Neuropsychological Criteria for Mild Cognitive Impairment Improves Diagnostic Precision, Biomarker Associations, and Progression Rates

Mark W. Bondi\textsuperscript{a,b,}, Emily C. Edmonds\textsuperscript{b}, Amy J. Jak\textsuperscript{a,b}, Lindsay R. Clark\textsuperscript{d}, Lisa Delano-Wood\textsuperscript{a,b}, Carrie R. McDonald\textsuperscript{b}, Daniel A. Nation\textsuperscript{e}, David J. Libon\textsuperscript{f}, Rhoda Au\textsuperscript{g}, Douglas Galasko\textsuperscript{a,c}, and David P. Salmon\textsuperscript{c} for the Alzheimer’s Disease Neuroimaging Initiative\textsuperscript{1}

\textsuperscript{a}Veterans Affairs San Diego Healthcare System, San Diego, CA, USA
\textsuperscript{b}Department of Psychiatry, University of California San Diego, School of Medicine, La Jolla, CA, USA
\textsuperscript{c}Department of Neurosciences, University of California San Diego, School of Medicine, La Jolla, CA, USA
\textsuperscript{d}San Diego State University/University of California San Diego Joint Doctoral Program in Clinical Psychology, San Diego, CA, USA
\textsuperscript{e}Department of Psychology, University of Southern California, Los Angeles, CA, USA
\textsuperscript{f}Department of Neurology, Drexel University, College of Medicine, Philadelphia, PA, USA
\textsuperscript{g}Department of Neurology and the Framingham Heart Study, Boston University, School of Medicine, Boston, MA, USA

Abstract

We compared two methods of diagnosing mild cognitive impairment (MCI): conventional Petersen/Winblad criteria as operationalized by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) and an actuarial neuropsychological method put forward by Jak and Bondi designed to balance sensitivity and reliability. 1,150 ADNI participants were diagnosed at baseline as cognitively normal (CN) or MCI via ADNI criteria (MCI: \( n = 846 \); CN: \( n = 304 \)) or Jak/Bondi criteria (MCI: \( n = 401 \); CN: \( n = 749 \)), and the two MCI samples were submitted to cluster and discriminant function analyses. Resulting cluster groups were then compared and further examined for APOE allelic frequencies, cerebrospinal fluid (CSF) Alzheimer’s disease (AD) biomarker levels, and clinical outcomes. Results revealed that both criteria produced a mildly...
impaired Amnestic subtype and a more severely impaired Dysexecutive/Mixed subtype. The neuropsychological Jak/Bondi criteria uniquely yielded a third Impaired Language subtype, whereas conventional Petersen/Winblad ADNI criteria produced a third subtype comprising nearly one-third of the sample that performed within normal limits across the cognitive measures, suggesting this method’s susceptibility to false positive diagnoses. MCI participants diagnosed via neuropsychological criteria yielded dissociable cognitive phenotypes, significant CSF AD biomarker associations, more stable diagnoses, and identified greater percentages of participants who progressed to dementia than conventional MCI diagnostic criteria. Importantly, the actuarial neuropsychological method did not produce a subtype that performed within normal limits on the cognitive testing, unlike the conventional diagnostic method. Findings support the need for refinement of MCI diagnoses to incorporate more comprehensive neuropsychological methods, with resulting gains in empirical characterization of specific cognitive phenotypes, biomarker associations, stability of diagnoses, and prediction of progression. Refinement of MCI diagnostic methods may also yield gains in biomarker and clinical trial study findings because of improvements in sample compositions of ‘true positive’ cases and removal of ‘false positive’ cases.

Keywords
Alzheimer’s disease; Alzheimer’s Disease Neuroimaging Initiative; biomarker; cluster analysis; dementia; mild cognitive impairment; neuropsychology; progression

INTRODUCTION
An explosion of research on genetic, imaging, and biomarker correlates of Alzheimer’s disease (AD) has occurred in the three decades since the publication of McKhann et al.’s [1] criteria for the diagnosis of AD. Over the years, this research has gravitated toward characterizing the period of mild cognitive impairment (MCI) that represents the borderland between normal aging and dementia [2–5]. The impact of this research is evident in recently revised criteria for AD [6], MCI [7], and “preclinical” AD [8] that increasingly rely upon biomarkers for disease detection, diagnosis, and prediction of clinical outcome. Unfortunately, this increased sophistication in the application of genetics and biomarkers to the study of mild forms of cognitive impairment has not been met with concomitant sophistication in profiling cognition. The revised criteria for MCI still largely rely on simple cognitive screening measures, clinical judgment, and limited neuropsychological assessment utilizing a ‘one test equals one domain’ methodology.

