



Published in final edited form as:

*J Int Neuropsychol Soc.* 2010 July ; 16(4): 721–729. doi:10.1017/S1355617710000512.

## Hierarchical cognitive and psychosocial predictors of amnestic mild cognitive impairment

S. DUKE HAN<sup>1</sup>, HIDEO SUZUKI<sup>2</sup>, AMY J. JAK<sup>3,4</sup>, YU-LING CHANG<sup>3</sup>, DAVID P. SALMON<sup>5</sup>, and MARK W. BONDI<sup>3,4</sup>

<sup>1</sup> Department of Behavioral Sciences, Rush University Medical Center, Chicago, Illinois

<sup>2</sup> Department of Biology, Loyola University Chicago, Chicago, Illinois

<sup>3</sup> Department of Psychiatry, University of California San Diego School of Medicine, San Diego, California

<sup>4</sup> Psychology Service, VA San Diego Healthcare System, San Diego, California

<sup>5</sup> Department of Neuroscience, University of California San Diego School of Medicine, San Diego, California

### Abstract

To identify neuropsychological and psychosocial factors predictive of amnestic Mild Cognitive Impairment (aMCI) among a group of 94 nondemented older adults, we employed a novel nonlinear multivariate classification statistical method called Optimal Data Analysis (ODA) in a dataset collected annually for 3 years. Performance on measures of memory and visuomotor processing speed or symptoms of depression in year 1 predicted aMCI status by year 2. Performance on a measure of learning at year 1 predicted aMCI status at year 3. No other measures significantly predicted incidence of aMCI at years 2 and 3. Results support the utility of multiple neuropsychological and psychosocial measures in the diagnosis of aMCI, and the present model may serve as a testable hypothesis for prospective investigations of the development of aMCI.

### Keywords

Amnestic mild cognitive impairment; aMCI; MCI; Neuropsychology; Memory; Visuomotor processing speed; D-KEFS; Depression; Optimal data analysis

## INTRODUCTION

Mild Cognitive Impairment (MCI) has garnered much attention in dementia research for its implication as a prodromal stage of Alzheimer's disease (AD) (see Morris, 2005). Since its establishment as an amnestic syndrome in the presence of otherwise intact cognition and ability to execute activities of daily living (Petersen et al., 1999), this well-studied condition has been revised to address and incorporate single-domain and multiple-domain deficits in cognitive abilities other than memory (Peterson & Morris, 2005). The revision therefore yielded four possible MCI conditions: single-domain amnestic, multiple-domain amnestic, single-domain nonamnestic, and multiple-domain nonamnestic. Research suggests that

amnestic MCI (aMCI) patients convert to AD at a rate of 16–41% per year (Gauthier et al., 2006) as opposed to a rate of 1–2% per year in the general population (Petersen et al., 2001). Some propose research criteria for very early AD that rely on a core diagnostic criterion of early episodic memory impairment, supportive features such as the presence of medial temporal lobe atrophy or abnormal cerebrospinal fluid markers, and exclusionary criteria like depression or sudden onset of symptoms (Dubois et al., 2007). Thus, the study of aMCI and its relationship to cognitive decline remains an important focus of neuropsychological inquiry.

We employed a novel nonlinear multivariate classification statistical method called Optimal Data Analysis (ODA; Yarnold & Soltysik, 2005) with the aim of identifying factors in the prediction of aMCI. Our prior work (Jak et al., 2009), as well as the work of others (see Twamley et al., 2006, for a review), suggests that specific performances on standardized clinical measures of memory, such as the Wechsler Memory Scale – Revised edition (WMS-R) Logical Memory and the California Verbal Learning Test – Second edition (CVLT-II), are highly predictive of aMCI status within a group of premorbidly nondemented older adults.

## METHOD

### Participants and Materials

All human data included in this article were obtained in compliance with regulations of the Internal Review Board of the University of California San Diego. Ninety-four participants were recruited by advertisements through various media sources in and around San Diego, CA (see Table 1). These participants were enrolled in a longitudinal aging study and had been tracked for three years. All were asked to complete an annual battery of psychosocial measures and neuropsychological tests. Participants were assessed for, and when appropriate diagnosed with, aMCI according to criteria delineated in Jak et al (2009). The Jak et al. (2009) method for assigning aMCI diagnoses is based on six variables (age-scaled scores of LMI, LMII, VRI, VRII, and CVLT Trials 1–5 Total and CVLT Long Delay Free Recall standard scores). If participants' performances on at least two of the memory measures fell one or more standard deviations below their age appropriate norms (i.e., single-domain aMCI), or if participants met criteria for a deficit in one or more cognitive domains in addition to single-domain aMCI (i.e., multiple-domain aMCI), the participants were classified as aMCI. Also, the participants with a deficit in one or more cognitive domains in the absence of memory problems (i.e., nonamnestic subtypes of MCI) were excluded from the analysis. Otherwise, participants were classified as "no MCI." At the initial wave of the longitudinal study, no participant qualified for a diagnosis of aMCI or AD. At the time of this investigation, 52 participants had completed the second wave, and 35 of these also had completed the third wave.

