Neuropsychology
2007, Vol. 21, No. 2, 158–169

0894-4105/07/$12.00 DOI: 10.1037/0894-4105.21.2.158

Longitudinal Change in Cognitive Performance Among Individuals With Mild Cognitive Impairment

Marilyn Albert
Johns Hopkins University School of Medicine

Deborah Blacker
Massachusetts General Hospital, Harvard Medical School

Mark B. Moss
Boston University

Rudolph Tanzi
Massachusetts General Hospital, Harvard Medical School

John J. McArdle
University of Southern California

The authors used mixed-effects growth models to examine longitudinal change in neuropsychological performance over a 4-year period among 197 individuals who were either normal or had mild cognitive impairment (MCI) at baseline. At follow-up, the participants were divided into 4 groups: (a) controls: participants who were normal at both baseline and follow-up ($n = 33$), (b) stables: participants with MCI whose Clinical Dementia Rating—Sum of Boxes (CDR–SB) score did not differ between the first and last evaluations ($n = 22$), (c) decliners: participants with MCI whose CDR–SB score declined between the first and last evaluations ($n = 95$), and (d) converters: participants who received a clinical diagnosis of Alzheimer's disease during the follow-up period ($n = 47$). Only the Episodic Memory factor showed a significantly greater rate of decline over the follow-up period among the converters. Two other factors were significantly lower in converters at baseline in comparison with other groups (the executive function factor and the general knowledge factor), but the rate of decline over time did not differ. Individuals with an APOE ε4 allele scored lower on the episodic memory and executive function factors at baseline.

Keywords: longitudinal cognition, prodromal AD, MCI, memory, executive function

There is increasing evidence that the pathology of Alzheimer's disease (AD) can take many years, if not decades, to evolve (Morris et al., 2001). Because the clinical hallmark of AD is a progressive decline in cognitive function, a number of research groups have recruited nondemented individuals with mild cognitive impairment (MCI; Petersen et al., 1999) and have followed them over time with the goal of examining the nature of the cognitive changes that occur during the transitional phase between normality and frank dementia.

Among these studies, there is considerable consensus that tests of memory are significantly different among nondemented individuals with mild memory deficits who receive a diagnosis of AD on follow-up in comparison with those who also have memory problems but do not progress to AD within a few years' time (Albert, Moss, Tanzi, & Jones, 2001; Bondi et al., 1994; Chen et al., 2000; Howieson et al., 2003; Jacobs et al., 1995; Kluger, Ferris, Golomb, Mittelman, & Reisberg, 1999; Newmann, Warrington, Kennedy, & Rossor, 1994; Petersen, Smith, Ivnik, Kokmen, & Tangalos, 1994; Rubin et al., 1998; G. Small, LaRue, Komo, Kaplan, & Mandelkern, 1995; Tierney et al., 1996; Tuokko, Vernon-Wilkinson, Weir, & Beattie, 1991).

Less is known about the longitudinal nature of the cognitive changes that occur during prodromal AD. A few studies have gathered longitudinal data and demonstrated declines in memory over time (Bennett et al., 2002; Chen et al., 2001; Lange et al., 2002; Storandt, Grant, Miller, & Morris, 2002), whereas some researchers have reported that memory performance may be stable for a period of time before becoming progressively worse within a few years prior to the diagnosis of AD (Backman, Small, & Fratiglioni, 2001).

Moreover, there is less consensus about which cognitive domains, other than memory, are impaired during prodromal AD. The discrepancies among studies are due, at least in part, to the fact that few studies have examined a wide variety of cognitive domains, thus limiting the types of associations that can be found.

In addition, most studies that have examined change in cognitive function have focused on change within the cohort as a whole. Few have compared such changes among individuals who have differing trajectories of progression over time. With increasing evidence that the rate and nature of progression vary considerably in MCI (Petersen, 2004), such analyses have assumed greater importance.

We used a lengthy neuropsychological battery to examine longitudinal neuropsychological data in a large cohort of participants. At baseline, some of the participants had normal cognition, and some had evidence of memory problems but were not demented. Participants were followed annually with a clinical assessment. The neuropsychological battery was readministered after an average follow-up period of 3–4 years. We could therefore examine

Marilyn Albert, Department of Neurology, Johns Hopkins University School of Medicine; Deborah Blacker, Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School; Mark B. Moss, Department of Anatomy and Neurobiology, Boston University; Rudolph Tanzi, Department of Neurology, Massachusetts General Hospital, Harvard Medical School; John J. McArdle, Department of Psychology, University of Southern California.

Correspondence concerning this article should be addressed to Marilyn Albert, 1620 McElderry Street, Reed Hall East 2, Baltimore, MD 21205. E-mail: malbert9@jhmi.edu

the neuropsychological changes that occurred during the intervening years among participants on the basis of their trajectory of change during the follow-up period. We could also examine the relationship of these changes to demographic and genetic status.

Method

Selection of Participants

A total of 327 individuals were participants in the present study. All of them were participants in an ongoing longitudinal study of the evolution of AD. We recruited them through the print media (rather than from a clinic or other medical referral source) with advertisements indicating that a research study was seeking individuals both with and without memory difficulty.

Volunteers underwent a multistage screening procedure. The details of the screening procedures have been described elsewhere (Albert et al., 2001). Briefly, to be included in the study, participants needed to be 65 and over; have an informant who could provide information about their daily function; be free of significant underlying medical, neurologic, or psychiatric illness; and be willing to participate in the study procedures. In addition, individuals with evidence of major vascular risk factors (e.g., atrial fibrillation, insulin dependent diabetes, cerebral infarcts) were excluded. All participants were required to be either cognitively normal or nondemented but mildly impaired (i.e., to have a Clinical Dementia Rating [CDR]—Hughes, Berg, Danziger, Coben, & Martin, 1982—of either 0.0 or 0.5).