This approach to diagnosing MCI is epitomized in several clinical trials targeting MCI (e.g., [9]) and in the Alzheimer’s Disease Neuroimaging Initiative (ADNI; http://www.adni-info.org), which largely identify the amnestic form of MCI based on impaired performance on delayed recall of a single prose passage [10]. This methodology leads to several problems. First, shortening a test and detaching it from its standardized administration, scoring, and normative referencing may make it less sensitive or reliable. Second, story recall may be less sensitive to MCI or an evolving dementia than tests of verbal list learning [11–16] or perhaps visual memory [17]. Third, the use of a single memory measure violates
the statistical maxim that multiple measures provide a more reliable estimate of a cognitive construct than any single measure [18]. Finally, the lack of objective tests of other cognitive domains reduces the ability to detect cognitive impairment profiles that might identify distinct subtypes of MCI that vary in clinical and biological characteristics [19–21].

Several studies have used a full range of neuropsychological test measures with actuarial decision-making to improve diagnostic rigor for MCI. Saxton et al. [22] showed that a neuropsychological test-based algorithm for MCI diagnosis produced fewer ‘false positive’ diagnostic errors and provided better prediction of progression than classification based on the Clinical Dementia Rating (CDR), a structured clinical interview method that stages decline across the spectrum of AD. This result is consistent with the susceptibility of clinical judgment to biases and faulty assumptions that can lead to diagnostic errors [23]. Jak and Bondi and colleagues [24, 25] have found that the percentage of older adults classified as MCI ranges greatly from 11% to 74% depending on the number of tests considered in the diagnosis and the cutoffs used to define objective cognitive impairment (see also [26]). Neuropsychological criteria that balanced sensitivity (which defined impairment below −1 SD as opposed to −1.5 or −2 SDs) and reliability (required two impaired scores within a domain as opposed to a single impaired score), and incorporated instrumental activities of daily living (ADL) scores, produced more stable diagnoses over time and the strongest associations between cognition and hippocampal volume or stroke risk [20].

Another actuarial approach neuropsychologically characterizes MCI using cluster analytic statistical techniques that determine how individuals group together based on their patterns of performance across tests [27–31]. Applying these methods to different samples of individuals with clinically-diagnosed MCI, Delano-Wood et al. [27] and Libon et al. [28] independently identified three clusters that represented amnestic, dysexecutive, and mixed (i.e., memory and language deficits) MCI subtypes. Subsequent analyses using this actuarial approach further shows the distinct nature of these subtypes is evident in differences on cognitive dimensions not included in the cluster analyses (e.g., temporal gradients of forgetting, susceptibility to interference, qualitative errors; [29, 30]) and unique brain-behavior associations (e.g., the dysexecutive MCI subtype is associated with deep white matter lesions [27]; cortical thinning of the temporal cortex is associated with the amnestic MCI subtype [31]).

Clark et al. [31] recently compared the efficacy of the actuarial neuropsychological and conventional “one test” diagnostic approaches to detecting MCI by examining the nature of cluster analysis-derived subtypes identified with each method. The conventional “one test” MCI criteria identified more participants as having MCI (n = 134) than did the actuarial neuropsychological criteria (n = 80). Cluster analysis identified amnestic, dysexecutive, and mixed MCI subtypes when MCI had been determined using the actuarial neuropsychological method, consistent with the results of Delano-Wood et al. [27] and Libon et al. [28]. In contrast, when MCI had been determined using the conventional “one test” MCI criteria, Clark et al. [31] observed amnestic and mixed MCI subtypes, and a third cluster-derived normal subtype that performed within normal limits across all neuropsychological measures. This Cluster-Derived Normal group included nearly half of the “one test” MCI sample and did not differ from a normal control group in terms of cognition or neuroimaging measures.
of cortical thickness in areas usually affected in MCI or AD. These results suggest that the conventional “one test” method routinely used to diagnosis MCI may be highly susceptible to false positive diagnostic errors.

To further address these issues, we applied actuarial neuropsychological criteria for the diagnosis of MCI developed by Jak and Bondi [24, 25] to the baseline neuropsychological test data of 1,150 non-demented individuals (i.e., cognitively normal and MCI) participating in ADNI. We then compared the resulting MCI group to the original ADNI-diagnosed MCI group that had been defined using the conventional “one test” approach. Each MCI group was independently submitted to cluster and discriminant function analyses (as in [31]) to determine if similar MCI subtypes would be derived. The MCI subtypes derived within each method were then compared for APOE allelic frequencies and cerebrospinal fluid (CSF) concentrations of amyloid-β (Aβ)1–42, total tau, and hyperphosphorylated tau (p-tau181). Clinical outcomes were also compared to evaluate the stability of the MCI diagnoses. Based on previous results [31], we expected the actuarial neuropsychological method to provide more accurate and reliable identification of MCI compared to the conventional “one test” method used for ADNI diagnostic classification. We expected that MCI participants would have a higher frequency of the APOE ε4 allele and AD-related CSF biomarkers when diagnosed with actuarial neuropsychological criteria than with the ADNI criteria. Finally, MCI participants diagnosed with the actuarial neuropsychological method would be less likely to revert to a less impaired state of cognition and more likely to progress to a more impaired state of cognition than those MCI diagnosed by the ADNI classification scheme.