The demographic information, genetic measures (apolipoprotein E genotype), psychosocial measures, and neuropsychological tests that comprised the battery included: age, education, gender, apolipoprotein E genotype, the Logical Memory (LM) subtest and the Visual Reproduction (VR) subtest from the Wechsler Memory Scale–Revised edition (WMS-R), the California Verbal Learning Test–Second edition (CVLT-II), the Dementia Rating Scale (DRS), the Digit Span and Block Design subtests from the Wechsler Adult Intelligence Scale–Revised edition (WAIS-R), Trials A and B, the Draw-A-Clock test, the Boston Naming Test (BNT), Verbal Fluency, Category Fluency, Color-Word Interference, Tower Test, Sorting Test, and Trail-Making Test from the Delis-Kaplan Executive Functions System (D-KEFS), the 48-card version of the Wisconsin Card Sorting Test (WCST), the American National Adult Reading Test (ANART), the Independent Living Scale (ILS), and the Geriatric Depression Scale (GDS). In addition, the participants were asked to submit to a

cheek buccal swabbing to determine their APOE allele genotype (see Saunders, Strittmatter, & Schmechel, 1993). In the ODA statistical analyses, all of the above measures collected at the first wave were used as the independent variables to predict the occurrence of aMCI at the second wave. Furthermore, the measures assessed at the first and second waves were examined to predict the occurrence of aMCI at the third wave. The dependent variable was the diagnosis of aMCI at the second and third waves, respectively.

### Analysis Strategy

Optimal Data Analysis (ODA) was used to explore whether there were any demographic (including APOE genotype), psychosocial, or neuropsychological factors that predicted diagnosis of aMCI in the second and third waves. The specific variables included in the analysis are listed in the Appendix. ODA was performed by the Windows-based computer analysis software (Yarnold & Soltysik, 2005). This nonlinear multivariate classification method provides a hierarchical classification tree model in which cases are categorized into each group of a dichotomous dependent variable (“aMCI” or “no MCI” in the current study) by pathways branched by independent variables or “nodes.” An advantage of ODA is that there are no necessary assumptions such as multivariate normality, additivity, equality of group sizes, number of variables, or multicollinearity (see Yarnold, Soltysik, & Bennett, 1997, for details).

ODA refers to an independent variable as an *attribute* and a dependent variable as a *class variable* (Soltysik & Yarnold, 1993; Yarnold & Soltysik, 2005). The class variable must be categorical (either dichotomous or multicategorical), whereas attributes may have any scale of measurement. ODA first sets the best categorical borderline for each attribute, called *cutpoint* or *decision rule*, which classifies cases with the maximum percentage accuracy (*percentage accuracy in classification* or PAC) into each category of a class variable. ODA uses a special index, called *effect strength for sensitivity* (ESS), to indicate the percentage of how many cases belonging to a group are correctly classified. In other words, higher ESS indicates that an obtained cutpoint achieves higher PACs in classifying cases into each category. Next, ODA employs a *leave-one-out* (LOO) *validity* approach to evaluate the stability of classification performance. This entails repeatedly analyzing classification performance and checking its consistency across subsamples every time one observation is occasionally excluded. Finally, to evaluate the significance level of classification performance, Fisher’s exact probability test is used.

An attribute that shows the highest ESS, LOO stability, and significant *p*-value is considered the strongest attribute, which is entered as the top *node* of the hierarchical tree model (Soltysik & Yarnold, 1993; Yarnold & Soltysik, 2005). Once the top attribute is selected, the same procedure is performed again within a subsample classified by the top attribute. Consequently, the model gradually builds a tree of several nodes branched out from the top attribute. If there is no significant attribute, the classification performance is stopped. To finalize the classification tree model, the significance levels of all attributes are retested by a sequentially rejecting Sidak Bonferroni-type multiple comparisons procedure. The purposes of this procedure are to control Type I error rate per comparison and maximize statistical power. If any significance levels are beyond *p*-value per comparison, these attributes are pruned from the model.

Lastly, it should be noted that, in spite of its unique approach being different from traditional classification methods, the indices used by ODA are compatible with traditional classification method indices, such as the goodness-of-fit index, effect size, and significance level. Therefore, models produced by ODA may be tested according to these parameters. For example, the goodness-of-fit index is comparable to overall classification accuracy, the

effect sizes can be calculated by ESS or overall effect strength in ODA, and the significance level is tested by Fisher's exact probability test.