The primary goal of the study as a whole was the examination of cognitive, brain structure/function, and genetic factors involved in the prodromal phase of AD. The study population was therefore intentionally enriched with mildly impaired participants. For purposes of comparison, a smaller number of normal participants were also enrolled. The absolute number of individuals in the cohort was based on estimates of the power needed to address the primary hypotheses of interest.

At baseline, the study procedures included a medical evaluation (consisting of a physical exam and medical history, an electrocardiogram, and standard laboratory tests), a semistructured interview, neuropsychological testing, an MRI scan, a single-photon-emission computed tomography scan, and blood drawn for genetic analysis. All participants provided informed consent prior to the initiation of the study, in accordance with the requirements of the Human Research Committee of Massachusetts General Hospital (Boston, MA).

Assessment of Clinical Severity

The degree of clinical severity of the participants was evaluated by the semistructured interview. This interview generates both an overall CDR rating and a measure known as the CDR—Sum of Boxes (CDR–SB). The interview was derived from the initial subject protocol that was used in the development of the CDR scale (Hughes et al., 1982). It includes a set of questions regarding functional status (asked of the participant and a collateral source [e.g., family member, friend]) and a standardized neurologic, psychiatric, and mental status evaluation of the participant. To be sensitive to clinical impairments at the mildest end of the spectrum, we added a special set of questions to the interview and established additional guidelines for the rating procedures (Daly et al., 2000). The mean interrater reliability of the CDR ratings was high ($r = .99, p < .0001$), as was the interrater reliability of the six CDR subcategories ($r = .90$) that were used to generate the overall CDR rating (Daly et al., 2000). The CDR–SB represents the sum of the ratings in each of the six CDR subcategories. In the current study, each interview was administered by a master's- or doctoral-level clinician (e.g., psychiatrist, neuropsychologist, or physician's assistant) and was performed without knowledge of the other study procedures, including the neuropsychological test findings. The interview took approximately 1–2 hr to complete.

Longitudinal Assessment of Participants

The semistructured interview was repeated annually for each participant. The remaining study procedures were repeated in a subset of the participants on the basis of a priori criteria concerning progression in level of clinical severity (i.e., participants who crossed specific thresholds on the CDR–SB were reevaluated, and those whose CDR–SB remained stable for specific amounts of time were also reevaluated). For example, participants whose CDR–SB increased from < 2.0 to ≥ 2.0 were reevaluated following the visit at which this change was observed, and participants whose CDR–SB did not change after a 3-year interval were also reevaluated. This approach, which intentionally oversamples participants who are changing over time, has been recommended as an optimal longitudinal study design (Cnaan, Laird, & Slasor, 1997; McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002; McArdle & Nesselroade, 2003). It permits intensive evaluation of those who are changing the most and comparison with those who remain relatively stable. Modern statistical techniques have been developed to specifically allow for such a study design, which, by its nature, contains unbalanced data with the potential of bias because of incompleteness. These techniques aim to provide a robust assessment of change over time and reduce any potential bias as much as possible (Cohen, Cohen, West, & Aiken, 2003; Diggle & Kenward, 1994; Little, 1995; McArdle, Prescott, Hamagami, & Horn, 1998; McDonald, 1985; Verbeke, Molenberghs, Krickeberg, & Fienberg, 2000).

Group Characteristics at Baseline and at Follow-Up

Baseline. On the basis of their initial CDR interview, participants were divided into two groups. One group consisted of 101 participants with normal cognition (CDR = 0.0); their mean CDR–SB score was 0.009 ± 0.070. (The nonzero mean CDR–SB reflects a few individuals with slight difficulties in a domain other than memory; a score of 0.5 in the memory domain is required for an overall rating of 0.5.) The other group consisted of 226 individuals who were mildly impaired but nondemented; their mean CDR–SB score was 1.300 ± 0.770. The mean age, educational status, gender, percent *APOE*-4, Mini-Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975) score, and CDR–SB score of the groups are shown in Table 1.

As shown in Table 1, the distribution of CDR–SB scores among the mildly impaired participants was broad. At the mild end of the spectrum (i.e., CDR–SB = 0.5–1.5), many participants would not meet psychometric cut-offs commonly used to select MCI participants in epidemiological studies and clinical trials (Davis & Rockwood, 2004; Petersen et al., 2005). The participants at the

more impaired end of the spectrum (i.e., CDR–SB $\geq$ 2) are comparable with MCI participants recruited from these settings, on the basis of the likelihood of progression to a diagnosis of AD (Daly et al., 2000). We use MCI here to refer to the entire group of mildly impaired participants.

Follow-up. For those who remain alive, the annual follow-up rate is high (i.e., 93%). As enrollment in the study has occurred over many years, there is considerable variation in the number of years of follow-up and thus in the number of participants who have been readministered the neuropsychological battery and other study procedures. Those individuals who have had neuropsychological testing at two or more points in time were divided into four groups on the basis of their trajectory of functional change over the intervening years:

1. Controls (n = 33): Participants who had normal cognition at baseline (CDR = 0) and continued to be categorized as normal (CDR = 0) at follow-up.
2. Stables (n = 22): Participants who, on the basis of their CDR–SB score, had mild memory difficulty at baseline and remained stable over time (i.e., the CDR–SB scores of these participants ranged from 0.5 to 3.0 at baseline, and the CDR–SB scores of each participant did not change over the follow-up period). Thus, although the participants were mildly impaired at baseline, they did not demonstrate a decline in functional status on follow-up.
3. Decliners (n = 95): Participants who, on the basis of an increase in their CDR–SB score, demonstrated functional decline over time but did not meet clinical criteria for dementia during the follow-up period, as described below (i.e., the CDR–SB scores of these participants ranged from 0.0 to 2.5 at baseline; thus, they were both normal and mildly impaired at baseline).
4. Converters (n = 47): Study participants who developed significant functional impairment over time and met clinical criteria for AD during the follow-up interval. On follow-up, these participants received a consensus diagnosis to determine whether they had sufficient impairment for a diagnosis of dementia, and, if so, whether the dementia was consistent with research criteria for AD (McKhann et al., 1984) or another known diagnostic entity (e.g., frontotemporal dementia, multi-infarct dementia). Diagnoses were based on a combination of clinical history, medical records, laboratory evaluation, and neuroimaging studies (i.e., the CDR–SB scores of these participants ranged from 0.0 to 2.5 at baseline).