METHODS

Participants

The present study included 304 cognitively normal (CN; mean age = 74.9 years, SD = 5.8; mean education = 16.3 years, SD = 2.7; gender: 150 women/154 men) and 846 MCI (mean age = 73.2 years, SD = 7.7; mean education = 16.0 years, SD = 2.8; gender: 343 women/503 men) participants from ADNI. Criteria for ADNI eligibility and diagnostic classifications are described at http://www.adni-info.org/Scientists/ADNIGrant/ProtocolSummary.aspx. All participants were 55–91 years old, non-depressed, had a modified Hachinski score [32] of 4 or less, and had a study partner able to provide an independent evaluation of functioning. Individuals with a history of significant neurological or psychiatric disease, substance abuse, or metal in their body other than dental fillings were excluded. The CN group included all who had at least one year of follow-up and who remained classified as normal throughout their participation in ADNI. ADNI criteria for MCI were: 1) subjective memory complaints reported by themselves, study partner, or clinician; 2) objective memory loss defined as scoring below an education-adjusted cut-off score on delayed recall of Story A of the WMS-R Logical Memory Test (score = 8 for those with ≥16 years of education; score = 4 for those with 8–15 years of education; score = 2 for those with 0–7 years of education); 3) global CDR score of 0.5; and 4) general cognitive and functional performance sufficiently preserved such that a diagnosis of dementia could not be made by the site physician at the time of screening.
Materials and procedure

Following recruitment and ADNI diagnosis, all participants completed a battery of neuropsychological tests at baseline and yearly thereafter. The battery included tests of episodic memory (Rey Auditory Verbal Learning Test [AVLT]), language (Category Fluency Test (‘animals’), Boston Naming Test (30-items)), and speed/executive function (Trail-Making Test Parts A and B). The Geriatric Depression Scale (GDS; [33]), Functional Assessment Questionnaire (FAQ; [34]), and Neuropsychiatric Inventory-Quick (NPI-Q; [35]) were also administered. Presence or absence of the APOE ε4 allele was determined for each participant. CSF levels of Aβ1–42, total tau, p-tau, and the ratio of Aβ/p-tau were obtained for a majority of participants (see Supplementary Material). Clinical outcome was assessed for all participants who had follow-up examinations (up to 7 years) following baseline assessment. Clinical outcome is a diagnosis at each visit made initially by the site investigator and adjudicated by ADNI’s Clinical Core when a change of diagnostic category is proposed. Clinical outcomes were operationalized as: 1) “No change” for MCI participants who remained diagnosed as MCI; 2) “Reversion” for MCI participants who reverted back to a CN diagnosis on subsequent exams; and 3) “Progression” for MCI participants who progressed to a diagnosis of AD on subsequent exams.

Actuarial neuropsychological diagnostic classification

All ADNI CN and MCI participants were diagnostically reclassified using Jak/Bondi actuarial neuropsychological test methods [24, 25] applied to their baseline data. Six neuropsychological measures from ADNI were chosen because of their routine use in assessing early cognitive manifestations of AD and administration across all three ADNI grant periods (ADNI-1, -GO, and -2). The six measures were Category Fluency and Boston Naming Test scores for the language domain, Trail-Making Test Parts A and B scores for the graphomotor speed/executive function domain, and Rey AVLT delayed recall and delayed recognition scores for the episodic memory domain. None of these cognitive measures were used in making the initial ADNI diagnostic classification. Each measure was converted to an age-corrected standard score using published normative data: Mayo Older Americans Normative Study data (n = 530; [36]) for the Rey AVLT and National Alzheimer’s Coordinating Center normative data (n = 3,286; [37, 38]) for the remaining neuropsychological measures. Participants were considered to have MCI if any one of the following three criteria were met: 1) they had an impaired score, defined as >1 SD below the age-corrected normative mean, on both measures within at least one cognitive domain (i.e., memory, language, or speed/executive function); 2) they had one impaired score, defined as >1 SD below the age-corrected normative mean, in each of the three cognitive domains sampled; or 3) they had a score on the FAQ =9 indicating dependence in three or more daily activities. This latter criterion approximated Jak, Bondi et al.’s [25] incorporation of instrumental ADL assessment to diagnosis and reflects significant study partner-rated difficulties in everyday function. If none of these criteria were met, the participant was diagnosed as CN.
Statistical analyses