## RESULTS

There were 8 participants categorized as aMCI (5 single-domain) at the second wave, and 5 categorized as aMCI at the third wave (2 single-domain). Three cases from the second wave and one case from the third wave were dropped in accordance with the pairwise deletion method, because these cases had missing data on measures that were significant in the model (i.e., WMS-R LMII % retention, D-KEFS Trail-Making Number Sequencing scaled score, Geriatric Depression Scale score, and WMS-R LMI MOANS age standard score). Figures 1a and 1b summarize the ODA hierarchical classification tree model of baseline data to predict the occurrence of aMCI at the second wave of the longitudinal study. Forty-nine participants entered into the model as the result of a pairwise deletion method, and overall classification accuracy was 93.88% ( $p < .001$ ) with an overall effect strength of 79.85%. These values indicate that our model was strongly predictive (see Table 2; for the method to evaluate effect strength, see Yarnold & Soltysik, 2005). Figure 1 depicts that the classification tree model predicted the development of aMCI with 87.5% accuracy; the participants were highly likely to develop aMCI at the second wave if their memory retention rate on WMS-R LM Delayed Recall *versus* Immediate Recall was lower than or equal to 78.5% at the first wave, and if they had a scaled score of less than or equal to 14.5 on D-KEFS Trail-Making Number Sequencing scale at the first wave. On the other hand, if the participants scored higher than 78.5% of their memory retention rate on WMS-R LMII at the first wave, aMCI was less likely to occur at the second wave with 94.74% accuracy. In addition, even if the memory retention rate was lower than or equal to 78.5% on WMS-R LMII at the first wave, a higher score than 14.5 on the D-KEFS Trail-Making Number Sequencing scale at the first wave predicted the low likelihood of the occurrence of aMCI at the second wave with 100% accuracy.

It was also found that the occurrence of aMCI at the second wave was predicted with the same classification accuracy if the Geriatric Depression Scale (GDS) score was used as the second predictor (see Figure 1b). In this case, the first attribute was still memory retention rate on WMS-R LMII, such that a higher score than 78.5% of their memory retention rate predicted a low likelihood of developing aMCI at the second wave with 94.74% accuracy. On the other hand, if memory retention rate was lower than 78.5%, GDS alternatively predicted the likelihood of developing aMCI in the following way: A participant was less likely to develop aMCI at the second wave if their GDS score was less than or equal to 2.5; otherwise, a participant was likely to develop aMCI at the second wave. Note that both Figures 1a and 1b predicted the occurrence of aMCI with the same accuracy of classification performance.

The predictors of the development of aMCI two years later were also examined by ODA. The ODA hierarchical classification tree model for this prediction is more parsimonious with greater classification accuracy than the first model (see Figure 1c). If participants had a score lower than 8.5 as a Mayo's Older American Normative Scales (MOANS) age standard score on WMS-R LMI at the first wave, they were diagnosed as aMCI at the third wave; otherwise, participants did not qualify for aMCI at the third wave. Note that both prediction endpoints were predicted with 100.00% accuracy. In other words, the overall classification accuracy was 100.00% ( $p < .001$ ), and the overall effect strength was also 100.00%, which means that the model perfectly predicted the occurrence of aMCI two years later (see Table 3).

## DISCUSSION

We employed a novel nonlinear multivariate classification statistical method called Optimal Data Analysis to identify possible predictive factors of developing aMCI in a dataset of neuropsychological and psychosocial measures collected annually for three years from 94 originally nondemented participants. With this method we found that story learning or retention, visuomotor processing speed, and depression were predictive of aMCI one to two years later. No other neuropsychological or psychosocial factors predicted development of aMCI.

Two statistical classification methods have been widely utilized in the literature to conduct exploratory classification analyses: logistic regression analysis (LRA) and discriminant function analysis (DFA). However, these methods assume linearity, where the variability of human behavior is forcefully fit into a mathematical approximation. Specifically, LRA assumes a linear relationship between independent variables and the log odds of a dependent variable, whereas DFA assumes linear combinations of independent variables (i.e., discriminant functions, see Agresti, 2007 and Stevens, 2002). However, the linearity assumption presumes that all observed data should be the same in terms of (1) the set of independent variables, (2) the direction of influence (i.e., positively or negatively predictive), and (3) the coefficient values (or weight) of each independent variable (Yarnold, Soltysik, & Bennett, 1997). If these characteristics are not present, the classification accuracy level is constrained or biased (Soltysik & Yarnold, 1993; Yarnold & Soltysik, 2005). In addition to these assumptions, LRA and DFA assume (1) no gross outliers, (2) low multicollinearity of independent variables, (3) the inclusion of independent variables that are all conceptually relevant to a dependent variable, (4) equal and adequate group size, and (5) normality (Agresti, 2007; Jaccard, 2001; Menard, 1995; Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996; Tabachnick & Fidell, 1989).