Table 1
Descriptive Statistical Information From the Initial Visit of the Entire Sample at Baseline Testing

Variable	Overall	Group 1: Normal	Group 2: Questionable
Sample size	327	101	226
Years of age	72.40 (5.50)	71.40 (4.60)	72.90 (5.80)
Years of education	15.50 (2.90)	15.60 (2.80)	15.40 (2.90)
Percent female	57.0%	58.8%	56.2%
Percent *APOE*-4	28.4%	24.1%	30.2%
MMSE	29.20 (1.10)	29.50 (0.70)	29.10 (1.30)
CDR–SB	0.91 (0.90)	0.01 (0.07)	1.32 (0.78)

Note. Standard deviations are listed in parentheses. *APOE*-4 = *Apolipoprotein* ε4; MMSE = Mini-Mental State Exam; CDR–SB = Clinical Dementia Rating—Sum of Boxes.

Over the course of the longitudinal study, 15 participants have come to autopsy; 9 of those 15 had a clinical diagnosis of probable AD, and this diagnosis was confirmed in 6 of the participants on the basis of the National Institute on Aging and Reagan Institute Working Group (1997) criteria for AD. An additional 3 participants showed AD changes insufficient for an autopsy diagnosis of definite AD. Thus, the converters group consisted of individuals with either a clinical or a pathological diagnosis of AD on follow-up.

Individuals selected for repeat neuropsychological testing. Table 2 provides baseline data concerning each of the follow-up groups outlined above, including age, education, gender, percent *APOE*-4, MMSE score, and CDR–SB score. The mean interval between baseline and follow-up neuropsychological testing is also presented.

Individuals not selected for repeat neuropsychological testing. It should be noted that a substantial number of participants (n = 130) have not yet been asked to repeat the neuropsychological testing because they have not met the a priori criteria for reassessment required by the study design. Descriptive statistics for this subsample of individuals are presented in Table 3. To determine if there were systematic differences between the sample with repeated testing and the subsample not yet reevaluated, we fit a series of hierarchical logistic regression models. We chose this approach as opposed to multiple t tests because the variables of interest are correlated. The logit models were based on the prediction of whether participants met the study criteria for readministration of the neuropsychological battery (yes or no) from the primary variables available at the baseline testing. The following variables were entered in the models, in order: demographic variables, CDR–SB and MMSE scores, and *APOE* ε4 status (present vs. absent). All models exhibited significant overall fit but only small explained variance (pseudo R^2 = .03 to .05). The only variable with a significant coefficient in any model reflected the small initial differences in educational level between those with repeated testing and those without (15.2 years vs. 15.8 years; see Table 3). Although the selection bias is relatively small, baseline data from the subsample with only one data point were included in the analyses of the baseline data presented below. The inclusion of participants with baseline data only improves the estimates of model parameters and is consistent with modern approaches to longitudinal mixed model analysis (Cnaan et al., 1997; McArdle et al., 2002).

Neuropsychological Assessment

All participants were administered a neuropsychological battery consisting of 22 tests (yielding 23 test scores) including (a) five memory tests: California Verbal Learning Test (Delis, Kramer, Kaplan, & Ober, 1987), Free and Cued Selective Reminding Test (Grober & Buschke, 1987), Rey Ostereith Complex Figure Test (delayed recall of 30 min; Rey, 1941), Delayed Word Recall Test (Knopman & Ryberg, 1989), and Visual Reproduction subtest of the Wechsler Memory Scale (Wechsler, 1945); (b) six tests of

Table 2
Descriptive Statistical Information for the Total Follow-Up Sample and the Four Groups at Baseline

Variable	Total sample	Group 1: Normal	Group 2: Stables	Group 3: Decliners	Group 4: Converters
Sample size	197	33	22	95	47
Years of age	71.90 (5.40)	70.30 (4.00)	69.70 (6.00)	72.10 (5.40)	73.60 (5.40)
Years of education	15.20 (2.80)	15.10 (2.70)	14.50 (2.90)	15.80 (2.60)	14.60 (2.90)
Percent female	55.8%	69.7%	59.1%	53.4%	48.9%
Percent *APOE*-4	28.4%	25.5%	26.2%	26.3%	40.0%
MMSE	29.30 (1.10)	29.50 (0.70)	29.50 (0.8)	29.50 (0.90)	28.60 (1.60)
CDR–SB	0.88 (.88)	0.00 (0.00)	0.93 (0.73)	0.74 (0.61)	1.77 (0.93)
Years between testing	3.30 (1.40)	3.80 (1.10)	3.10 (0.70)	3.10 (1.50)	2.80 (2.00)

Note. Standard deviations are listed in parentheses. *APOE*-4 = *Apolipoprotein* ε4; MMSE = Mini-Mental State Exam; CDR–SB = Clinical Dementia Rating—Sum of Boxes.

executive function (EF): Trail Making Test–Part B (Reitan, 1958), Stroop Interference Test (Stroop, 1935), Self-Ordering Test (Petrides & Milner, 1982), Porteus Mazes (Porteus, 1959), Alpha Span Test (Craik, 1986), and Digit Span Backward (Wechsler, 1988); (c) three language tests: Controlled Word Association Test for letters and for categories (Benton & Hamsher, 1976) and 15 items from the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1982); (d) two tests of spatial ability: Rey Complex Figure copying (Rey, 1941) and Wechsler Memory Scale figure copying (Wechsler, 1945); (e) three tests of sustained attention: Digit Span Forward (Wechsler, 1988), Trail Making Test–Part A (Reitan, 1958), and Cued Reaction Time (Baker, Letz, & Fidler, 1985); and (f) three subtests of the Wechsler Adult Intelligence Scale—Revised to assess general knowledge (GK; Vocabulary, Information, and Similarities; Wechsler, 1945).

