Alzheimer's & Dementia 11 (2015) 415-424

Susceptibility of the conventional criteria for mild cognitive impairment to false-positive diagnostic errors

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

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

Abstract

Background: We assessed whether mild cognitive impairment (MCI) subtypes could be empirically derived within the Alzheimer's Disease Neuroimaging Initiative (ADNI) MCI cohort and examined associated biomarkers and clinical outcomes.

Methods: Cluster analysis was performed on neuropsychological data from 825 MCI ADNI participants.

Results: Four subtypes emerged: (1) *dysnomic* (n = 153), (2) *dysexecutive* (n = 102), (3) *amnestic* (n = 288), and (4) *cluster-derived normal* (n = 282) who performed within normal limits on cognitive testing. The cluster-derived normal group had significantly fewer *APOE* ε4 carriers and fewer who progressed to dementia compared with the other subtypes; they also evidenced cerebrospinal fluid Alzheimer's disease biomarker profiles that did not differ from the normative reference group.

Conclusions: Identification of empirically derived MCI subtypes demonstrates heterogeneity in MCI cognitive profiles that is not captured by conventional criteria. The large cluster-derived normal group suggests that conventional diagnostic criteria are susceptible to false-positive errors, with the result that prior MCI studies may be diluting important biomarker relationships.

Keywords: Mild cognitive impairment; MCI; Alzheimer's disease; Dementia; Neuropsychology; Misdiagnosis; Misclassification; Cluster analysis

1. Introduction

Mild cognitive impairment (MCI), conceptualized as a transitional state between normal aging and dementia, is defined by objective evidence for cognitive impairment along with a subjective memory complaint in the context of preserved global cognition and activities of daily living [1–3]. Objective impairment is typically operationalized as 1.5 standard deviations (SDs) or more below normative means on at least one measure in a neuropsychological battery. MCI has been further divided as "amnestic," characterized

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of the ADNI and/or provided data but did not participate in analysis or writing of this article. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

*Corresponding author. Tel.: +1-858-552-8585x2809; Fax: +1-858-642-1218.

E-mail address: mbondi@ucsd.edu

1552-5260/$ - see front matter © 2015 The Alzheimer's Association. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jalz.2014.03.005

by predominant memory impairment, and "nonamnestic," which involves deficits in other cognitive domains such as executive functions or language. However, recent research using cluster analytical techniques has demonstrated that individuals with MCI can be grouped based on similarities in their neuropsychological profiles, providing an actuarial method of describing MCI subtypes without being confined to the amnestic/nonamnestic distinction [4–6].

One critical finding from a recent cluster analytic study was the identification of a large subgroup who performed within normal limits on neuropsychological testing despite their MCI diagnosis [4]. This cluster-derived normal group did not differ from a normal control group in terms of cognition or imaging measures of cortical thickness in areas usually affected in MCI or Alzheimer's disease (AD). These results suggest that the conventional diagnosis of MCI may be highly susceptible to false-positive diagnostic errors, which is consistent with previous reports of high reversion rates or lack of progression in those with MCI [7–12].

To replicate and extend our previous findings to a large cohort with longitudinal clinical outcomes, we assessed whether distinct MCI subtypes could be empirically derived within the Alzheimer's Disease Neuroimaging Initiative (ADNI) MCI cohort and, if present, examined associated clinical characteristics, biological markers, and longitudinal outcomes.

2. Methods

Data were obtained from the ADNI database (adni.loni.usc.edu). The primary goal of ADNI is to test whether neuroimaging, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. ADNI is the result of efforts of many coinvestigators from a range of academic institutions and private corporations, and subjects have been recruited from more than 50 sites across the United States and Canada. Participants are recruited via newsletters, Web-based communication, direct mail, and press releases. Inclusion criteria include: age 55 to 90 years, permitted medications stable for 4 weeks, study partner who can accompany participant to visits, Geriatric Depression Scale less than 6, Hachinski Ischemic Score less than or equal to 4, adequate visual and auditory acuity, good general health, 6 grades of education or work history equivalent, and ability to speak English or Spanish fluently. Exclusion criteria for cognitively normal and MCI participants include any significant neurologic disease or history of significant head trauma. For more information, see www.adni-info.org.

2.1. Participants

Participants were 1109 ADNI participants who completed a neuropsychological evaluation: 825 diagnosed as MCI at their initial screening evaluation based on ADNI diagnostic criteria [2,13] and 284 classified as cognitively normal. Nearly all of the 825 MCI participants were classified as "amnestic MCI" by ADNI, with only two being coded as "nonamnestic MCI." Criteria for MCI were (1) subjective memory complaint reported by participant or study partner; (2) Mini-Mental State Examination (MMSE) scores between 24 and 30 (inclusive); (3) global Clinical Dementia Rating (CDR) score of 0.5; (4) abnormal memory function documented by scoring below education-adjusted cutoffs for delayed free recall on story A of the Wechsler Memory Scale-Revised (WMS-R) Logical Memory II subtest [14]; and (5) general cognition and functional performance sufficiently preserved to an extent that one could not qualify for a diagnosis of AD. Importantly, we retained in the normal control group all participants who had at least 1 year of follow-up data and who remained classified as normal for the duration of their participation in the study (range of 1–7 years of follow-up). The normal control group of 284 participants did not differ from the MCI group in terms of age, education, or gender (P-values $>.05$).