Cluster analyses require scores on a common metric, thus raw scores on each of the neuropsychological measures were first transformed to age-corrected z-scores based on the means and standard deviations of the normal reference group specific to each neuropsychological measure. Two hierarchical cluster analyses [27] were then conducted using the z-scores. The first cluster analysis included all subjects diagnosed as MCI using the ADNI diagnostic classification. The second cluster analysis included participants diagnosed as MCI using the actuarial neuropsychological criteria. Ward’s method was used in both analyses to calculate the distance between each cluster (squared Euclidean distance) and merge clusters together that produced the smallest increase in overall distances within clusters. The number of clusters was chosen based on examination of the resulting cluster structure using the dendogram plot. Because cluster analysis is a descriptive approach, discriminant function analyses (DFA) were conducted to quantitatively demonstrate the ability of the neuropsychological measures to discriminate the clustered subgroups. DFAs used the cognitive measures as predictors and the cluster analysis-derived subgroups as the outcome. A series of ANOVAs/ANCOVAs with Bonferroni-corrected post hoc tests compared the CSF AD biomarkers across the respective subtypes generated from each of the diagnostic methods. Finally, chi-square analyses were conducted to statistically compare the APOE ε4 allelic frequencies and clinical outcomes associated with each diagnostic method.

RESULTS

In contrast to the 846 MCI and 304 CN participants identified with the ADNI diagnostic procedures, the actuarial neuropsychological method identified 401 participants with MCI and 746 as cognitively normal (see Table 1). Of the 401 participants diagnosed with MCI based on actuarial neuropsychological criteria, 346 participants (86.3%) met the 1st criterion (impaired score on both measures within at least one cognitive domain), 16 (4.0%) met the 2nd criterion (one impaired score in each of the three cognitive domains sampled), and 39 met the 3rd criterion (score on the FAQ =9). Of these latter 39 participants, 20 also demonstrated impairment on one of the six neuropsychological tests, 10 were impaired on two neuropsychological measures (but not in the same domain), and 9 were unimpaired on all six cognitive scores. Excluding the 9 participants who had a high FAQ score but no impaired test scores had no effect on the overall pattern of the results. A McNemar’s chi-square analysis showed that the actuarial neuropsychological criteria classified significantly fewer individuals as MCI and more as CN compared to the original ADNI criteria (p < 0.001).

MCI subtype classification using conventional ADNI diagnostic criteria

Cluster analysis—Cluster analysis of the 846 participants diagnosed with MCI using the ADNI criteria resulted in three distinct subgroups (see Table 2 and Fig. 1A). Four- and five-cluster solutions resulted in poorer DFA classification percentages (85% versus the 90% with the three cluster solution) and were judged to be non-optimal. The first subgroup (n = 477) was considered Amnestic MCI based on mildly impaired performance on both memory measures (Rey AVLT delayed recall and recognition). The second subgroup (n = 104) was considered Dysexecutive/Mixed MCI based on severely impaired scores on speed/executive...
function (Trail-Making Test, Parts A and B) and mildly impaired memory and language scores. The third subgroup (n = 265), representing nearly one-third (31.3%) of the ADNI MCI sample, performed within normal limits on all six neuropsychological test measures included in the cluster analysis (mean z-scores ranged from −0.1 to +0.4, SDs = 0.6–0.9) and was considered Cluster-Derived Normal.

The three identified subgroups differed in years of education, as the Dysexecutive/Mixed subgroup had fewer years of education than both the Amnestic and Cluster-Derived Normal subgroups, and the Amnestic subgroup had fewer years of education than the Cluster-Derived Normal subgroup. The Cluster-Derived Normal subgroup also had a higher percentage of women, whereas the Amnestic subgroup had more men.

**Discriminant function analyses**—Two discriminant functions were obtained: The first accounted for 67.8% of the variability and the second discriminant function accounted for 32.2% of the variance among the three subgroups; see Fig. 2A. Overall, the six neuropsychological measures accurately classified 90.0% of the cases into the three subgroups. A cross-validation procedure using leave-one-out classification was used to validate results. Overall, classification fell minimally to 89.5% upon cross-validation.

**MCI subtype classification using actuarial neuropsychological criteria**

**Cluster analysis**—A cluster analysis of the 401 participants diagnosed with MCI using the actuarial neuropsychological method resulted in three distinct subgroups (see Table 2 and Fig. 1B). Four- or five-cluster solutions produced less evenly distributed classification percentages, did not improve the DFA classification percentages (91% versus 92% for the three cluster solution) and thus were considered non-optimal. The first subgroup (n = 236) was considered Amnestic MCI based on impaired delayed recall and recognition with normal performance on other cognitive measures. The second subgroup (n = 86) was considered Impaired Language MCI based on impairments on both language tasks (BNT and Category Fluency). The third subgroup (n = 79) was considered Dysexecutive/Mixed MCI based on severely impaired z-scores on measures of speed/executive function and mildly impaired z-scores on language and memory measures. In contrast to the results obtained with ADNI criteria, there was no evidence for a cluster derived normal group in any of the cluster solutions based on the neuropsychological method. The Dysexecutive/Mixed and Impaired Language MCI subgroups were significantly older and less educated than the Amnestic MCI group but did not differ from one another (see Table 2).