Genetic Assessment

The *APOE* gene was also examined in the participants because the ε4 allele of this gene is overrepresented in AD patients in comparison with the general population (Saunders et al., 1993) and is now widely recognized as a risk factor for AD. *APOE* genotypes were determined using the methods of Hixson and Vernier (1990), as previously described (Blacker et al., 1997). We sought to determine whether *APOE* ε4 status among the groups mentioned above was related to neuropsychological performance at baseline or to change over time.

Table 3
Comparative Statistics for the Longitudinal Subsample and the Nonrepeated Subsample at Baseline Testing

Variable	Overall	Longitudinal subsample	Nonrepeated subsample
Sample size	327	197	130
Years of age	72.40 (5.50)	72.10 (5.40)	73.20 (5.60)
Years of education	15.50 (2.90)	15.20 (2.80)	15.80 (3.00)
Percent female	57.0%	56.1%	59.1%
Percent *APOE*-4	28.4%	29.6%	26.4%
MMSE	29.20 (1.10)	29.20 (1.20)	29.10 (1.20)
CDR-SB	0.91 (0.90)	0.90 (0.90)	0.97 (0.90)

Note. Standard deviations are listed in parentheses. *APOE*-4 = *Apolipoprotein* ε4; MMSE = Mini-Mental State Exam; CDR–SB = Clinical Dementia Rating—Sum of Boxes.

Longitudinal Mixed-Effects Growth Curve Modeling

Longitudinal mixed-effects growth modeling was used to examine changes among the individuals over time. This modeling approach is based on the statistical method termed *multilevel, random coefficients,* or *latent growth models* (McArdle & Nesselroade, 2003). These statistical procedures are used to fit a model directly to the observed scores on the basis of various kinds of mathematical forms of growth (the statistical details of these methods are presented in the Appendix). Standard forms of mixed-effects growth models are based on a trajectory over age. Unlike the standard calculation of rates of change, this model allows us to consider unobserved or latent scores representing an individual's initial level, an individual's latent slope over age, and unobserved but independent residual errors of measurement at each occasion. In this latent growth model, the age at any particular time is not considered as a random variable—instead the changes over age in the scores are considered as random variables, and the average group changes are described by the fixed parameters. This is an age-basis model that implies that the score measured on a person at each age gives us some indication of their likely scores at the ages not measured, and persons measured at specific ages represent the age-based scores for anyone. The model is also designed to take the time interval between measurements into account by using the exact age at testing as the basis of the changes.

In the typical model fitted here, we also fit second-order models as a set of regressions in which the latent initial levels (the expected scores at age 70) and slopes (the expected changes per decade) were impacted by other variables (e.g., group membership, education, *APOE* status; Cohen et al., 2003). From these parameters, we calculated other informative statistics about individual differences between persons as well as goodness-of-fit (chi-square tests).

These parameters of latent growth curve models can be estimated using many available mixed-effects computer packages (e.g., SAS MIXED and NLMIXED). This approach to growth curve analysis offers techniques for dealing with the problem of unbalanced and nonrandomly incomplete data. This approach can

be used when the data are incomplete because of attrition or are not missing at random (MAR). The latter was the case in the present study because of the a priori criteria concerning who should be retested. In most of the models evaluated, we used the MAR assumption to deal with incomplete longitudinal records and tested the assumptions whenever possible (Little, 1995). This approach allowed us to deal with nonrandom missing data by including all the longitudinal and cross-sectional data to provide the best estimate of the model parameters as if everyone had been reevaluated over time (Diggle & Kenward, 1994; Little, 1995; McArdle et al., 1998). This guarded against the possibility that persons who had not yet been retested differed in performance at baseline from those who were reevaluated on follow-up. Additional details regarding the statistical methods may be found in the Appendix.

The strategic use of these MAR assumptions and pattern-mixture considerations allowed us to include more models over the wide age range analyzed here. In this data collection, there was no common starting point of specific interest (e.g., drug intervention; $t = 0$), so it seemed most natural to use a timing of observation based on the observed or chronological age at the occasion of measurement (i.e., t = age). The age variable used in the models was then recentered (age = 70) and rescaled in terms of change over decades (age = 70/10). This approach provides an estimate of the latent variable means and regressions (i.e., *fixed effects*) and the latent variable variances and covariances (i.e., *random effects*).

Results

Structural Factor Analysis Modeling

The neuropsychological data were first analyzed using techniques of structural factor analysis (McDonald, 1985). The factor analysis demonstrated that a four-factor model was the most appropriate representation of the overall test battery (root-mean-square error of approximation $< .05$). The four common factors were GK, episodic memory (EM), spatial skill (SS), and EF.

Factor scores were then generated for each participant at baseline and represented in standard score units ($M = 0$, $SD = 1$). The factor weights from the baseline analysis were then used to generate factor scores for those individuals with repeated neuropsychological testing. In this sense, the interpretation of each factor was considered to be invariant over time, but the mean and variance of the factor scores were allowed to change over time and over subgroups of individuals.

Modeling Factor Scores in Relation to Demographic Variables

In the first phase of the longitudinal analysis, we developed a mixed model of each of the four factor scores in relation to two demographic variables: years of age (centered at 70) and years of education (centered at 12). Table 4 presents a summary of these initial models. Subsequently, we investigated gender effects and interactions among the variables and considered separate models for each individual factor score. Most of these more complex models did not add anything significant to these results; only those that did are described here.