2.2. Materials and procedure

2.2.1. Neuropsychological battery

Cognitive measures consisted of six scores from each participant's baseline neuropsychological evaluation: (1) Animal Fluency, total score; (2) 30-item Boston Naming Test (BNT) total score; (3) Trail Making Test (TMT), part A, time to completion; (4) TMT, part B, time to completion; (5) Rey Auditory Verbal Learning Test (AVLT) 30-minute delayed free recall, number of words recalled; and (6) AVLT recognition, number of words correctly recognized. These variables were selected because they were administered to all participants and they assessed three different domains of cognitive ability—language (Animal Fluency and BNT), attention/executive function (TMT, parts A and B), and memory (AVLT recall and recognition).

2.2.2. Cerebrospinal fluid and genetic biomarkers

Biological markers included cerebrospinal fluid (CSF) concentrations of hyperphosphorylated tau (p-tau$_{181p}$), β-amyloid (A$\beta_{1\text{-}42}$), the ratio of p-tau$_{181p}$ to A$\beta_{1\text{-}42}$ [15], and frequency of the apolipoprotein E (*APOE*) ε4 allele [16–18].

2.3. Statistical analyses

Raw neuropsychological scores for each MCI participant were converted into age- and education-adjusted z scores based on regression coefficients derived from the normal control group. A hierarchical cluster analysis was performed on the z scores using Ward's method, consistent with previous MCI studies [4,5]. A discriminant function analysis (DFA) was conducted to more quantitatively examine the ability of the six neuropsychological measures to discriminate the cluster subgroups. The stability of the cluster solution was

also examined using the leave-one-out cross-validation procedure, a method that reduces the potential bias of using the same individuals to develop the classification matrix and to compute the discriminant function. Following these analyses, differences between groups (i.e., cluster and normal control groups) were examined using a series of analysis of variance/analyses of covariance with post hoc *t* tests and chi-squares. Bonferroni correction was used to adjust for multiple comparisons. Survival curves and Cox regression were used to explore progression and reversion rates.

3. Results

3.1. Cluster analysis and DFA

A cluster analysis of the neuropsychological scores from 825 MCI participants resulted in four distinct subgroups based on the mean performance for each group (see Fig. 1): (1) *dysnomic* MCI (n = 153; 18.5%) with a significant deficit in naming; (2) *dysexecutive* MCI (n = 102; 12.4%) with a significant deficit in executive function, as well as impairments in attention, naming, and memory; (3) *amnestic* MCI (n = 288; 34.9%) with isolated memory impairment; and (4) a *cluster-derived normal* group (n = 282; 34.2%) that performed within normal limits on cognitive testing.

DFA using the six neuropsychological measures to predict group membership in the four cluster groups identified three discriminant functions: the first accounted for 74.0% of the variance between groups, the second for 17.1%, and the third for 8.8%. The full predictive model accurately classified 88.0% of participants, and cross-validation of the four-cluster solution using the leave-one-out method showed only mild expected reduction in correct classification (87.3%). A four-cluster solution was determined to be optimal relative to a three-cluster solution that combined the dysnomic and amnestic groups into one group (as this did not allow us to examine how the traditional "amnestic MCI" subtype compared with other cognitive phenotypes), or a five-cluster solution that produced unbalanced groups (i.e., one group had only 10 participants). Notably, all the cluster solutions produced an invariant cluster-derived normal group of 282 participants.

3.2. Clinical characteristics of the cluster and normal control groups

3.2.1. Demographic characteristics

As shown in Table 1, the five groups differed in terms of age and education ($P \leq .001$). For age, the participants in the amnestic group were significantly younger than those in the dysnomic, dysexecutive, and normal control groups, and the participants in the cluster-derived normal group were younger than those in the dysnomic group. For education, the participants in the dysexecutive group were significantly less educated than those in all other groups. There was no gender difference between groups ($P > .05$). All further analyses used age and education as covariates.

3.2.2. Neuropsychological performance

As shown in Table 1, there were significant group differences on all six neuropsychological measures ($P < .001$). Post hoc *t* tests with Bonferroni correction confirmed that the dysnomic group performed worse than all other groups on measures of language, with the exception of equivalent

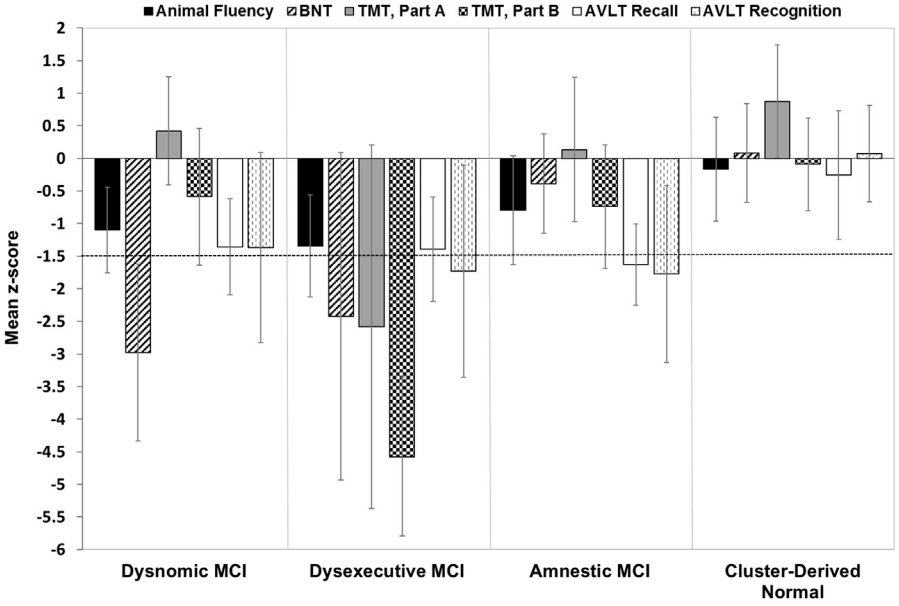

Fig. 1. Neuropsychological performance for the cluster groups. Error bars denote standard deviations (SDs). The horizontal dotted line indicates the typical cutoff for impairment (−1.5 SDs). BNT, Boston Naming Test; TMT, Trail Making Test; AVLT, Rey Auditory Verbal Learning Test; MCI, mild cognitive impairment.

