li Y, Boppana M, Rusinek H, de Leon MJ. Hippocampal hypometabolism predicts cognitive decline from normal aging. *Neurobiology of Aging* 2008;29:676–692. [PubMed: 17222480]
- Norman MA, Evans JD, Miller SW, Heaton RK. Demographically corrected norms for the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology* 2000;22:80–94. [PubMed: 10649547]
- O'Hara R, Yesavage JA, Kraemer HC, Mauricio M, Friedman LF, Murphy GM Jr. The APOE epsilon4 allele is associated with decline on delayed recall performance in community-dwelling older adults. *Journal of American Geriatric Society* 1998;46:1578–1579.
- Palmer BW, Boone KB, Lesser IM, Wohl MA. Base rates of "impaired" neuropsychological test results among healthy older adults. *Archives of Clinical Neuropsychology* 1998;13:503–511. [PubMed: 14590634]
- Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, et al. Current concepts in mild cognitive impairment. *Archives of Neurology* 2001;58:1985–1992. [PubMed: 11735772]
- 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]
- Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris SH, et al. Vitamin E and donepezil for the treatment of mild cognitive impairment. *The New England Journal of Medicine* 2005;353:2305–2315.
- Powell MR, Smith GE, Knopman DS, Parisi JE, Boeve BF, Petersen RC, et al. Cognitive measures predict pathologic Alzheimer disease. *Archives of Neurology* 2006;63:865–868. [PubMed: 16769868]
- Prichep LS, John ER, Ferris SH, Rausch L, Fang Z, Cancro R, et al. Prediction of longitudinal cognitive decline in normal elderly with subjective complaints using electrophysiological imaging. *Neurobiology of Aging* 2006;27:471–481. [PubMed: 16213630]
- Reisberg B. Dementia: A systematic approach to identifying reversible causes. *Geriatrics* 1986;41:30–46. [PubMed: 3949165]
- Reisberg B, Ferris SH. The Brief Cognitive Rating Scale (BCRS). *Psychopharmacology Bulletin* 1988;24:629–636. [PubMed: 3249764]
- Reisberg B, Ferris SH, de Leon MJ, Crook T. The global deterioration scale for assessment of primary degenerative dementia. *American Journal of Psychiatry* 1982;139:1136–1139. [PubMed: 7114305]
- Reisberg B, Ferris SH, de Leon MJ, Sinaiko E, Franssen EH, Kluger A, et al. Stage-specific behavioral, cognitive, and in vivo changes in community residing subjects with age-associated memory impairment (AAMI) and primary degenerative dementia of the Alzheimer type. *Drug Development Research* 1988;15:101–114.
- Reisberg, B.; Ferris, SH.; Steinberg, G.; Shulman, E.; de Leon, MJ.; Sinaiko, E. Longitudinal study of dementia patients and aged controls: An approach to methodologic issues. In: Lawton, MP.; Herzog, AR., editors. *Special research methods for gerontology*. Amityville, NY: Baywood; 1989. p. 195–231.