The first column of Table 4 gives labels for the parameters fitted and for the goodness-of-fit indices. The next eight columns are a listing of parameters for two models for each of the four factors. The first model, labeled "all," indicates the results when all the data from any individual are used (as described above). The second model, labeled "follow," indicates the results when only the data from individuals measured at least twice are used. In general, there are very few differences between the two models for each variable, so we focused on the first and more powerful model (i.e., the one that used all the data).

For example, the model listed under column 1 gives the mixed-model results for the GK factor score in standard score units. The four fixed parameters are listed as follows: the mean score at age 70 for 12 years of education ($-.62$, $p < .001$), the mean change for 1 decade of age ($-.15$, *ns*), the effect of 1 year of education on scores at age 70 (.18, $p < .001$), and the effect of 1 year of education on the change in scores over 1 decade (.01, *ns*). These are followed by the following four random parameters: the variance at age 70 for 12 years of education (.49, *ns*), the variance of the change for 1 decade of age (.02, *ns*), the correlation of level and slope scores (.46, *ns*), and the residual error variance (.18, $p < .001$). Taken as a whole, the GK factor exhibits systematic individual differences between persons (0.73) and has positive effects of education, but there is no evidence for systematic linear slopes over age. The same mixed model was fitted to the three Wechsler Adult Intelligence Scale subtest scores that were the best markers of the GK factor—Information, Similarities, and Vocabulary—and all results were essentially the same as found for the GK factor score (data not shown).

The initial model results for the other three factors are listed in the same way in the other columns of Table 4. The EM factor (column 2) exhibits systematic individual differences (0.70), has large systematic declines over age (-1.08, $p < .001$) and variance (.30, $p < .05$), and has positive effects of education (.05, $p < .01$). The SS factor (column 3) exhibits some systematic individual differences ($-.32$, $p < .001$), has systematic slope declines over age ($-.65$, $p < .001$) and variance (.40, $p < .01$), and has positive effects of education (.13, $p < .001$). The EF factor (column 4) exhibits systematic individual differences ($-.28$, $p < .01$), has systematic declines over age ($-.61$, $p < .001$), and has positive effects of education (.12, $p < .001$). In all cases, the three best test markers for each factor score were examined in the same way, and no notable differences were found (data not shown).

Differences in Factors Scores Among Groups

Longitudinal mixed-model analyses were then performed to examine factor score differences among the four follow-up groups mentioned above. In these analyses, each of the groups was compared with the converters group. This resulted in three contrasts: converters versus controls, converters versus stables, and converters versus decliners. Table 5 presents a summary of these group comparisons as they impact the levels and slopes of the initial mixed model. The maximum likelihood estimates of the mean and variance of each of the four factors at age 70 and at 12 years of education and the mean and variance of the change in each of the four factors projected over a decade interval are all included in the table. The mixed-model results for these analyses are summarized below and are shown graphically in Figures 1–4.

1. GK (see Table 5, column 1 and Figure 1): The GK factor was lower at the projected intercept (age 70 and 12 years

Table 4
Age-Based Linear Mixed-Effect/Latent Growth Models for the Longitudinal Factor Scores With Education Effects

Parameters and fits	General knowledge		Episodic memory		Spatial skills		Executive functioning	
	All	Follow	All	Follow	All	Follow	All	Follow
Fixed (overall) parameters								
Age 70 intercept	$-.62^{***}$	$-.66^{***}$	$-.00$	.12	$-.32^{***}$	$-.26^{**}$	$-.28^{**}$	$-.24^{*}$
10-year slope	−0.15	−0.20	-1.08^{***}	-1.29^{***}	-0.65^{***}	-0.79^{***}	-0.61^{***}	-0.79^{***}
Education on intercept	$.18^{***}$	$.19^{***}$	$.05^{**}$	.05	$.13^{***}$	$.11^{***}$	$.12^{***}$	$.12^{***}$
Education on slope	.01	.03	.03	.04	.04	$.08^{*}$	$-.00$	.03
Random (variance) parameters								
Age 70 variance	$.49^{***}$	$.44^{***}$	$.61^{***}$	$.75^{***}$	$.30^{***}$	$.23^{***}$	$.62^{***}$	$.64^{***}$
10-year slope variance	.02	.01	$.30^{*}$	$.46^{*}$	$.40^{**}$	$.33^{**}$	.10	.08
Level-slope correlation	.46	1.0?	.19	.14	.24	$.58^{**}$	$-.13$	$-.11$
Residual error variance σ_e^2	$.18^{***}$	$.18^{***}$	$.26^{***}$	$.25^{***}$	$.32^{***}$	$.31^{***}$	$.25^{***}$	$.25^{***}$
Goodness-of-fit								
Overall likelihood / parameters = 8	1,184	840	1,403	1,060	1,343	962	1,336	980
No random covariance as baseline $\chi^2(3)$	164	166	185	191	111	107	164	164

Note. "All" indicates the overall number of participants ($N = 327$) and the number of data points ($N = 571$), whereas "Follow" indicates only the followed-up participants ($N = 197$) and number of data points ($N = 415$). The parameters for age are centered at Age[t] = 70 and scaled by $\Delta t = 10$ years (i.e., per decade change), and the parameters for education are centered at 12 years. Parameters are maximum likelihood estimates from minimizing the likelihood function $f = -211$ of the raw data using the SAS PROC MIXED program. The question mark under General Knowledge, Follow, in the Level-slope correlation row indicates that a parameter could not be estimated.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

of education) in participants who subsequently received a diagnosis of AD (i.e., converters) in comparison with either controls ($p < .002$) or decliners ($p < .002$). The rate of decline over time on GK was not significantly greater among converters than in the other groups.