Table 1
Demographic, neuropsychological, biomarker, and clinical outcome characteristics of the cluster groups and normal control group

Variable	Dysnomic MCI (n = 153)	Dysexecutive MCI (n = 102)	Amnestic MCI (n = 288)	Cluster-derived normal (n = 282)	Normal control (n = 284)	F or χ^2	Sig.	Effect size
Demographics*								
Age (y)	75.5 (6.8)	74.7 (7.3)	72.5 (6.9)	73.1 (7.8)	74.2 (5.2)	$F = 6.4$	$P < .001$	$\eta_p^2 = 0.02$
Education (y)	16.1 (2.9)	15.0 (3.3)	16.1 (2.6)	16.2 (2.6)	16.3 (2.7)	$F = 4.8$	$P = .001$	$\eta_p^2 = 0.02$
Gender (% male)	59.5	56.9	62.2	54.6	51.8	$\chi^2 = 7.3$	$P > .05$	$\varphi_c = 0.08$
Cognitive measures (raw)*								
Language								
Animal Fluency	14.5 (3.9)	12.6 (4.2)	16.8 (4.4)	20.3 (4.7)	21.0 (5.6)	$F = 103.7$	$P < .001$	$\eta_p^2 = 0.27$
BNT	22.1 (2.8)	23.0 (5.4)	27.5 (1.7)	28.4 (1.6)	28.3 (2.0)	$F = 254.9$	$P < .001$	$\eta_p^2 = 0.48$
Attention/executive function								
TMT, part A (s)	38.7 (9.5)	71.3 (30.9)	40.3 (12.7)	32.5 (10.3)	34.0 (11.0)	$F = 157.9$	$P < .001$	$\eta_p^2 = 0.36$
TMT, part B (s)	107.2 (41.5)	258.5 (49.9)	106.2 (38.9)	82.9 (29.8)	81.6 (38.0)	$F = 498.1$	$P < .001$	$\eta_p^2 = 0.64$
Memory								
AVLT recall	2.6 (2.8)	2.4 (2.8)	2.0 (2.2)	7.1 (4.0)	7.9 (3.8)	$F = 188.1$	$P < .001$	$\eta_p^2 = 0.41$
AVLT recognition	9.8 (3.4)	9.0 (3.8)	9.0 (3.1)	13.2 (1.7)	13.0 (2.3)	$F = 142.1$	$P < .001$	$\eta_p^2 = 0.34$
Diagnostic measures (raw)*								
LM II recall	4.5 (3.2)	3.9 (3.1)	5.0 (3.3)	7.6 (2.8)	13.6 (3.2)	$F = 97.4$	$P < .001$	$\eta_p^2 = 0.59$
MMSE	27.1 (1.8)	26.7 (1.7)	27.4 (1.8)	28.4 (1.5)	29.1 (1.2)	$F = 74.4$	$P < .001$	$\eta_p^2 = 0.21$
CDR: sum of boxes	1.6 (0.9)	1.8 (0.9)	1.6 (0.9)	1.2 (0.7)	0.0 (0.1)	$F = 224.7$	$P < .001$	$\eta_p^2 = 0.45$
CSF†/genetic‡ biomarkers								
% high p-tau$_{181p}$	67.6	82.0	59.7	37.8	31.4	$\chi^2 = 66.2$	$P < .001$	$\varphi_c = 0.34$
% low A$\beta_{1\text{-}42}$	66.2	84.0	67.5	35.3	31.4	$\chi^2 = 84.9$	$P < .001$	$\varphi_c = 0.38$
% high p-tau$_{181p}$/A$\beta_{1\text{-}42}$	73.0	84.0	68.2	40.4	36.5	$\chi^2 = 72.2$	$P < .001$	$\varphi_c = 0.35$
% *APOE* ε4 carriers	53.6	60.4	58.6	37.8	27.7	$\chi^2 = 64.0$	$P < .001$	$\varphi_c = 0.25$
Clinical outcome§								
% progression to dementia	40.6	55.6	34.7	10.7	—	$\chi^2 = 100.6$	$P < .001$	$\varphi_c = 0.26$
% reversion to normal	1.4	1.0	2.2	9.2	—			

Abbreviations: MCI, mild cognitive impairment; BNT, Boston Naming Test; TMT, Trail Making Test; AVLT, Rey Auditory Verbal Learning Test; LM, Logical Memory; MMSE, Mini-Mental State Examination; CDR, Clinical Dementia Rating; p-tau$_{181p}$, hyperphosphorylated tau; A$\beta_{1\text{-}42}$, β-amyloid; *APOE*, apolipoprotein E.

*Data are summarized as mean (standard deviation), unless otherwise indicated.

†Number of participants for CSF analysis: dysnomic: n = 74, dysexecutive: n = 50, amnestic: n = 154, cluster-derived normal: n = 156, and normal control: n = 156.

‡Number of participants for *APOE* analysis: dysnomic: n = 151, dysexecutive: n = 101, amnestic: n = 285, cluster-derived normal: n = 278, and normal control: n = 282.