In contrast to the linear classification methods, a hierarchical classification tree analysis (CTA) is a nonlinear approach (Yarnold & Soltysik, 2005; Yarnold, Soltysik, & Martin, 1994). The major methods of CTA include classification and regression tree models (e.g., CART; see Breiman, Friedman, Olshen, & Stone, 1984) and Optimal Data Analysis (ODA; Soltysik & Yarnold, 1993; Yarnold & Soltysik, 2005). These nonlinear methods show some advantages over the linear methods, especially for exploratory analyses. First, CTA theoretically provides a better classification accuracy level than the linear methods, because CTA constructs a hierarchical tree model in which a different set of independent variables with different directions and/or weights are suggested across different partitions of a given sample (i.e., no requirement of forcefully fitting variance into a mathematical estimation). This also means that CTA (1) is less sensitive to gross outliers and (2) detects an interaction effect automatically, without having to create a cross-product variable, which occur in linear classification methods (Bremner & Taplin, 2002; Fox, 2000; Sonquist & Morgan, 1964).

Furthermore, CTA repeatedly analyzes the overall effect size of each independent variable and enters only the best variable(s) into a model (Breiman et al., 1984; Soltysik & Yarnold, 1993; Yarnold & Soltysik, 2005), whereas the linear methods compute the partial effect size of each predictor simultaneously to fit all predictors into an overall model. CTA's unique approach enables (1) selection of a set of independent variables that are all statistically relevant, (2) the ability to ignore a multicollinearity of independent variables, (3) minimization of a loss of observed data by using a pairwise deletion method (rather than a listwise deletion method), and (4) examination of as many independent variables as needed.

Finally, group size is an issue for LRA and DFA because unequal group size can diminish statistical power. In contrast, regardless of group size, CTA maximizes statistical power by



using cross-validation (for CART; Breiman et al., 1984) or a sequentially rejective Sidak Bonferroni-type multiple comparisons procedure (for ODA; Soltysik & Yarnold, 1993; Yarnold & Soltysik, 2005). These procedures determine the size of a CTA model. Thus, CTA does not necessarily assume equality or adequacy of group size to maximize statistical power.

Therefore, CTA (e.g., CART and ODA) is conceptually advantageous over LRA and DFA. But, what is the difference between CART and ODA? CART relies on the least squares and maximum likelihood estimation to evaluate “impurity,” an index that indicates the heterogeneity of given categories (e.g., the Gini index, the towing index, the deviance of nodes; see Breiman et al., 1984; Clark & Pregibon, 1992; Bremner & Taplin, 2002), whereas ODA employs percentage accuracy in classification (PAC) and Fisher’s exact probability test. In other words, CART uses parametric tests as classification criteria for a given sample (i.e., the normality and linearity are assumed within a category). However, ODA does not require the assumptions of normality and linearity. Thus, Yarnold et al. (1997) believe that the nonlinear methods using the least squares/maximum likelihood (e.g., CART) “fail to maximize classification accuracy explicitly for the training sample” (p. 1452), compared to ODA, if the assumptions of normality and linearity are seriously violated within a training sample.

Previous studies revealed that ODA yielded better classification performance accuracy on predicting cardiac death (Yarnold, Soltysik, & Martin, 1994) and mortality of patients with cardiopulmonary resuscitation (Yarnold, Soltysik, Lefevre, & Martin, 1998) than LRA. For these and the reasons detailed above, ODA was selected in the present study to achieve our goal – exploring neuropsychological and other predictors of aMCI.

Our findings suggest that lower, *and not necessarily impaired*, performances on measures of story learning and memory, visuomotor processing speed, and depressive symptoms are predictive of subsequent memory decline in a normal population. These findings, at first glance, appear to be in accord with prior studies that have reported the utility of either delayed recall (Albert, Moss, Tanzi, & Jones, 2001; Arnaiz & Almkvist, 2003; Bäckman et al., 2005; Twamley et al., 2006) or learning measures (Grober & Kawas, 1997; Rabin et al., 2009) in providing strong diagnostic sensitivity for aMCI. However, it is important to note that the results showed that relatively lower scores on either WMS-R LM Delayed Recall, D-KEFS Trail-Making Number Sequencing scale, or Geriatric Depression Scale alone did not provide good predictive value of the occurrence of aMCI at follow-up visits, whereas the predictive power improved significantly when Delayed Recall and either D-KEFS Trail-Making Number Sequencing or depression scores were taken into account. Our model suggests that consideration of additional cognitive features beyond memory buttresses the prediction of progression to aMCI.

Studies of aMCI have relied almost exclusively on delayed recall or retention measures in rendering the diagnosis (Arnaiz & Almkvist, 2003). Our findings, however, suggest that the diagnosis of aMCI may be aided by the incorporation of other cognitive and psychosocial functioning measurement strategies. A number of studies have specifically shown the sensitivity of Trail-Making test procedures (Chen et al., 2001), as well as depressive features (Teng, Lu, & Cummings, 2007) in the years preceding a diagnosis of Alzheimer’s disease. As Jak and colleagues (2009) have pointed out, the use of comprehensive neuropsychological assessment when diagnosing MCI subtypes will help to improve the stability and reliability of diagnosis, as will the use of multiple measurements within a cognitive domain, such as episodic memory. These results may suggest that the conventional practice of relying solely on the use of a delayed recall or retention measure, or rating scale summaries of a single delayed recall measure, may lead to more false positive errors (i.e.,

misdiagnosing healthy individuals as aMCI; Saxton et al., 2009) than using a procedure based on multiple measures.