**Discriminant function analysis**—Two discriminant functions were found: The first accounted for 80.3% of the variability and the second accounted for 19.7% of the variance among the three subgroups; see Fig. 2B. The six neuropsychological measures accurately classified 91.8% of the cases into the three subgroups. Classification accuracy fell minimally to 90.3% when using the leave-one-out cross-validation procedure.

**APOE allelic frequencies, CSF AD biomarker levels, and progression rates**

**APOE ε4 allelic frequencies**—For MCI participants identified by ADNI criteria, a Genotype (APOE ε4 versus non-ε4) × Subgroup (Amnestic MCI, Dysexecutive/Mixed MCI,
Cluster-Derived Normal) chi-square analysis showed significant differences in APOE ε4 allelic frequencies across the three subgroups (see Table 2). The Amnestic and Dysexecutive/Mixed MCI subgroups demonstrated significantly higher ε4 frequencies than the Cluster-Derived Normal subgroup. In contrast, a Genotype × Subgroup (Amnestic, Impaired Language, Dysexecutive/Mixed) chi-square analysis with MCI defined using actuarial neuropsychological criteria showed no significant difference in APOE ε4 allelic frequencies across the three subgroups (see Table 2).

**CSF biomarkers**—For MCI participants identified using the ADNI diagnostic criteria, analyses of covariance showed that CSF biomarker values for the Amnestic and Dysexecutive/Mixed MCI subgroups were significantly different from those of the Cluster-Derived Normal subgroup, but did not differ from one another (see Table 2). Follow-up analyses showed that all CSF biomarker levels of the Amnestic and Dysexecutive/Mixed MCI subgroups were significantly different from those of a group of ADNI’s CN participants (n = 168; all p-values bold>0.001; see Fig. 3). In contrast, CSF biomarker levels in the Cluster-Derived Normal subgroup and the CN participants did not differ (all p-values >0.05). Thus, the Cluster-Derived Normal MCI subgroup demonstrated CSF biomarker levels comparable to the CN participants and significantly different from the abnormal levels found in the two cognitively impaired MCI subgroups based on the ADNI diagnostic criteria.

For MCI participants identified using the actuarial neuropsychological criteria, analyses of covariance showed no significant differences among the subgroups for all CSF biomarkers. Follow-up analyses showed that the Amnestic subgroup was significantly different from the CN participants (n = 405) for all CSF biomarker levels (all p-values <0.001). The Impaired Language subgroup had different levels of Aβ1–42, total tau, and p-tau181/Aβ1–42 compared to the CN participants. The Dysexecutive/Mixed subgroup had different levels of Aβ1–42, p-tau, and p-tau181/Aβ1–42 compared to the CN participants. Thus, when MCI was defined by actuarial neuropsychological criteria, all three MCI subgroups had abnormal levels of almost all the CSF AD biomarkers (see Fig. 3).

**Clinical outcomes**—For MCI participants identified using the conventional ADNI diagnostic criteria, a Type-of-Change (MCI to AD [progression], MCI to CN [reversion], or no change) ×Subgroup (Amnestic, Dysexecutive/Mixed, Cluster-Derived Normal) chi-square analysis revealed significant differences in prevalence of various types of change across the three subgroups (p < 0.001) (see Table 3). Overall, 30.3% of those diagnosed with MCI progressed to AD and 4.2% reverted back to CN classification. For the 239 participants who progressed to AD, the mean time point at which a dementia diagnosis was made was 22.5 months post-screening. For the 33 participants who reverted, the mean time point at which this classification was made was 19.5 months post-screening.

For MCI participants identified using the actuarial neuropsychological criteria, a Type-of-Change ×Subgroup (Amnestic, Impaired Language, Dysexecutive/Mixed) chi-square analysis showed no significant differences in clinical outcomes for the three subgroups (p = 0.10) (see Table 3). Overall, 49.0% of those diagnosed with MCI progressed to AD and less than 1% reverted back to a cognitively normal classification. For the 179 participants who...
progressed to AD, the mean time point at which a dementia diagnosis was made was 19.6 months post-screening. For the 2 participants who reverted, the mean time point at which this classification was made was 15.0 months post-screening.