§Number of participants progression/reversion analysis: dysnomic: n = 138, dysexecutive: n = 99, amnestic: n = 274, and cluster-derived normal: n = 262.

performance between the dysnomic and dysexecutive groups on Animal Fluency. The dysexecutive group performed worse than all other groups on measures of attention/executive functioning. The amnestic group performed worse than the cluster-derived normal and normal control groups on both measures of memory and worse than the dysnomic group on one measure of memory (AVLT recognition). There was no significant difference between the cluster-derived normal and the normal control groups on five of the six neuropsychological measures ($P > .05$); although there was a statistically significant difference in performance on AVLT recall ($P < .01$), the cluster-derived normal group's performance was less than a one-word difference (7.1 vs. 7.9 words), and their mean score on this measure fell well within normal limits (z score = -0.26).

3.2.3. Performances on ADNI's diagnostic measures

On the WMS-R Logical Memory II subtest, which was used in ADNI's MCI diagnosis and thus not included in the cluster analysis, the dysnomic, dysexecutive, and amnestic groups performed similarly to each other ($P > .05$) but worse than the cluster-derived normal group ($P < .001$). A similar pattern was found on the MMSE, as the three impaired groups performed worse than the cluster-derived normal group ($P < .001$). Global CDR scores were 0.5 for all cluster groups, as this was a criterion for an MCI diagnosis; however, the dysnomic, dysexecutive, and amnestic groups scored higher on the CDR sum of boxes compared with the cluster-derived normal group ($P < .001$).

3.3. Biomarker characteristics of the cluster and normal control groups

3.3.1. CSF biomarkers

CSF data were available for 53.2% of the sample (see footnote of Table 1). Based on established CSF cut point concentrations for p-tau$_{181p}$, A$\beta_{1\text{-}42}$, and the p-tau$_{181p}$-to-A$\beta_{1\text{-}42}$ ratio [19], participants were classified into

dichotomous groups (high/low) for each variable. Chi-square analysis showed significant differences between groups for all three CSF measures ($P < .001$; see Table 1). Specifically, all MCI groups demonstrated a greater percentage of individuals with positive CSF AD biomarkers (i.e., high p-tau$_{181p}$, low A$\beta_{1\text{-}42}$, high p-tau$_{181p}$-to-A$\beta_{1\text{-}42}$ ratio) compared with the cluster-derived normal and the normal control groups, whereas percentages were comparable between the cluster-derived normal and normal control groups. In addition, the dysexecutive group had higher percentages of individuals with positive CSF AD biomarkers compared with the dysnomic and amnestic groups. When CSF measures were analyzed as continuous variables, the same pattern was found for all three measures (see Fig. 2): no differences between the dysnomic, dysexecutive, and amnestic groups ($P > .05$), but all had higher p-tau$_{181p}$, lower A$\beta_{1\text{-}42}$, and larger p-tau$_{181p}$-to-A$\beta_{1\text{-}42}$ ratios compared with the cluster-derived normal and normal control groups ($P < .001$). No differences were observed between the cluster-derived normal and normal control groups for any CSF measure ($P > .05$).

A DFA was conducted with only the subgroup of MCI cases who had CSF data available (n = 434). The model accurately classified 86.6% of participants, and cross-validation fell minimally to 85.3%. Thus, the classification rates with this subset were comparable with the rates in the entire MCI sample.

3.3.2. APOE

APOE genotypes were available for 98.9% of the sample (see footnote of Table 1). A 2 (*APOE* ε4 vs. non-ε4) × 5 (group) chi-square analysis revealed significant group differences in *APOE* ε4 frequencies (see Table 1). The dysnomic, dysexecutive, and amnestic groups had significantly more *APOE* ε4 carriers (53.6%–60.4%) than the cluster-derived normal (37.8%) and normal control (27.7%) groups, although the percentage in the cluster-derived normal group was also significantly higher than that in the normal control group.

3.4. Longitudinal clinical outcomes

3.4.1. Progression/reversion rates

Longitudinal data (mean follow-up, 22.9 months; range, 6–84 months), which were available for 93.7% of the MCI sample (see footnote of Table 1), showed that a subset of participants in ADNI's MCI sample progressed to meet criteria for a diagnosis for probable AD, whereas a smaller subset reverted to normal (i.e., no longer met criteria for MCI) over time. The dysexecutive group had slightly less follow-up (18.6 months) than the other three cluster groups (22–25 months; $P < .05$). A 3 (no change, progression from MCI to AD, and reversion from MCI to normal) × 4 (cluster group) chi-square analysis revealed significant differences between the cluster groups (see Table 1), with the cluster-derived normal group showing the lowest rate of progression to dementia (10.7%) and the highest rate of reversion to normalcy (9.2%). The cluster-derived normal group also showed a different survival curve compared with the other cluster groups (see Fig. 3). The dysexecutive group showed the highest rate of progression to dementia (55.6%). Cox regression including demographic, neuropsychological, and biomarker variables showed that reduced risk of progression to dementia was associated with better scores on the AVLT delayed recall ($P < .001$, hazard ratio = 0.504) and TMT, part B ($P < .01$, hazard ratio = 0.848). The participants in the normal control group were not included in these analyses because they were selected on the basis of remaining normal (did not progress/revert) throughout the course of their participation in ADNI.

Post hoc analysis showed that, within the cluster-derived normal group, the 28 individuals who progressed to dementia were slightly older ($P = .03$), performed worse on memory testing ($P = .001$), and had lower A$\beta_{1\text{-}42}$ ($P < .01$) and a slightly higher p-tau$_{181p}$-to-A$\beta_{1\text{-}42}$ ratio ($P = .04$) compared with those who did not progress. Also, 15 of the 28 (53.6%) who progressed carried the *APOE* ε4 allele. The mean time point at which a dementia diagnosis was made for these 28 individuals was 33.2 months after screening.