Of particular note is the fact that apolipoprotein E (APOE) genotype and gender were not predictive of aMCI in our sample. The APOE genotype, more specifically possession of the epsilon 4 allele, has been associated with earlier age of onset of Alzheimer's disease (Corder et al., 1993) and with impairments in aMCI (Ramakers et al., 2008). However, it was not identified as a significant predictive factor in our model. Our results suggest that neurocognitive and possibly psychological factors may be more predictive of aMCI than the APOE genotype. In regard to gender, some studies have identified a gender difference in MCI incidence (e.g., Das et al., 2007), although others have not (e.g., Panza et al., 2005). Our results suggest gender is not a factor in the incidence of aMCI, at least when considering neurocognitive and psychosocial factors, supporting the refutation of gender as a risk factor for aMCI.

Limitations of the present study include potential sources of sampling error, such as demographic factors that may not be generalizable to the population as a whole. Our study group's age range was particularly circumscribed (mean = 77.23,  $SD = 7.30$ ), and our group had a relatively high level of education (mean = 15.87,  $SD = 2.49$ ). Our neuropsychological and psychosocial variables were also limited to the battery incorporated for our longitudinal study and may not have addressed factors that could have had an impact on development of aMCI (e.g., neurovascular factors). It is also unknown how many of our aMCI-diagnosed participants will progress to AD. The size of our study sample was not a limitation because ODA as a statistical approach is not limited by traditional sample size power considerations. A final limitation is that our results may be viewed as "circular" given that we examined performances on the same memory measures utilized one or two years later in the diagnosis of aMCI. We do not regard this possibility as reflecting criterion contamination given that we investigated performances on memory measures that were not used in the diagnosis of aMCI at the time that aMCI was diagnosed. In other words, even though the same tests of memory may have been used in the diagnosis of aMCI, the actual test score performances entered into our predictive model were from a different time than diagnosis (i.e., one or two years prior to diagnosis). In addition, the Jak et al. (2009) method for assigning aMCI diagnoses were based on six variables (age-scaled scores of LMI, LMII, VRI, VRII, and CVLT Trials 1–5 Total and CVLT Long Delay Free Recall standard scores), whereas our predictive models considered a total of 26 memory variables (see Appendix), six of which overlapped with the assignment method of Jak et al. (2009), although, again, the use of these six test score performances antedated the diagnosis of aMCI – which was based on different test scores from these same tests – by one to two years. As a final remedy to inspect for the possibility of criterion contamination, we again performed ODA analyses excluding those six memory measures used in the Jak et al. (2009) aMCI classification method. The resulting model trees were identical.

In conclusion, our results have interesting implications for models of the aMCI construct and provide some comparative value to the various definitional schemes recently proposed (see Petersen & Morris, 2005; Dubois et al. 2007, Jak et al. 2009). Some of the advantages of ODA as a statistical approach are that it yields specific cutpoints and a decision tree model that can be cross-validated and empirically tested in future prospective studies. Future research is needed to investigate whether these performance cutpoints in this age range are indeed predictors of aMCI and ultimately of progression to dementia.

## Acknowledgments

This work was supported by grant IIRG 07-59343 from the Alzheimer's Association (M.W.B.), and National Institute on Aging grants P30 AG10161 (S.D.H), R01 AG012674 (M.W.B.), K24 AG026431 (M.W.B.) and P50 AG05131 (D.P.S.).