**DISCUSSION**

The application of a neuropsychological method of actuarial diagnostic decision-making based on a minimal set of six neuropsychological variables and one functional measure provided a more accurate and better characterization of MCI than did conventional criteria based on subjective memory complaints, clinical interviews, cognitive screening and rating scales, and impairment on a single objective memory measure. When participants with MCI were identified through our actuarial neuropsychological method, they dissociated into three distinct cognitive phenotypes that varied in the salience of impairment in memory, language, and/or executive functions. Regardless of cognitive phenotype, the participants with MCI diagnosed in this way tended to remain as MCI or progress to dementia, rarely reverted to a cognitively normal status, had higher than normal APOE-ε4 allelic frequencies, and had abnormal CSF levels of Aβ1–42, total tau, and p-tau181 biomarkers associated with AD. In contrast, when participants with MCI were identified through the conventional criteria used by ADNI, they could be dissociated into two cognitive phenotypes that varied in the salience of impairment in memory and executive functions, and a third “cognitively normal” phenotype (i.e., they performed within normal limits on the six neuropsychological tests) that comprised nearly one-third of their MCI sample. While the MCI participants in the two impaired phenotypes tended to remain as MCI or progress to dementia, rarely revert to a CN status, have a higher than normal APOE ε4 allelic frequency, and abnormal CSF Aβ1–42 and tau biomarkers, those in the “cognitively normal” phenotype had a very low rate of progression (approximately four to five times less than the impaired phenotypes), were as likely to revert as to progress, had only a slightly higher than normal APOE ε4 allelic frequency, and demonstrated normal levels of CSF Aβ1–42 and tau biomarkers. They also reported more depressive symptoms but less ADL concerns than the two impaired phenotypes.

These findings suggest that the conventional criteria for MCI used in ADNI, based in part on a cutoff from one test of memory, together with other cognitive and informant-derived tests and clinical consensus, produce a relatively high rate of “false positive” diagnostic errors. This may arise from an over-reliance on a single impaired test score. Several studies have shown that a majority of neurologically normal adults score in the impaired range on at least one measure when tested with an extensive battery of cognitive tests [39, 40]. In these studies the median number of impaired scores in a neurologically normal sample was 10%. More specific to aging, Palmer et al. [41] found that more than 20% of healthy older adults tested with a battery of tests containing multiple measures for each cognitive domain obtained one impaired score in two different cognitive domains, whereas less than 5% had two or more impaired scores in the same domain. Brooks et al. [42] showed that 26% of the older adults in the standardization sample for the WMS-III [43] obtained one or more age-adjusted standard score that was more than 1.5 SDs below the mean.
Several additional factors may contribute to inaccuracy of the conventional “one-test” criteria in identifying MCI in older adults. More strict cut-points on cognitive tests of -1.5 or -2 SDs below normative means generally trade modest gains in specificity for larger losses in sensitivity [44, 45], and several studies suggest that a cut-point for impairment of -1 SD below normative means provides an optimal balance of sensitivity and specificity [25, 44, 45]. Using an actuarial neuropsychological method to circumvent the difficulty of interpreting an isolated impaired score on a single cognitive test, applying cut-off scores for impairment that optimize classification rates, understanding the base rates of ‘impaired’ low scores, and assigning less weight to subjective ratings of cognitive impairment [46–49], might reduce over-interpretation of isolated low scores and minimize the potential for a false positive diagnosis of MCI (see [24, 25, 50] for discussion).

The susceptibility of the conventional criteria for MCI to false positive errors could have unfortunate ‘downstream’ consequences. If a high number of cognitively normal individuals without significant amyloidosis or neurodegeneration are incorrectly identified as MCI in studies of potential genetic, imaging, or other biomarkers, the perceived accuracy of these biomarkers could be greatly reduced. This possibility is also true for clinical trials of drug therapies that target the underlying biology of AD, such that including inaccurately identified “false-positive” MCI cases in the cohort could dilute their results.

The susceptibility of the conventional criteria for MCI to diagnostic error is likely to be amplified if used to characterize and stage preclinical AD—a pre-MCI categorization based on subtle cognitive decline [8]. Initial studies in this emerging area have identified subgroups of older adults with subtle cognitive decline that do not conform to expected patterns and investigators have designated them as Suspected Non-Alzheimer Pathology (SNAP) and Unclassified groups [51, 52]. It may be the case that some of these designations are false positive errors since the measurement strategies in these studies have continued to identify cases on the basis of a single cognitive composite and global CDR scores. Reliance on these conventional “one test” methods to assign diagnoses based on fine-grained distinctions of subtle cognitive decline may perpetuate error-prone diagnostic decision-making and obscure assessment of the effectiveness of potentially useful therapeutics or biomarkers.

Many MCI studies diagnose participants on the basis of a single impaired test score, the most prevalent of which is an impaired memory score. An important implication of the present results is that subtle decline in cognitive abilities other than an assessment of delayed free recall obtained from a single episodic memory test should be considered when making the diagnosis of MCI. The identification of clusters of MCI participants with prominent executive dysfunction and language/semantic impairment in the present study supports this contention, as do previous demonstrations of multi-domain cognitive declines in individuals with both preclinical AD or MCI (for reviews, see [53, 54]) as well as across the spectrum of AD- and vascular-related dementias [55].