4. Discussion

We empirically derived subgroups from the ADNI MCI cohort using cluster analysis based on performances on six neuropsychological measures. Four MCI subgroups emerged: dysnomic, dysexecutive, amnestic, and a cluster-derived normal group who performed within normal limits on all six neuropsychological measures (mean *z* scores ranged from −0.26 to +0.87) despite their other performances on logical memory, MMSE, and global CDR scores that lead to their ADNI MCI diagnosis. The cluster-derived normal group comprised one-third (34%) of the ADNI MCI sample and was comparable with a robust normal control group in neuropsychological test performance and percentage of individuals with positive CSF biomarkers for AD. In addition, the cluster-derived normal group had fewer *APOE* ε4 carriers and fewer individuals with positive CSF biomarkers of AD than the dysnomic, dysexecutive, and amnestic MCI groups. The cluster-derived normal group was also less likely to progress to AD and more likely to revert to normal than the other three MCI groups.

These results are consistent with those of previous cluster analytic studies showing heterogeneity in neuropsychological [4] and biomarker profiles [20] in MCI. Despite nearly all participants being classified as "amnestic MCI" by ADNI, results suggest that only one-third of the ADNI MCI cohort was solely amnestic, with another third representing primarily dysnomic or dysexecutive subtypes. It is

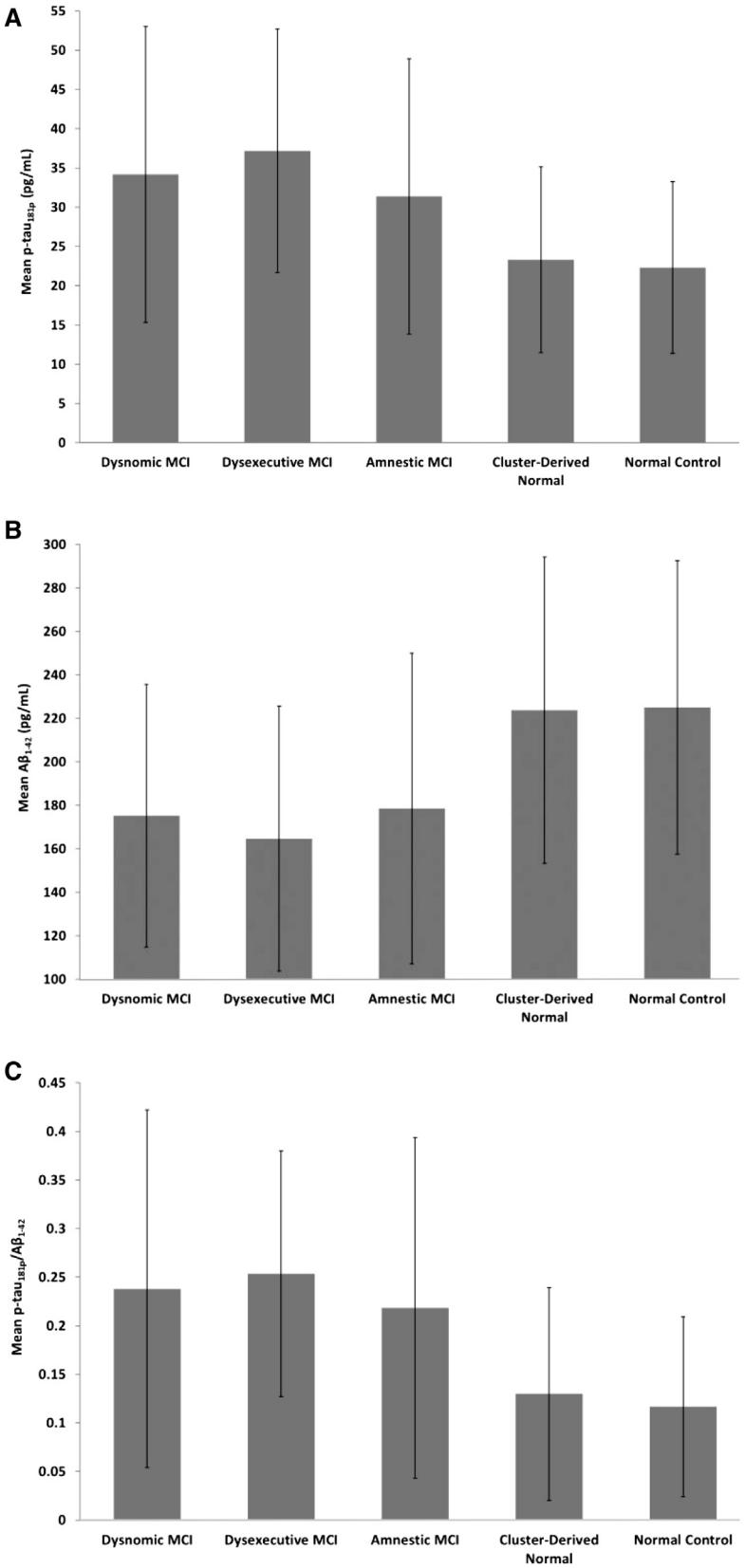

Fig. 2. CSF (A) hyperphosphorylated tau (p-tau$_{181p}$) concentrations (B) β-amyloid (Aβ_{1-42}) concentrations, and (C) ratio of p-tau$_{181p}$ to Aβ_{1-42} for the cluster groups and normal control group. Error bars denote standard deviations. MCI, mild cognitive impairment.