2. EM (see Table 5, column 2 and Figure 2): The EM factor was lower at the projected intercept (age 70 and 12 years of education) in converters in comparison with controls ($p < .0001$), stables ($p < .01$), and decliners ($p < .0001$). In addition, the rate of decline over time in EM was greater in converters in comparison with controls ($p < .006$) and stables ($p < .009$).
3. SS (see Table 5, column 3 and Figure 3): The SS factor was lower at the projected intercept (age 70 and 12 years of education) among converters in comparison with controls ($p < .01$). The rate of decline over time on SS was not significantly greater among converters than in the other groups.
4. EF (see Table 5, column 4 and Figure 4): The EF factor was lower at the projected intercept (age 70 and 12 years of education) among converters in comparison with controls ($p < .008$) and decliners ($p < .0001$). The rate of decline over time on EF approached significance for the comparison of converters versus the controls and stables ($p = .08$).

As noted above, when the three best test markers for each of the factor scores were examined, the group differences were similar, but there was more precision (i.e., accuracy and smaller standard errors) in the factor score mixed-model estimates.

As it has been widely demonstrated that educational level is related to cognitive test scores (e.g., correlation of years of education with GK factor = .54, $p < .01$), we sought to determine whether educational level was operating in different ways for any of the groups by examining the education by group interaction in the models described above. These coefficients were not significant in any of the models.

Relationship of Factor Scores to Genetic Status

Mixed-effects model analyses were then performed to examine the relationship of *APOE* ε4 status to the factor scores. In the initial analyses, *APOE* was coded as a contrast of the number of ε4 alleles (i.e., 0, 1, or 2; see Table 6). In subsequent analyses, we compared sets of genotypes by using contrasts that reflected questions about group comparisons. These latter contrasts were not dichotomous and allowed us to use all available information, thus maximizing statistical power (Cohen et al., 2003). The contrasts were (a) [33 > (24 + 34 + 44)}, (b) [(24 + 34) > 44], and (c) [33 > (22 + 23)]. The additional results of the mixed models with these three effect codes are also presented in Table 6 and can be summarized as follows:

1. GK (Table 6, column 1): The levels and slopes of the GK factor had no significant associations with the three *APOE* effect-coded variables.
2. EM (Table 6, column 2): The EM factor was lower at the projected intercept (age 70 and 12 years of education) in participants who were *APOE* ε4 positive (0.89, $p = .001$) and who were ε4 homozygotes (0.64, $p < .01$). In addition, participants with ε2 showed a higher score than participants with the ε3 homozygotes (−0.46, $p < .01$). No slope effects were found.
3. SS (Table 6, column 3): The SS factor showed no strong

Table 5
Age-Based Linear Mixed-Effect/Latent Growth Models for the Longitudinal Factor Scores With Group Contrasts

Parameters and fits	General knowledge	Episodic memory	Spatial skills	Executive functioning
Fixed (overall) parameters				
Age 70 intercept	−.94***	−.57***	−.53***	−.72***
10-year slope	−0.23	−1.33***	−0.71***	−0.73***
Education on intercept	.18***	.05*	.13***	.11***
Education on slope	.00	.02	.04	−.01
Fixed (group) parameters				
Control vs. convert @70	.53**	.95***	.41*	.65***
Stable vs. convert @70	.21	.44*	.07	.29
Decline vs. convert @70	.46**	.68***	.28	.68***
Control vs. convert on slope	.06	.80**	.13	.45
Stable vs. convert on slope	.20	.75**	.45	.47
Decline vs. convert on slope	.15	.25	.05	.09
Random parameters				
Age 70 variance	.46***	.55***	.28***	.55***
10-year slope variance	.00	.12	.27*	.08
Level-slope correlation	>.99	.15	.41	−.26
Residual error variance	.18***	.27***	.32***	.25***
Goodness-of-fit				
Overall likelihood / parameters = 14	1,139	1,327	1,296	1,274
No random variance as baseline $\chi^2(3)$	148	142	99	144

Note. The parameters for age are centered at Age[t] = 70 and scaled by $\Delta t = 10$ years (i.e., per decade change), and the parameters for education are centered at 12 years. Parameters are maximum likelihood estimates from minimizing the likelihood function $f = -211$ of the raw data using the SAS PROC MIXED program. Dummy codes for the four groups formed into Matrix C = [1, 0, 0, 0; 0, 1, 0, 0; 0, 0, 1, 0, 0].
* $p < .05$. ** $p < .01$. *** $p < .001$.

relationships to *APOE* ε4 status. There was a small but significant difference in decline for ε34 versus ε44 (−.65, $p < .05$), but no additional slope effects were found.

4. EF (Table 6, column 4): The EF factor was lower at the projected intercept (age 70 and 12 years of education) in participants who were *APOE* ε4 positive (.46, $p < .05$). No slope effects were found.

These trajectories for the different *APOE* allele groupings show clear differences in overall levels at any age but show no additional group differences in slopes. The results for the EM factor showed the largest differences based on *APOE* ε4 status, so the numerical estimates for those groupings are presented in Figure 5. The differences in level suggest that members of the ε4 positive groups (i.e., both ε44 and ε34) will reach lower score levels at earlier ages.

As one previous report found a more rapid decline (i.e., greater slope) in memory among MCI cases who were *APOE* ε4 carriers (Wilson, Schneider, et al., 2002) and we did not, we sought to examine our data with an alternate modeling strategy to directly compare our data with this previous report. Rather than the age-

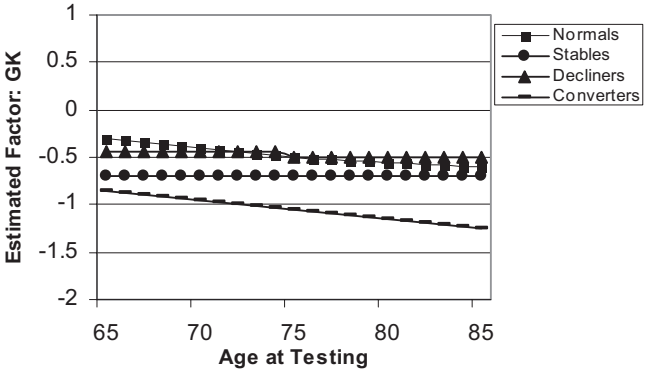

Figure 1. Mixed-model estimates for general knowledge (GK) factor scores in each group.