- Reisberg B, Gauthier S. Current evidence for subjective cognitive impairment (SCI) as a pre-mild cognitive impairment (MCI) stage of subsequently manifest Alzheimer's disease. *International Psychogeriatrics*. in press
- Reisberg B, Sclan SG, Franssen EH, de Leon MJ, Kluger A, Torossian CL, et al. Clinical stages of normal aging and Alzheimer's disease: The GDS staging system. *Neuroscience Research Communications* 1993;13(Suppl 1):551–554.
- Rentz DM, Huh TJ, Faust RR, Budson AE, Scinto LFM, Sperling RA, et al. Use of IQ-adjusted norms to predict progressive cognitive decline in highly intelligent older individuals. *Neuropsychology* 2004;18:38–49. [PubMed: 14744186]
- Ritchie K, Kildea D. Is senile dementia “age-related” or “ageing-related”? Evidence from meta-analysis of dementia prevalence in the oldest old. *Lancet* 1995;335:931–934. [PubMed: 7564727]
- Ritchie LJ, Frerichs RJ, Tuokko H. Effective normative samples for the detection of cognitive impairment in older adults. *The Clinical Neuropsychologist* 2007;21:863–874. [PubMed: 17853155]
- Rosen WG, Terry RD, Fuld PA, Katzman R, Peck A. Pathological verification of ischemia score in differentiation of dementias. *Annals of Neurology* 1980;7:486–488. [PubMed: 7396427]
- Rubin EH, Storandt M, Miller P, Kinschler DA, Grant EA, Morris JC, et al. A prospective study of cognitive function and onset of dementia in cognitively healthy elders. *Archives of Neurology* 1998;55:395–401. [PubMed: 9520014]
- Salloway S, Ferris S, Kluger A, Goldman R, Griesing T, Kumar D, et al. Efficacy of donepezil in mild cognitive impairment: A randomized placebo-controlled trial. *Neurology* 2004;63:651–657. [PubMed: 15326237]
- Sheridan LK, Fitzgerald HE, Adams KM, Puttler LI, Wong MM, Zucker RA. Normative Symbol Digit Modalities Test performance in a community-based sample. *Archives of Clinical Neuropsychology* 2006;21:23–28. [PubMed: 16139470]
- Sliwinski M, Buschke H, Stewart WF, Masur D, Lipton RB. The effect of dementia risk factors on comparative and diagnostic selective reminding norms. *Journal of the International Neuropsychological Society* 1997;3:317–326. [PubMed: 9260441]
- Sliwinski M, Lipton RB, Buschke H, Stewart W. The effects of preclinical dementia on estimates of normal cognitive functioning in aging. *Journal of Gerontology: Psychological Sciences* 1996;51B:P217–P225.
- Tabert MH, Manly JJ, Liu X, Pelton GH, Rosenblum S, Jacobs M, et al. Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. *Archives of General Psychiatry* 2006;63:916–924. [PubMed: 16894068]
- Testa JA, Ivnik RJ, Boeve B, Petersen RC, Pankratz VS, Knopman D, et al. Confrontation naming does not add incremental diagnostic utility in MCI and Alzheimer's disease. *Journal of the International Neuropsychological Society* 2004;10:504–512. [PubMed: 15327729]
- Tierney MC, Szalai JP, Snow WG, Fisher RH, Nores A, Nadon G, et al. Prediction of probable Alzheimer's disease in memory-impaired patients: A prospective longitudinal study. *Neurology* 1996;46:661–665. [PubMed: 8618663]
- Tierney MC, Yao C, Kiss A, McDowell I. Neuropsychological tests accurately predict incident Alzheimer's disease after 5 and 10 years. *Neurology* 2005;64:1853–1859. [PubMed: 15955933]
- Wechsler, D. Wechsler Adult Intelligence Scale. New York: Psychological Corporation; 1955.
- Wechsler, D. Wechsler Adult Intelligence Scale–Revised. New York: Harcourt Brace Jovanovich; 1981.
- Zec RF, Burkett NR, Markwell SJ, Larsen DL. Normative data stratified for age, education, and gender on the Boston Naming Test. *The Clinical Neuropsychologist* 2007;21:617–637. [PubMed: 17613981]
- Zec RF, Markwell SJ, Burkett NR, Larsen DL. A longitudinal study of confrontation naming in the “normal” elderly. *Journal of the International Neuropsychological Society* 2005;11:716–726. [PubMed: 16248907]

GROUP	BASELINE	1 st FOLLOW-UP	LAST FOLLOW-UP
Conventional Norm	NL	-	-
Robust Norm	NL	NL	NL
Healthy Test	NL	NL	-
Normal Decline { <div style="display: inline-block; vertical-align: middle; margin-left: 5px;"> Late Early </div>	NL	NL	MCI (58%) AD (42%)
	NL	NL (22%) MCI (78%)	MCI (63%) AD (37%)
Stable MCI	MCI	MCI	MCI
MCI Decline	MCI	MCI (27%) AD (73%)	AD
AD – AD	AD	AD	AD

NL = cognitively normal, MCI = Mild Cognitive Impairment, AD = Alzheimer's Disease

Figure 1.
Study groups with diagnoses at three time points.

Figure 2.

Impairment in healthy test and NL→DEC for all domains using Conventional and Robust norms.

Figure 3.
 Impairment in the delayed memory domain using Conventional and Robust norms.

Table 1

Baseline Demographics for Normative Groups and Healthy Test Group

	Normative groups		Test group
	Conventional	Robust	HT
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Group size	256	113	139
Age at baseline	62.7 (11.5)	62.8 (9.8)	62.6 (10.7)
Education	16.3 (2.3)	16.5 (2.2)	16.2 (2.0)
MMSE	29.3 (1.1)	29.4 (1.1)	29.4 (1.0)
Time stable	N/A	8.3 (4.0)	3.3 (2.3)
Time to decline	N/A	N/A	N/A
% women	65	68	63
% White	92	95	89
% ApoE: % e4+	36	40	39

Note. There were no significant differences between the groups for any demographic variable with the exception of time stable. By design participants in the healthy test (HT) group had two visits and therefore less follow-up time than the Robust Norm group where participants had three or more visits. MMSE = Mini Mental Status Examination Score; N/A = not applicable; ApoE4 = apolipoprotein E, e4+ = carrier of e4 allele.