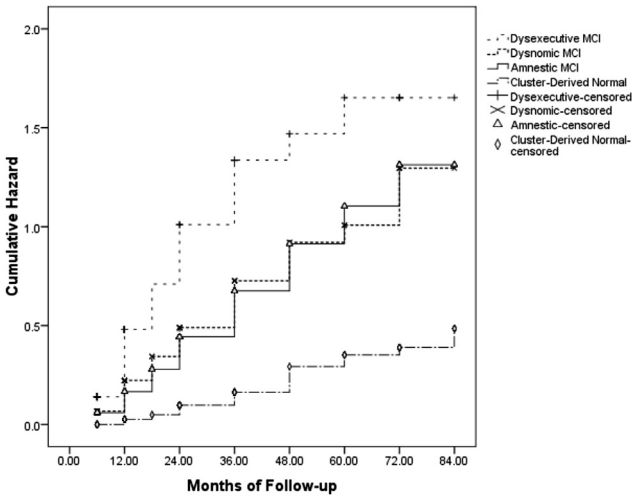

Fig. 3. Hazard function showing risk of progression to dementia across time for the cluster groups. MCI, mild cognitive impairment.

possible that combining subtypes of MCI may limit the generalizability of research findings. In addition to identifying subtypes of MCI, our results also suggest that a significant proportion of individuals in the ADNI MCI sample are cognitively normal once detailed testing is taken into account (i.e., a false-positive error in classification) and do not represent prodromal AD. It is plausible that at least a subset of the cluster-derived normal group may represent a group of individuals who are nevertheless at risk for cognitive decline and AD, particularly given their lower performance on Logical Memory and their higher prevalence of the *APOE* ε4 allele relative to the robust normal control group, although as a group, their intact performances across the neuropsychological tests indicate that a diagnosis of MCI is not warranted.

This statistical method of classifying MCI based on neuropsychological test scores resulted in a significant improvement in the specificity of the diagnosis as it identified 282 participants with potentially false-positive diagnoses. However, it was at a cost of some modest corresponding decline in sensitivity as a subset of 28 individuals (10.7%) in the cluster-derived normal group did progress to dementia over time. (Fig. 3 shows the increase in risk of dementia over time for the cluster-derived normal group due to the inclusion of these 28 participants.) Thus, their original ADNI diagnosis of MCI could be considered accurate. However, it is important to note that nearly an equal number of individuals (24 participants; 9.2%) in the cluster-derived normal group reverted to a cognitively normal classification by ADNI at follow-up, suggesting roughly equal diagnostic errors in the opposite direction. All told, our findings suggest that the very modest loss in sensitivity (i.e., 28 of the 282 participants) is far outweighed by the large gains in specificity (i.e., 254 of the 282 participants). In addition, the progression rate of the cluster-derived normal group might be best considered in the context of base rates of cognitive decline for the overall ADNI cohort. Examination of the base rate of cognitive decline in ADNI's entire normal control group of 404 participants with neuropsychological and follow-up data (not just the 284 participants retained for the robust normal group in the present study) was found to be 13%, with 2% of the normal control sample progressing to dementia and 11% progressing to MCI.

There are several possible shortcomings to the diagnostic criteria used by ADNI that could account for low specificity and large numbers of false-positive misclassifications. First, abnormal memory function was determined by a single memory score (delayed recall of story A from WMS-R Logical Memory), despite evidence showing that isolated low memory test scores are quite common in older adult populations (e.g., 39% of healthy older adults in the WMS-III standardization sample scored in the impaired range on at least one memory measure) [21–23]. Such findings emphasize the importance of considering normal variability and base rates of low memory scores in healthy older adults when interpreting a single test score. Second, only half of the Logical Memory test was administered to ADNI participants (story A), potentially diminishing its reliability. In addition, there is evidence that measures of story memory may be less sensitive to incipient dementia relative to verbal list learning tasks [24,25], suggesting that a list learning test may be a better screening measure. Third, general cognitive function was assessed only with the MMSE, a crude measure with limited ability to differentiate healthy controls versus MCI, or MCI versus AD [26]. Finally, the MCI diagnosis required a global score of 0.5 on the CDR [27]. Given the wide range of cognitive and biomarker profiles seen within the four clusters that emerged from the ADNI MCI cohort, it is clear that a global CDR score of 0.5 does not capture variability in the cognitive phenotype or the level of severity of MCI. This conclusion is supported by previous research showing that global CDR scores of 0.5 in an MCI sample masked variability in cortical thinning and activities of daily living and was not sensitive to the level of MCI severity or in predicting progression to AD [28]. Other research also shows that reliance on global CDR scores in MCI diagnosis results in a high rate of false-positive diagnostic errors [29]. The CDR may be susceptible to recall bias or influenced by psychiatric factors, and it is possible that "worried well" individuals could report enough difficulties to obtain a CDR score of 0.5 [29]. Subjective memory complaints can also be related to depressive symptoms, personality features [30], or knowledge that one carries a risk factor for AD [31]. In the present sample, the cluster-derived normal group reported more depressive symptoms than normal controls on a self-report measure of depression ($P < .001$), but there were no differences between the four cluster groups ($P > .05$). This finding supports the possibility that reliance on subjective memory complaints in diagnosis may be another source of variability that contributes to false-positive diagnostic errors.