## References

- Agresti, A. An introduction to categorical data analysis. 2. Hoboken, NJ: Wiley; 2007.
- Albert MS, Moss MB, Tanzi R, Jones K. Preclinical prediction of AD using neuropsychological tests. *Journal of the International Neuropsychological Society* 2001;7:631–639. [PubMed: 11459114]
- Arnaiz E, Almkvist O. Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease. *Acta Neurologica Scandinavica* 2003;179(Suppl):34–41.
- Bäckman L, Jones S, Berger AK, Laukka EJ, Small BJ. Cognitive impairment in preclinical Alzheimer's disease: A metaanalysis. *Neuropsychology* 2005;19:520–531. [PubMed: 16060827]
- Breiman, L.; Friedman, JH.; Olshen, RA.; Stone, CJ. Classification and regression trees. Belmont, CA: Wadsworth; 1984.
- Bremner AP, Taplin RH. Modified classification and regression tree splitting criteria for data with interactions. *Australian and New Zealand Journal of Statistics* 2002;44:169–176.
- Chen P, Ratcliff G, Belle SH, Cauley JA, DeKoskey ST, Ganguli M. Patterns of cognitive decline in presymptomatic Alzheimer disease: A prospective community study. *Archives of General Psychiatry* 2001;58:853–858. [PubMed: 11545668]
- Clark, LA.; Pregibon, D. Treebased models. In: Chambers, JM.; Hastie, TJ., editors. Statistical models in S. Pacific Grove, CA: Wadsworth & Brooks/Cole; 1992. p. 377-419.
- Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993;261:921–923. [PubMed: 8346443]
- Das SK, Bose P, Biswas A, Dutt A, Banerjee TK, Hazra AM, et al. An epidemiologic study of mild cognitive impairment in Kolkata, India. *Neurology* 2007;68:2019–2026. [PubMed: 17548552]
- Dubois B, Feldman HH, Jacova C, DeKoskey ST, BarbergerGateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS–ADRDA criteria. *Lancet* 2007;6:734–746.
- Fox, J. Quantitative Applications in the Social Sciences Series, No.131. Thousand Oaks, CA: Sage Publications; 2000. Multiple and generalized nonparametric regression.
- Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, et al. Mild cognitive impairment. *Lancet* 2006;367:1262–1270. [PubMed: 16631882]
- Grober E, Kawas C. Learning and retention in preclinical and early Alzheimer's disease. *Psychology and aging* 1997;12(1):183–8. [PubMed: 9100279]
- Jaccard, J. Quantitative Applications in the Social Sciences Series, No. 135. Thousand Oaks, CA: Sage Publications; 2001. Interaction effects in logistic regression.
- Jak AJ, Bondi MW, DelanoWood L, Wierenga C, CoreyBloom J, Salmon DP, Delis DC. Quantification of five neuropsychological approaches to defining mild cognitive impairment. *American Journal of Geriatric Psychiatry* 2009;17:368–375. [PubMed: 19390294]
- Menard, S. Applied logistic regression analysis. Thousand Oaks, CA: Sage Publications; 1995.
- Morris JC. Mild cognitive impairment and preclinical Alzheimer's disease. *Geriatrics* 2005;(Suppl):9–14. [PubMed: 16025770]
- Panza F, D'Introno A, Colacicco AM, Capurso C, Del Parigi A, Caselli RJ, et al. Current epidemiology of mild cognitive impairment and other predementia syndromes. *American Journal of Geriatric Psychiatry* 2005;13:633–644. [PubMed: 16085779]
- Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* 1996;49:1373–1379. [PubMed: 8970487]
- Petersen RC, Morris JC. Mild cognitive impairment as a clinical entity and treatment target. *Archives of Neurology* 2005;62:1160–1163. [PubMed: 16009779]



- 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]
- Rabin LA, Pare N, Saykin AJ, Brown MJ, Wishart HA, Flashman LA, Santulli RB. Differential memory test sensitivity for diagnosing amnesic mild cognitive impairment and predicting conversion to Alzheimer's disease. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* 2009;16:357–376. [PubMed: 19353345]
- Ramakers IHGB, Visser PJ, Aalten P, Bekers O, Slegers K, van Broeckhoven CL, et al. The association between APOE genotype and memory dysfunction in subjects with mild cognitive impairment is related to age and Alzheimer pathology. *Dementia and Geriatric Cognitive Disorders* 2008;26:101–108. [PubMed: 18617739]
- Saunders AM, Strittmatter WJ, Schmechel DE. Association of apolipoprotein E allele e4 with lateonset familial and sporadic Alzheimer's disease. *Neurology* 1993;43:1467–1472. [PubMed: 8350998]
- Saxton J, Snitz BE, Lopez OL, Ives DG, Dunn LO, Fitzpatrick A, et al. Functional and cognitive criteria produce different rates of mild cognitive impairment and conversion to dementia. *Journal of Neurology, Neurosurgery, & Psychiatry* 2009;80:737–743.
- Soltysik, RC.; Yarnold, PR. ODA 1.0: Optimal data analysis for DOS. Chicago: Optimal Data Analysis; 1993.
- Sonquist, JA.; Morgan, JN. The detection of interaction effects. Institute of Social Research, University of Michigan; Ann Arbor: 1964. Monograph No. 35
- Stevens, JP. Applied multivariate statistics for the social sciences. 4. Mahwah, NJ: Erlbaum; 2002.
- Tabachnick, BG.; Fidell, LS. Using multivariate statistics. 2. New York: Harper & Row; 1989.
- Teng E, Lu PH, Cummings JL. Neuropsychiatric symptoms are associated with progression from mild cognitive impairment to Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders* 2007;24:253–259. [PubMed: 17700021]
- Twamley E, Ropacki S, Bondi M. Neuropsychological and neuroimaging changes in preclinical Alzheimer's disease. *J Int Neuropsychol Soc* 2006;12:707–735. [PubMed: 16961952]
- Yarnold, PR.; Soltysik, RC. Optimal Data Analysis: A guidebook with software for Windows. Washington, DC: American Psychological Association; 2005.
- Yarnold PR, Soltysik RC, Bennett CL. Predicting in-hospital mortality of patients with AIDS-related *pneumocystis carinii* pneumonia: An example of hierarchically optimal classification tree analysis. *Statistics in Medicine* 1997;16:1451–1463. [PubMed: 9249918]
- Yarnold PR, Soltysik RC, Lefevre F, Martin GJ. Predicting in-hospital mortality of patients receiving cardiopulmonary resuscitation: Unit-weighted multiODA for binary data. *Statistics in Medicine* 1998;17:2405–2414. [PubMed: 9819836]
- Yarnold PR, Soltysik RC, Martin GJ. Heart rate variability and inducibility for sudden cardiac death: An example of optimal discriminant analysis. *Statistics in Medicine* 1994;13:1015–1021. [PubMed: 8073197]