One caveat of our study was the absence of assessing other cognitive domains like that of visuospatial functions, particularly since we have previously identified a visuospatial MCI subtype [31] and Ferman et al. [19] have shown that baseline MCI diagnoses based on
visuospatial deficits reliably predict development of dementia with Lewy bodies. Another limitation of our study includes an inability to examine false negative diagnostic errors due to our decision to use a ‘robust’ normal control sample. In other words, participants misclassified as CN but found to have cognitive impairment on more extensive testing or who subsequently declined were not included in the CN sample. Future efforts to more fully profile accuracies of the diagnoses will shed additional light on the utility of the varying MCI diagnostic approaches. Strengths of our study include a large well characterized sample, an empirical statistical approach to the identification of MCI phenotypes, employing a robust normative reference group, and relating the actuarial diagnostic approach to CSF AD biomarkers and longitudinal outcomes.

Additional directions for future research will be to examine a fuller sampling within and across cognitive domains, examine different normative referencing methods to examine their differential impact on MCI diagnosis and progression, as well as to use neuroimaging to compare the structural and functional underpinnings of empirically-derived subtypes (i.e., “clusters”) of MCI. As pointed out by Gorelick et al. [56], a comprehensive assessment strategy is necessary for examining vascular contributions to subtle cognitive impairment, MCI, and dementia; and it notably differs from those strategies expressed by McKhann et al. [6] that bedside mental status testing—though not optimal—is acceptable or by the DSM-5 which recommends cognitive impairment be assessed either by neuropsychological testing or some other (unspecified) “clinical assessment” strategy. We suggest these latter types of approaches will miss meaningful numbers of individuals who have subtle cognitive decline that does not fit the typical profile for AD, and possibly lead to “false positive” diagnoses of MCI in some who are cognitively normal when tested with a comprehensive battery of neuropsychological tests.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institute on Aging (NIA) grants R01 AG012674 (MB), R01 AG16495 (RA), P50 AG05131 (DG), K24 AG026431 (MB), and by grant NIRG 13-281806 (CM) from the Alzheimer’s Association. Dr. Salmon serves as a consultant for CHDI Foundation, Novartis, and Bristol-Meyers Squibb. The authors gratefully acknowledge the assistance of Ivy Ewald in the preparation of this manuscript.

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by NIA, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuroimaging at the University of Southern California.
References


Fig. 1. Mean z-scores for the three MCI subtypes on neuropsychological measures included in cluster analyses of conventional Petersen/Winblad ADNI criteria (A) and neuropsychological Jak/Bondi criteria (B). Error bars denote standard deviations. TMT, Trail Making Test.
Fig. 2.
Individual scores on discriminant functions for MCI participants classified according to (A) the conventional criteria and (B) the neuropsychological criteria.
Fig. 3.
CSF biomarker levels of (A) $\text{A}\beta_{1-42}$, (B) total tau, and (C) p-tau$_{181}$ for the cluster subgroups and CN participants according to the conventional criteria and actuarial neuropsychological criteria. Error bars denote standard deviations.
Table 1
Comparison of participants classified as MCI versus cognitively normal based on the conventional Petersen/Winblad ADNI criteria and actuarial neuropsychological Jak/Bondi criteria

<table>
<thead>
<tr>
<th>Neuropsychological Criteria</th>
<th>ADNI criteria</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MCI</td>
<td>Normal</td>
<td>Total</td>
</tr>
<tr>
<td>MCI</td>
<td>386 (33.6%)</td>
<td>15 (1.3%)</td>
<td>401</td>
</tr>
<tr>
<td>Normal</td>
<td>460 (40.0%)</td>
<td>289 (25.1%)</td>
<td>749</td>
</tr>
<tr>
<td>Total</td>
<td>846</td>
<td>304</td>
<td>1150</td>
</tr>
</tbody>
</table>
Table 2

Demographic, clinical, genetic, CSF biomarker, and neuropsychological characteristics of MCI subtypes resulting from cluster analyses of participants diagnosed via conventional ADNI MCI criteria (left) and actuarial neuropsychological criteria (right)