The observed difficulties in conventional criteria for MCI diagnosis have implications for both practice and research. From a practice perspective, Diagnostic and Statistical Manual of Mental Disorders, 5th edition criteria for mild neurocognitive disorder require a "modest impairment in cognitive performance" but do not state specifically how the determination of cognitive impairment should be made ("preferably documented by standardized neuropsychological testing or, in its absence, another quantified clinical assessment"). Our results suggest that false-positive errors in diagnosis are more likely if such determination relies on a single cognitive measure, subjective complaints, or subjective rating scales, rather than based on more detailed neuropsychological evaluation. From a research perspective, findings from studies of the natural history or potential treatment of MCI could be diluted or obscured by the inclusion of individuals who are better classified as cognitively normal by a more thorough sampling of neuropsychological functions (i.e., false-positive diagnostic errors). These implications will only assume greater importance as studies begin to examine "preclinical" AD [32] and assign such diagnoses based on fine-grained distinctions of "subtle cognitive declines."

With regard to the three cognitively impaired MCI subgroups, the groups were similar in terms of *APOE* status, CSF AD biomarkers, and performance on measures used in ADNI's diagnosis (e.g., MMSE, CDR, Logical Memory). However, the dysexecutive MCI group was older, demonstrated impairment in multiple cognitive domains, had higher percentages of individuals with positive CSF AD biomarkers (when CSF was used as a dichotomous measure), and showed the highest rate of progression to dementia compared with the dysnomic and amnestic groups. It is not clear whether this group represents a more "severe" form MCI or whether primary deficits in attention/executive functioning are impacting performance in other cognitive domains. Additional research is needed to explore whether these different cognitive phenotypes of MCI are associated with distinct clinical outcomes.

Strengths of the present study include a large well-characterized sample and use of robust norms [33] that were age-and education-adjusted and derived from a sample of normal control participants that excluded individuals with preclinical dementia (based on 1–7 years of follow-up). A limitation of the present study was the lack of visuospatial measures in the cluster analyses, particularly since a "visuospatial" MCI subgroup was identified by Clark et al. [4]. If this additional cognitive domain had been included, it is possible that some of the 28 individuals in the cluster-derived normal group who ultimately progressed to AD might have been identified as belonging to a non-normal cluster. The possibility that including more or different neuropsychological measures could modify cluster solutions and potentially identify more individuals at risk for progression to AD will be explored in future studies, in addition to examining the effect of different normative reference methods on cluster solutions. Another future direction will be to compare the conventional MCI diagnostic criteria to actuarial neuropsychological MCI criteria put forth by Jak et al. [34] to determine whether this method reduces the number of false-positive diagnostic errors. The overarching aim of these efforts is to improve diagnostic accuracy and better characterize distinct prodromal cognitive phenotypes, as the determination of biomarkers cannot substitute for accurate characterization of the clinical syndrome of MCI or prodromal AD. It is hoped that improving diagnostic accuracy will enhance biomarker study findings, opportunities for earlier interventions, and better clinical decision making.

Acknowledgments

This work was supported by NIH grants R01 AG012674 (M.W.B.), K24 AG026431 (M.W.B.), P50 AG05131 (D.G.), R01 AG16495 (R.A.), and an Alzheimer's Association grant NIRG-13-281806 (C.R.M.). 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 ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer Drug Discovery Foundation; BioClinica, Inc; Biogen Idec; Bristol-Myers Squibb; Eisai Inc; Elan Pharmaceuticals, Inc; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech; GE; 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; Pfizer; Piramal Imaging; Servier; Synarc Inc; and Takeda Pharmaceuticals North America. 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 Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

M.W.B. serves as an associate editor for the *Journal of the International Neuropsychological Society.* D.G. serves as an editor for *Alzheimer's Research and Therapy* and as a paid consultant on Data Safety Monitoring Boards for Pfizer, Elan, and Balance Pharmaceuticals, Inc. D.P.S. serves as a

consultant for CHDI Foundation, Novartis, and Bristol-Meyers Squibb. The other authors report no disclosures.

RESEARCH IN CONTEXT

1. Systematic review: The authors searched PubMed for studies related to misdiagnosis or misclassification of mild cognitive impairment. Review of the literature revealed that the conventional MCI diagnostic criteria are susceptible to errors. Specifically, isolated low scores on cognitive measures can result in false-positive errors. The use of subjective memory complaints can elevate both false-positive and false-negative rates of MCI diagnoses.
2. Interpretation: Our study supports previous findings showing high rates of diagnostic errors based on conventional criteria, as one-third of our sample was misclassified as MCI. Results further show that this misdiagnosed subgroup had different CSF profiles, *APOE* allelic frequencies, and rates of progression to dementia compared with other MCI subtypes.
3. Future directions: Future research is needed to improve diagnostic accuracy and better characterize distinct prodromal cognitive phenotypes, including determining whether more comprehensive neuropsychological assessment reduces the number of false-positive diagnostic errors.