## APPENDIX

### List of attributes analyzed by ODA

1. age as of test date
2. gender
3. handedness
4. examiner
5. education (yrs)
6. ethnicity

7. subject referral
8. ANART VIQ
9. ANART errors
10. WAIS-R digit span forward
11. WAIS-R digit span backwards
12. WAIS-R digit span scaled score
13. WAIS-R digit span MOANS
14. WISC-R block design raw
15. WISC-R block design *T* score
16. WISC-R block design broken configuration
17. WISC-R block design over time
18. DRS total
19. DRS total *T* score
20. DRS attention
21. DRS attention *T* score
22. DRS initiation/perseveration
23. DRS initiation/perseveration *T* score
24. DRS supermarket items
25. DRS supermarket items *T* score
26. DRS construction
27. DRS construction *T* score
28. DRS conceptualization
29. DRS conceptualization *T* score
30. DRS memory
31. DRS memory *T* score
32. ADRC form (1 or 2)
33. Boston Naming Test total correct
34. Boston Naming Test total correct *T* score
35. Boston Naming Test total correct MOANS scaled score
36. BNT spontaneous correct (total)
37. BNT stimulus cues given (total)
38. BNT stimulus cues correct (total)
39. BNT phonemic cues given (total)
40. BNT phonemic cues correct (total)
41. WCST-48 number of categories

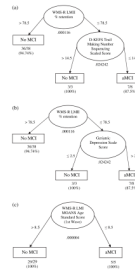
42. WCST-48 categories *T* score
43. WCST-48 nonperseverative errors
44. WCST-48 nonperseverative errors *T* score
45. WCST-48 perseverative errors
46. WCST-48 perseverative errors *T* score
47. WCST-48 set losses
48. WCST-48 total errors
49. Trails A
50. Trails A *T* score
51. Trails A MOANS
52. Trails A no. of errors
53. Trails B
54. Trails B *T* score
55. Trails B MOANS
56. Trails B no. of errors
57. draw a clock command
58. draw a clock copy
59. verbal fluency version (standard/alternate)
60. letter fluency (f)
61. letter fluency (a)
62. letter fluency (s)
63. letter fluency total raw
64. D-KEFS verbal fluency scaled score
65. letter fluency total *T* score
66. category fluency (animals) raw
67. D-KEFS category fluency scaled score
68. category fluency (animals) *T* score
69. D-KEFS color-word interference inhibition scaled score
70. D-KEFS color-word interference inhibition/switch scaled score
71. D-KEFS tower total achievement scaled score
72. D-KEFS sorting test confirmed correct sorts scaled score
73. D-KEFS sorting test sort recognition description scaled score
74. D-KEFS trail-making visual scanning scaled score
75. D-KEFS trail-making number sequencing scaled score
76. D-KEFS trail-making letter sequencing scaled score

77. D-KEFS trail-making number-letter switch scaled score
78. D-KEFS trail-making motor sequencing scaled score
79. WMS-R LMI
80. WMS-R LMI age scaled score
81. WMS-R LMI MOANS age scaled score
82. WMS-R LMII
83. WMS-R LMII age scaled score
84. WMS-R LMII MOANS age scaled score
85. WMS-R LMII % retention
86. WMS-R LMII % retention MOANS age scaled score
87. WMS-R LM recognition %
88. WMS-R LM recognition discrimination percentage
89. WMS-R LM response bias
90. WMS-R VRI
91. WMS-R VRI age scaled score
92. WMS-R VRI MOANS age scaled score
93. WMS-R VR II
94. WMS-R VR II age scaled score
95. WMS-R VR II MOANS age scaled score
96. WMS VR II % retention
97. WMS VR II % retention MOANS age scaled score
98. WMS VR II recognition
99. WMS-R VR recognition discrimination percentage
100. WMS-R VR response bias
101. ILS managing money raw
102. ILS managing money *T* score
103. ILS managing money problem-solving
104. ILS managing money information
105. ILS health and safety raw
106. ILS health and safety *T* score
107. ILS health and safety problem-solving
108. ILS health and safety information
109. Geriatric Depression Scale score
110. Geriatric Depression Scale rating
111. CVLT-II