<table>
<thead>
<tr>
<th></th>
<th>Conventional Petersen/Winblad ADNI MCI criteria</th>
<th>Actuarial Jak/Bondi neuropsychological MCI criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amnestic (n = 477)</td>
<td>Dysexec/ Mixed (n = 104)</td>
</tr>
<tr>
<td><strong>Demographic Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>73.1 (7.2)</td>
<td>74.1 (7.9)</td>
</tr>
<tr>
<td>Education</td>
<td>16.0 (2.8)</td>
<td>14.8 (3.4)</td>
</tr>
<tr>
<td>Gender (F/M)</td>
<td>171/306</td>
<td>41/63</td>
</tr>
<tr>
<td><strong>Cluster Analysis Measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animal Fluency</td>
<td>−0.6 (0.8)</td>
<td>−1.3 (0.7)</td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>−0.3 (1.1)</td>
<td>−1.1 (1.5)</td>
</tr>
<tr>
<td>Speed/Executive Function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trail Making Test, Part A</td>
<td>−0.2 (0.8)</td>
<td>−2.7 (2.2)</td>
</tr>
<tr>
<td>Trail Making Test, Part B</td>
<td>−0.3 (0.8)</td>
<td>−3.6 (1.1)</td>
</tr>
<tr>
<td>Memory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVLT Recall</td>
<td>−1.4 (0.7)</td>
<td>−1.3 (0.8)</td>
</tr>
<tr>
<td>AVLT Recognition</td>
<td>−1.1 (0.8)</td>
<td>−1.1 (1.1)</td>
</tr>
<tr>
<td><strong>Genetic / CSF Biomarkers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOE (ε4/non-ε4)</td>
<td>261/210</td>
<td>60/43</td>
</tr>
<tr>
<td>Aβ1-42 (pg/ml)</td>
<td>177.4 (68.7)</td>
<td>164.9 (59.1)</td>
</tr>
<tr>
<td>Total tau (pg/ml)</td>
<td>99.4 (63.4)</td>
<td>104.4 (56.6)</td>
</tr>
<tr>
<td>p-tau (pg/ml)</td>
<td>31.9 (17.7)</td>
<td>36.8 (16.1)</td>
</tr>
<tr>
<td>p-tau/Aβ1-42</td>
<td>0.22 (0.18)</td>
<td>0.25 (0.13)</td>
</tr>
<tr>
<td><strong>Clinical Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDS</td>
<td>1.4 (1.4)</td>
<td>1.5 (1.5)</td>
</tr>
<tr>
<td>NPI-Q</td>
<td>1.9 (2.6)</td>
<td>2.1 (2.9)</td>
</tr>
<tr>
<td>FAQ</td>
<td>3.6 (4.4)</td>
<td>4.7 (4.6)</td>
</tr>
<tr>
<td>6-month MMSE follow-up</td>
<td>26.8 (2.4)</td>
<td>25.0 (3.0)</td>
</tr>
</tbody>
</table>
Aβ1–42, amyloid-β protein; ADNI, Alzheimer's Disease Neuroimaging Initiative; AVLT, Rey Auditory Verbal Learning Test; APOE, apolipoprotein E; CSF, cerebrospinal fluid; FAQ, Functional Activity Questionnaire; GDS, Geriatric Depression Scale; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; NPI-Q, Neuropsychiatric Inventory; p-tau, hyperphosphorylated tau; pg/ml, picograms/milliliters.

*Information regarding APOE genotype was not available for 10 participants classified as MCI by conventional ADNI criteria and 5 participants classified as MCI by actuarial neuropsychological criteria. For MCI participants classified using conventional ADNI criteria, CSF biomarker data were available for 244 participants in the Amnestic MCI subgroup (51.2%), 53 participants in the Dysexecutive/Mixed MCI subgroup (51.0%), and 146 participants in the Cluster-Derived Normal subgroup (55.1%). For MCI participants identified using the actuarial neuropsychological criteria, CSF biomarker data were available for 130 participants from the Amnestic subgroup (55.1%), 38 participants from the Impaired Language subgroup (44.2%), and 38 participants from the Dysexecutive/Mixed subgroup (48.1%).
### Table 3

Comparison of the progression, reversion, and stability rates of MCI participants with follow-up data who were diagnosed based on the conventional Petersen/Winblad ADNI criteria and neuropsychological Jak/Bondi criteria

<table>
<thead>
<tr>
<th></th>
<th>Progression (MCI to dementia)</th>
<th>Reversion (MCI to cognitively normal)</th>
<th>No change</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MCI based on ADNI criteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amnestic</td>
<td>239 (30.3%)</td>
<td>33 (4.2%)</td>
<td>518 (65.6%)</td>
<td>790 (100%)</td>
</tr>
<tr>
<td>Dysexecutive/Mixed</td>
<td>159 (35.8%)</td>
<td>11 (2.5%)</td>
<td>272 (61.7%)</td>
<td>444 (100%)</td>
</tr>
<tr>
<td>Cluster-Derived Normal</td>
<td>57 (57.0%)</td>
<td>1 (1.0%)</td>
<td>42 (42.0%)</td>
<td>100 (100%)</td>
</tr>
<tr>
<td><strong>MCI based on Neuropsychological criteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amnestic</td>
<td>179 (49.0%)</td>
<td>2 (0.5%)</td>
<td>184 (50.4%)</td>
<td>365 (100%)</td>
</tr>
<tr>
<td>Impaired Language</td>
<td>101 (46.5%)</td>
<td>0 (0.0%)</td>
<td>116 (53.5%)</td>
<td>217 (100%)</td>
</tr>
<tr>
<td>Dysexecutive/Mixed</td>
<td>33 (45.2%)</td>
<td>1 (1.4%)</td>
<td>39 (53.4%)</td>
<td>73 (100%)</td>
</tr>
</tbody>
</table>

Longitudinal data were available for 93.4% of the MCI sample (mean follow-up = 22.9 months; range 6–84 months).