Table 2
Baseline Demographics for NL→ Decline Subgroups

	NL→DEC	
	<i>M (SD)</i>	<i>M (SD)</i>
	Early	Late
Group size	41	43
Age at baseline	72.8 (7.4) *	69.5 (6.0) *
Education	15.7 (2.3)	15.6 (2.2)
MMSE	29.1 (1.3)	28.9 (1.3)
Time to decline	2.8 (0.9)	9.7 (3.6)
% women	68	53
% White	90	95
% ApoE: % e4 carriers	19 *	19 *

Note. NL = diagnostically normal; MMSE = Mini Mental Status Examination score; ApoE = apolipoprotein E, e4+ = carrier of e4 allele.

* Significantly different from HT: $p < .05$.

Table 3

Baseline Demographics for Test Groups

	HT	NL→DEC	Stable MCI	MCI→AD	AD→AD
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Group size	139	84	92	44	162
Age at baseline	62.6 (10.7)	71.1 (6.9) [*]	74.7 (8.6) [*]	74.9 (9.1) [*]	73.1 (8.3) [*]
Education	16.2 (2.0)	15.7 (2.3)	15.6 (2.3)	15.5 (2.5)	14.7 (2.4) [†]
MMSE	29.4 (1.0)	29.0 (1.3)	28.1 (1.4) [*]	27.7 (1.8) [*]	21.9 (4.5) [†]
Time stable	3.3 (2.3)	N/A	3.4 (2.2)	N/A	5.1 (3.4)
Time to decline	N/A	6.4 (4.4)	N/A	3.1 (2.0)	N/A
% women	63	61	52	61	60
% White	89	93	98 [*]	98	91
% ApoE: % e4+	39	19 [*]	27	57	60 [*]

Note. HT = healthy test group with two evaluations; NL→DEC = normal decline group; Stable MCI = mild cognitive impairment at all time points; MCI→AD = MCI decline group; AD→AD = stable Alzheimer's disease group; MMSE = Mini Mental Status Examination score; N/A = not applicable; ApoE4 = apolipoprotein E, e4+ = carrier of e4 allele.

[†] Significantly different from HT, NL→DEC, and Stable MCI: $p < .05$.

^{*} Significantly different from HT: $p < .05$.

Table 4

Baseline Cognitive Performance for Normative Groups and Healthy Test Group

Domain	Normative groups		Test group
	Conventional	Robust	HT
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Tests			
Working Memory			
Digits Forward	7.0 (1.2)	7.2 (1.1)	7.1 (1.2)
Digits Backward	5.5 (1.4)	5.7 (1.3)	5.6 (1.4)
Immediate Memory			
Paragraph Recall	7.2 (2.6)	7.5 (2.5)	7.9 (2.7)
Verbal Paired Associates	5.1 (2.5)	5.2 (2.3)	5.1 (2.5)
Visual Paired Associates	6.3 (2.6)	6.2 (2.4)	6.1 (2.3)
Delayed Memory			
Paragraph Recall	8.8 (3.4)	9.6 (2.8)	9.4 (2.8)
Verbal Paired Associates	5.6 (2.8)	5.8 (2.4)	5.5 (2.5)
Word List Recall	8.7 (1.7)	9.0 (1.3)	8.6 (1.7)
Attention-Language			
Digit Symbol			
Substitution	56 (12)	58 (11)	57 (11)
Cancellation Test	75 (14)	76 (13)	76 (13)
Object Naming Test	55 (6)	56 (4)	55 (4)

Note. HT = healthy test group with two evaluations.