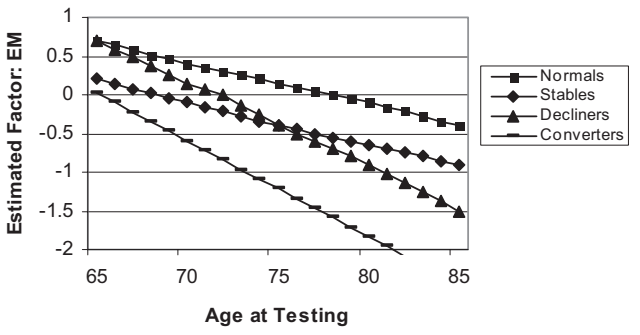

Figure 2. Mixed-model estimates for episodic memory (EM) factor scores in each group.

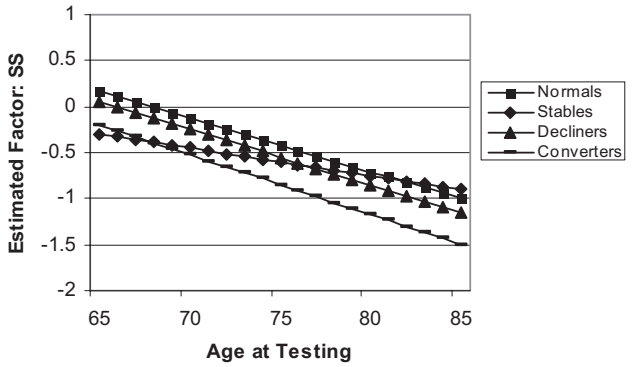

Figure 3. Mixed-model estimates for spatial skill (SS) factor scores in each group.

basis model used above, we substituted the time-basis model used by Wilson and colleagues and replicated their specific *APOE* contrasts (i.e., 33 vs. 34 + 44 and 33 vs. 22 + 23; Wilson, Schneider, et al., 2002). These analyses continued to demonstrate the absence of an *APOE* effect on time slopes (data not shown). As the *APOE* effect sizes of Wilson, Schneider, et al. (2002) were relatively small, it is possible that differences in the overall size and composition of the two samples explain the discrepancy between the two results.

Discussion

These findings indicate that factor scores related to EM show the most striking deficits both at baseline and at follow-up among participants with MCI who are subsequently diagnosed with AD (i.e., converters) compared with all other groups. The importance of changes in memory during prodromal AD is demonstrated by the fact that the EM factor was significantly different among converters in comparison with all other groups at baseline (i.e., controls, stables, decliners) and showed a greater rate of decline on follow-up than in both the control and stable groups. This is consistent with the hypothesis that prodromal AD is characterized by a progressive and profound impairment in EM.

These findings also suggest that deficits in EF are a prominent feature of prodromal AD. The MCI participants who were subsequently diagnosed with AD were significantly impaired at baseline on the EF factor and on the GK factor in comparison with both controls and decliners. These differences were largest in the EF factor between the controls and converters ($p < .001$; see Table 5, column 4). The rate of decline over time on the EF factor did not differ significantly among the groups (i.e., between converters and controls, $p = .08$). No previous study, to our knowledge, has examined rate of decline in EF in an MCI population. A number of studies have, however, reported comparable cross-sectional data using individual test scores (Albert et al., 2001; Chen et al., 2000; Tierney et al., 1996), and their findings are consistent with the present report. These findings are also consistent with the concept that individuals with amnestic MCI have a predominant memory deficit but may also have impairments in other domains (i.e., amnestic MCI, multiple domains impaired; Petersen, 2004).

In addition, both the EM and EF factors showed a significant association with *APOE* ε4 status at baseline. Both factors were significantly lower at baseline among those who were *APOE* ε4 carriers. No association was, however, seen between *APOE* ε4 status and rate of decline over time. This finding with respect to the memory factor is consistent with a previous report in MCI cases (Lange et al., 2002) and contrasts with one that showed a relationship with rate of decline, as noted above (Wilson, Schneider, et al., 2002). This is the first report, to our knowledge, of the impact of *APOE* ε4 carrier status on level of EF performance during prodromal AD, and it is in agreement with a recent meta-analysis incorporating a wide range of populations (B. Small, Rosnik, Fratiglioni, & Backman, 2004).

The association between *APOE* ε4 status and cognitive function at baseline (but not at follow-up) is consistent with the hypothesis that the ε4 genotype affects risk for AD primarily by accelerating the initial stages of the disease process and thereby lowering the age of onset. The differences in level of both memory and EF among the ε4 positive individuals indicate that, without any acceleration in slope, those who are ε4 positive will reach lower levels of performance at an earlier age (see Figure 5) and are therefore more likely to be diagnosed with AD at younger ages. The finding that *APOE* ε4 status is not associated with a more rapid rate of progression among those with established AD (Growdon, Locascio, Corkin, Gomez-Isla, & Hyman, 1996) suggests that this phenomenon is maintained throughout the course of disease. Taken in combination, these results suggest that once the pathophysiological abnormalities of AD are well established, *APOE* ε4 status is unrelated to their evolution. There is increasing evidence to suggest that *APOE* ε4 lowers age of onset because it reduces the clearance of the Aβ peptide from brain, leading to a more rapid accumulation of neuritic plaques (Holtzman et al., 2000). The lack of an effect of *APOE* ε4 status on rate of progression of established AD suggests that once this pathological process is full blown, other factors, above and beyond the rate of Aβ clearance, assume a major role in the evolution of pathology.

Although we did not apply statistics that could evaluate the sequence by which cognitive domains change during the prodromal phase of AD, our findings suggest the sequence that may occur (EM problems occur first, followed by problems in another major cognitive domain). The findings here suggest that this second domain pertains to EF.