- 112.**CVLT-II list A trials 1–5 total *T* score
- 113.**CVLT-II long delay free recall *T* score
- 114.**Overall Abilities
- 115.**Overall Attention
- 116.**Overall Language
- 117.**Overall Visuospatial Skills
- 118.**Overall Executive Functions
- 119.**Overall Memory
- 120.**Overall Living Skills
- 121.**APOE epsilon 4 positive

*Note.* All attributes listed above were collected at the first wave and the second wave, and each attribute at each wave was individually analyzed by ODA. Class variables were the diagnosis of aMCI at the second wave or the third wave.





**Fig. 1.**

Fig. 1a – 1c. (a) The Optimal Data Analysis (ODA) Hierarchical Tree Model 1 for predicting no MCI *versus* aMCI one year later based on neuropsychological and psychosocial variables ( $N = 49$ ); (b) Classification performance summary of Optimal Data Analysis prediction of aMCI one year later ( $N = 49$ ); (c) Classification performance summary of Optimal Data Analysis prediction of aMCI two years later ( $N = 34$ ).

*Note.* Ellipses represent nodes, arrow lines represent branches, and rectangles represent prediction endpoints. Numbers under each ellipse (node) indicate Fisher's exact p value for each node. Numbers next to arrows indicate the cutpoint for classifying cases into the categories (No MCI or aMCI) for each node. Finally, fractions and percentages below each prediction endpoint indicate the absolute number or percentage of the cases correctly classified.

**Table 1**

Demographic data for participants

<b>Demographic information</b>	<b>Mean</b>	<b>Standard Deviation</b>
<i>N</i>	94	
Age	77.23	7.303
Gender (M/F)	39/51	
Education	15.87	2.487
ANART VIQ	119.50	5.780
DRS Total	139.30	4.268
APOE (E4/Non-E4/Unknown)	24/60/10	

**Table 2**Classification performance summary of Optimal Data Analysis prediction of MCI one year later ( $N = 49$ )

<b>Performance Index</b>	<b>Performance Parameter</b>	
Overall classification accuracy	46/49 (93.88%)	
Sensitivity (No MCI)	39/41 (95.12%)	
Sensitivity (aMCI)	7/8 (87.50%)	
Effect strength for sensitivity	82.62%	
Predictive value (No MCI)	39/40 (97.50%)	
Predictive value (aMCI)	7/9 (77.78%)	
Effect strength for predictive value	75.28%	
Effect strength overall	78.95%	
Cross-Classification Table ( $p < .001$ )		
<b>Respondents' Actual Status</b>	<b>Respondents' Predicted Status</b>	
	<b>No MCI</b>	<b>aMCI</b>
No MCI	39	2
aMCI	1	7

*Note.* This classification performance was exactly replicated, regardless of whether the second attribute was (1) D-KEFS Trail-Making Number Sequencing scaled score or (2) Geriatric Depression Scale score. Overall classification accuracy is the percentage of the cases classified correctly. Sensitivity is the percentage of how many cases were correctly classified among cases that actually belong to a given category. Predictive value is the percentage of how many cases were correctly classified among cases that were predicted as a given category. Higher percentage indicates greater classification performance. Effect strength overall is the mean of effect strength for sensitivity and effect strength for predictive value. According to Yarnold & Soltysik (2005), the effect strength is strong ( $75\% < ES < 90\%$ ).

**Table 3**Classification performance summary of Optimal Data Analysis prediction of MCI one year later ( $N = 34$ )

<b>Performance Index</b>	<b>Performance Parameter</b>	
Overall classification accuracy	34/34 (100.00%)	
Sensitivity (No MCI)	29/29 (100.00%)	
Sensitivity (aMCI)	5/5 (100.00%)	
Effect strength for sensitivity	100.00%	
Predictive value (No MCI)	29/29 (100.00%)	
Predictive value (aMCI)	5/5 (100.00%)	
Effect strength for predictive value	100.00%	
Effect strength overall	100.00%	
Cross-Classification Table ( $p < .001$ )		
<b>Respondents' Actual Status</b>	<b>Respondents' Predicted Status</b>	
	<b>No MCI</b>	<b>aMCI</b>
No MCI	29	0
aMCI	0	5

*Note.* Overall classification accuracy is the percentage of the cases classified correctly. Sensitivity is the percentage of how many cases were correctly classified among cases that actually belong to a given category. Predictive value is the percentage of how many cases were correctly classified among cases that were predicted as a given category. Higher percentage indicates greater classification performance. Effect strength overall is the mean of effect strength for sensitivity and effect strength for predictive value. According to Yarnold & Soltysik (2005), the effect strength is very strong ( $95\% < ES$ ).