Table 5

Baseline Cognitive Performance for Test Groups

Domain	HT	NL→DEC	Stable MCI	MCI→AD	AD-AD
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Tests					
Working Memory					
Digits Forward	7.1 (1.2)	6.9 (1.2)	6.8 (1.2)	6.4 (1.3)	6.0 (1.3)
Digits Backward	5.6 (1.4)	5.4 (1.3)	5.0 (1.2)	5.0 (1.3)	3.8 (1.3)
Immediate Memory					
Paragraph Recall	7.9 (2.7)	7.4 (3.2)	5.8 (2.2)	4.2 (2.1)	2.2 (1.8)
Verbal Paired Associates	5.1 (2.5)	3.8 (2.1)	3.7 (2.1)	2.0 (1.8)	0.6 (1.1)
Visual Paired Associates	6.1 (2.3)	5.0 (2.1)	4.5 (2.3)	3.3 (1.7)	1.5 (1.6)
Delayed Memory					
Paragraph Recall	9.4 (2.8)	9.0 (3.8)	6.5 (2.5)	3.6 (2.5)	1.2 (1.9)
Verbal Paired Associates	5.5 (2.5)	3.8 (2.5)	4.0 (2.6)	2.3 (1.9)	0.6 (1.1)
Word List Recall	8.6 (1.7)	7.6 (2.7)	6.6 (2.2)	4.8 (3.2)	1.9 (2.3)
Attention-Language					
Digit Symbol					
Substitution	57 (11)	48 (11)	44 (11)	37 (13)	24 (15)
Cancellation Test	76 (13)	67 (11)	64 (14)	62 (19)	40 (22)
Object Naming Test	55 (4)	54 (5)	51 (9)	46 (10)	39 (13)

Note. HT = healthy test group with two evaluations; NL→DEC = normal decline group; Stable MCI = mild cognitive impairment at all time points; MCI→AD = MCI decline group; AD-AD = stable Alzheimer's disease group.

Table 6
Z Score Cutoffs and Criterion for Defining Impairment

Domain	Baseline		Change
	Conventional	Robust	Robust
Tests			
Working Memory	Criterion: 2 of 2 tests		Criterion: 1 of 2 tests
Digits Forward	-0.75	-1.75	-1.5
Digits Backward	-0.75	-1.75	-2.5
Immediate Memory	Criterion: 2 of 3 tests		Criterion: 1 of 3 tests
Paragraph Recall	-1.00	-1.00	-2.00
Verbal Paired Associates	-1.00	-1.00	-2.5
Visual Paired Associates	-1.00	-1.5	-2.00
Delayed Memory	Criterion: 2 of 3 tests		Criterion: 1 of 3 tests
Paragraph Recall	-0.75	-1.00	-1.5
Verbal Paired Associates	-1.25	-1.5	-4.00
List Recall	-0.75	-1.00	-1.5
Attention-Language	Criterion: 2 of 3 tests		Criterion: 1 of 3 tests
Digit Symbol			
Substitution	-1.00	-1.00	-1.5
Cancellation Test	-1.00	-1.00	-3.00
Object Naming Test	-1.00	-2.00	-3.5

Table 7
Baseline McNemar Results: Specificity and Sensitivity Using Conventional and Robust Norms for Diagnostically NL Groups

	Specificity		Sensitivity		
	HT	NL→DEC	→	Early	Late
Working Memory					
Conventional	91%	8%	→	7%	9%
Robust	91%	18%	→	22%	14%
Immediate Memory					
Conventional	94%	17%	→	22%*	12%
Robust	94%	13%	→	24%*	2%
Delayed Memory					
Conventional	93%	10%	→	15%	5%
Robust	93%	20%*+	→	32%*+	9%
Attention-Language					
Conventional	94%	7%	→	7%	7%
Robust	94%	18%*+	→	17%	19%

Note. HT = healthy test group; NL→DEC = normal decline group; early = normal baseline evaluation with early decline to mild cognitive impairment or Alzheimer's disease (less than 5 years after baseline); late = normal baseline evaluation with late decline to mild cognitive impairment or Alzheimer's disease (5 years or more after baseline).

+ Significantly different from Sensitivity using Conventional norms within group, $p < .05$.

* Significantly higher proportion impaired than HT group, $p < .01$.

Table 8

Baseline Results: Sensitivity Using Conventional and Robust Norms for MCI and AD Diagnostic Groups

	Stable MCI	MCI→AD	AD-AD
Working Memory			
Conventional	9%	14%	31% [*]
Robust	18%	27% [*]	35% [*]
Immediate Memory			
Conventional	15%	50% [*]	89% [*]
Robust	11%	45% [*]	89% [*]
Delayed Memory			
Conventional	26% [*]	66% [*]	93% [*]
Robust	34% ^{*+}	70% [*]	93% [*]
Attention-Language			
Conventional	22% [*]	34% ^{*+}	74% [*]
Robust	32% ^{*+}	41% [*]	78% [*]

Note. MCI = mild cognitive impairment; AD = Alzheimer's disease; Stable MCI = mild cognitive impairment at all time points; MCI→AD = MCI decline group; AD-AD = Stable Alzheimer's disease group.