References

[1] Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004;256:183–94.
[2] Petersen RC, Morris JC. Mild cognitive impairment as a clinical entity and treatment target. Arch Neurol 2005;62:1160–3.
[3] Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO, et al. Mild cognitive impairment—beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. J Intern Med 2004; 256:240–6.
[4] Clark LR, Delano-Wood L, Libon DJ, McDonald CR, Nation DA, Bangen KJ, et al. Are empirically derived subtypes of mild cognitive impairment consistent with conventional subtypes? J Int Neuropsychol Soc 2013;19:1–11.
[5] Delano-Wood L, Bondi MW, Sacco J, Abeles N, Jak AJ, Libon DJ, et al. Heterogeneity in mild cognitive impairment: differences in neuropsychological profile and associated white matter lesion pathology. J Int Neuropsychol Soc 2009;15:906–14.
[6] Libon DJ, Xie SX, Eppig J, Wicas G, Lamar M, Lippa C, et al. The heterogeneity of mild cognitive impairment: a neuropsychological analysis. J Int Neuropsychol Soc 2010;16:84–93.
[7] Ganguli M, Snitz BE, Saxton JA, Chang CC, Lee CW, Vander Bilt J, et al. Outcomes of mild cognitive impairment by definition: a population study. Arch Neurol 2011;68:761–7.
[8] Koepsell TD, Monsell SE. Reversion from mild cognitive impairment to normal or near-normal cognition. Neurology 2012;79:1591–8.
[9] Nordlund A, Rolstad S, Klang O, Edman A, Hansen S, Wallin A. Two-year outcome of MCI subtypes and aetiologies in the Goteborg MCI study. J Neurol Neurosurg Psychiatry 2010;81:541–6.
[10] Summers MJ, Saunders NL. Neuropsychological measures predict decline to Alzheimer's dementia from mild cognitive impairment. Neuropsychology 2012;26:498–508.
[11] Tyas SL, Salazar JC, Snowdon DA, Desrosiers MF, Riley KP, Mendiondo MS, et al. Transitions to mild cognitive impairments, dementia, and death: findings from the nun study. Am J Epidemiol 2007;165:1231–8.
[12] Petersen RC, Aisen P, Boeve BF, Geda YE, Ivnik RJ, Knopman DS, et al. Criteria for mild cognitive impairment due to Alzheimer's disease in the community. Ann Neurol 2013. http://dx.doi.org/10.1002/ana.23931 [Epub ahead of print].
[13] Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology 2010;74:201–9.
[14] Wechsler D. Wechsler memory scale-revised. New York: The Psychological Corporation; 1987.
[15] Landau SM, Harvey D, Madison CM, Reiman EM, Foster NL, Aisen PS, et al. Comparing predictors of conversion and decline in mild cognitive impairment. Neurology 2010;75:230–8.
[16] Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Rimmler JB, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993;261:921–3.
[17] Mayeux R, Stern Y, Ottman R, Tatemichi TK, Tang MX, Maestre G, et al. The apolipoprotein epsilon 4 allele in patients with Alzheimer's disease. Ann Neurol 1993;34:752–4.
[18] Saunders AM, Strittmatter WJ, Schmechel D, George-Hyslop PH, Pericak-Vance MA, Joo SH, et al. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. Neurology 1993;43:1467–72.
[19] Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in Alzheimer's Disease Neuroimaging Initiative subjects. Ann. Neurol 2009;65:403–13.
[20] Nettiksimmons J, Decarli C, Landau S, Beckett L. Biological heterogeneity in ADNI amnestic mild cognitive impairment. Alzheimers Dement 2014. http://dx.doi.org/10.1016/j.jalz.2013.09.003 [Epub ahead of print].
[21] Brooks BL, Iverson GL, White T. Substantial risk of "accidental MCI" in healthy older adults: base rates of low memory scores in neuropsychological assessment. J Int Neuropsychol Soc 2007;13:490–500.
[22] Brooks BL, Iverson GL, Holdnack JA, Feldman HH. Potential for misclassification of mild cognitive impairment: a study of memory scores on the Wechsler Memory Scale-III in healthy older adults. J Int Neuropsychol Soc 2008;14:463–78.
[23] Palmer BW, Boone KB, Lesser IM, Wohl MA. Base rates of "impaired" neuropsychological test performance among healthy older adults. Arch Clin Neuropsychol 1998;13:503–11.
[24] Rabin LA, Paré N, Saykin AJ, Brown MJ, Wishart HA, Flashman LA, et al. Differential memory test sensitivity for diagnosing amnestic mild cognitive impairment and predicting conversion to Alzheimer's disease. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2009; 16:357–76.
[25] Tremont G, Miele A, Smith MM, Westervelt HJ. Comparison of verbal memory impairment rates in mild cognitive impairment. J Clin Exp Neuropsychol 2010;32:630–6.
[26] Mitchell AJ. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. J Psychiatr Res 2009;43:411–31.
[27] Morris JC. Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993;43:2412–4.

[28] Chang YL, Bondi MW, McEvoy LK, Fennema-Notestine C, Salmon DP, Galasko D, et al. Global clinical dementia rating of 0.5 in MCI masks variability related to level of function. Neurology 2011;76:652–9.
[29] Saxton J, Snitz BE, Lopez OL, Ives DG, Dunn LO, Fitzpatrick A, et al. Functional and cognitive criteria produce different rates of MCI and conversion to dementia. J Neurol Neurosurg Psychiatry 2009;80:737–43.
[30] Reid LM, MacLullich AM. Subjective memory complaints and cognitive impairment in older people. Dement Geriatr Cogn Disord 2006; 22:471–85.
[31] Lineweaver TT, Bondi MW, Galasko D, Salmon DP. Effect of knowledge of APOE genotype on subjective and objective memory performance in healthy older adults. Am J Psychiatry 2014;171:201–8.
[32] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7:280–92.
[33] Sliwinski M, Lipton RB, Buschke H, Stewart W. The effects of preclinical dementia on estimates of normal cognitive functioning in aging. J Gerontol B Psychol Sci Soc Sci 1996;51:P217–25.
[34] Jak AJ, Bondi MW, Delano-Wood L, Wierenga C, Corey-Bloom J, Salmon DP, et al. Quantification of five neuropsychological approaches to defining mild cognitive impairment. Am J Geriatr Psychiatry 2009;17:368–75.