⁺ Significantly different from Sensitivity using Conventional norms (within group analysis), $p < .05$.

^{*} Significantly higher proportion impaired than HT group, $p < .01$.

Table 9

First Follow-Up Demographics for Normal Groups

	HT	NL-NL→DEC	→	Early decline	Late decline
	<i>M (SD)</i>	<i>M (SD)</i>		<i>M (SD)</i>	<i>M (SD)</i>
Group size	139	52	→	28	24
Age at first follow-up	64.9 (10.8)	73.4 (6.1)*	→	74.3 (6.4)*	72.3 (5.7)*
Education	16.2 (2.0)	15.6 (2.2)	→	15.5 (2.4)	15.7 (2.0)
MMSE	29.4 (1.0)	29.1 (1.2)	→	29.1 (1.1)	28.9 (1.4)
Time stable	3.3 (2.3)	N/A	→	N/A	N/A
Time from baseline	3.3 (2.3)	2.9 (1.2)	→	2.8 (1.3)	2.9 (1.1)
Time to decline	N/A	5.6 (3.8)	→	2.8 (1.2)	8.8 (3.1)
% women	63	60	→	64	54
% White	89	94	→	89	100
ApoE: % e4 +	39	20	→	23	16

Note. HT = healthy test group; NL-NL→DEC = normal baseline and normal first follow-up with subsequent decline; NL-NLDEC = and normal first follow-up evaluation with early decline to mild cognitive impairment or Alzheimer's disease (less than 5 years from first follow-up); NL-NLDEC = normal baseline and first follow-up evaluation with late decline to mild cognitive impairment or Alzheimer's disease (5 years or more after first follow-up); MMSE = Mini Mental Status Examination score; N/A = not applicable; ApoE4 = apolipoprotein E, e4+ = carrier of e4 allele.

* Significantly different from HT; $p < .05$.

Table 10

First Follow-Up Demographics for MCI Groups Compared to HT

	HT	Stable MCI	MCI-MCI→AD
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Group size	139	92	12
Age at first follow-up	64.9 (10.8)	76.6 (8.4) *	77.7 (10.1) *
Education	16.2 (2.0)	15.6 (2.3)	15.3 (2.6)
MMSE	29.4 (1.0)	28.1 (1.4) *	27.7 (2.0) *
Time stable	3.3 (2.3)	3.4 (2.2)	N/A
Time from baseline	3.3 (2.3)	2.0 (0.4)	2.2 (0.8)
Time to decline	N/A	N/A	3.1 (1.6)
% women	63	52	67
% White	89	98 *	92
ApoE: % e4 +	39	27	50

Note. HT = healthy test group; Stable MCI = mild cognitive impairment at all time points; MCI-MCI→AD = mild cognitive impairment at baseline and at first follow-up with subsequent decline; MMSE = Mini Mental Status Examination score; N/A = not applicable; ApoE4 = apolipoprotein E, e4+ = carrier of e4 allele.

* Significantly different from HT: $p < .05$.

Table 11
Results for the Addition of Longitudinal Change to Baseline Performance

	Immediate memory		Delayed memory		Attention-language	
	χ^2 Increment	Sensitivity	χ^2 Increment	Sensitivity	χ^2 Increment	Sensitivity
NL-NL→DEC	0.8	19%	4.8*	33%	6.9*	27%
NL-NL _L DEC	0.5	8%	0.2	13%	0.2	25%
NL-NL _E DEC	2.8	29%	7.5*	46%	10.2*	29%
Stable MCI	6.8*	34%	12.2*	53%	6.9*	40%
MCI-MCI→AD	4.2*	58%	9.8*	83%	13.1*	58%

Note. NL-NL→DEC = normal baseline and first follow-up with subsequent decline; NL-NL_EDEC = normal baseline and normal first follow-up evaluation with early decline to mild cognitive impairment (MCI) or Alzheimer's disease (AD); less than 5 years after first follow-up; NL-NL_LDEC = normal baseline and normal first follow-up evaluation with late decline to MCI or AD (5 years or more after first follow-up); Stable MCI = MCI at all time points; MCI-MCI→AD = MCI at baseline and first follow-up with subsequent decline to Alzheimer's disease.

* Significant χ^2 increment of addition to baseline model; $p < .05$.