Moreover, the declines in cognition that occur during the prodromal phase of AD appear to be fairly steady. There does not

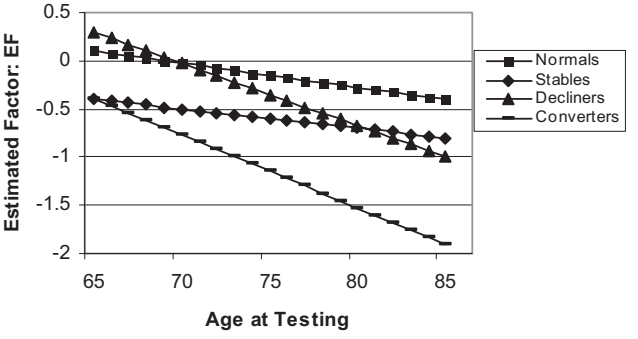

Figure 4. Mixed-model estimates for executive function (EF) factor scores in each group.

Table 6
Age-Based Linear Mixed Effect/Latent Growth Models for the Longitudinal Factor Scores With APOE Contrasts

Parameters and fits	General Knowledge	Episodic Memory	Spatial Skills	Executive Functioning
		Fixed parameters (0, 1, 2)		
Age 70 intercept	−.78***	−.34*	−.40***	−.42**
10-year slope	−0.04	−1.04***	−0.40	−0.78***
Education on intercept	.21***	.06**	.14***	.12***
Education on slope	−.01	.01	.02	−.01
		Fixed *APOE* contrasts		
APOE 33 vs. 34 + 44 @70	.22	.89***	.15	.46*
APOE 34 vs. 44 @70	.18	.64**	.21	.24
APOE 33 vs. 22 + 23 @70	−.07	−.46**	−.05	−.22
APOE 33 vs. 34 + 44 on slope/decade	−.38	.12	−.31	.29
APOE 34 vs. 44 on slope/decade	−.10	−.05	−.65*	.36
APOE 33 vs. 22 + 23 on slope/decade	.26	−.13	.20	−.14
		Random parameters		
Age 70 variance	.45***	.59***	.24***	.59***
10-year slope variance	.01	.25	.42**	.13
Level-slope correlation	>.99	.29	.33	−.12
Residual error variance	.18***	.27***	.31***	.24***
		Goodness-of-fit		
Overall likelihood / parameters = 15	1,082	1,310	1,229	1,229
Random as baseline $\chi^2(3)$	156	170	113	161

Note. Due to missing data on *Apolipoprotein* (*APOE*), participants $N = 308$ and data points $N = 513$. The parameters for age are centered at Age[t] = 70 and scaled by $\Delta t = 10$ years (i.e., per decade change), and the parameters for education are centered at 12 years. Parameters are maximum likelihood estimates from minimizing the likelihood function $f = -211$ of the raw data using the SAS PROC MIXED program. Effect codes for the four groups formed into Matrix E = [0, 1, −.5, −.5; 0, 0, 1, −1; 0, −1, 1, 0, 0]. See Figure 5 for plot of episodic memory expectations.
* $p < .05$. ** $p < .01$. *** $p < .001$.

appear to be a time when performance is stable just prior to a clinical diagnosis of AD, as suggested by some prior studies (Backman et al., 2001). Thus, if one were examining a medication to determine whether or not it altered disease progression among participants in the prodromal phase of AD, one should consider monitoring changes in EF (in addition to memory), as deficits in this cognitive domain appear to be a harbinger of AD.

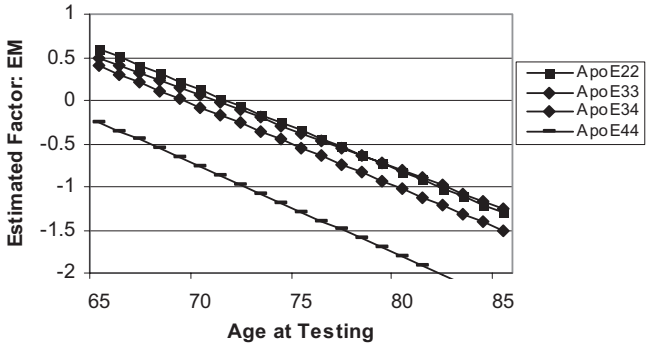

Figure 5. Mixed-model estimates for episodic memory (EM) factor scores for participants grouped by *APOE* status. The differences in the level of the episodic memory factor suggest that members of the ε4 positive groups (i.e., both ε44 and ε34) will reach lower score levels of memory performance at earlier ages than those who are ε4 negative.

The underlying cause of the cognitive declines observed among individuals in the prodromal phase of AD is worthy of note. There is substantial evidence that AD pathology develops first in the medial temporal lobe, leading to selective but profound neuronal loss. This is particularly evident in the entorhinal cortex and CA1 region of the hippocampus (Gomez-Isla et al., 1996), which are brain regions critical for normal memory (Squire & Zola, 1996). This conclusion is supported by in vivo neuroimaging studies showing significant atrophy of these brain regions in MCI cases (for reviews, see Atiya, Hyman, Albert, & Killiany, 2003; Kantarci & Jack, 2003). The pathological changes that are responsible for the additional cognitive deficits that occur during prodromal AD are less well understood. We hypothesize that neuronal loss in the cingulate gyrus (particularly the rostral portion of the anterior cingulate) underlies the EF deficits that develop during this phase of disease, but additional work is needed to clarify this issue.

The declines in EM over time, even among controls, are also of considerable interest. They are consistent with multiple reports that healthy elderly individuals (as well as nonhuman primates and rodents) show mild declines in cognitive performance over time (Moss, Rosene, & Peters, 1988; Wilson, Beckett, et al., 2002). Several factors may explain the difference between the present

findings and those of studies showing no significant cognitive declines in normal aging, such as the nature of the tests that were administered, the fact that individual cognitive domains (such as EM and EF) were not examined separately, or the fact that tests in the present study were not administered annually, thus minimizing practice effects.

The findings among the stable participants are also noteworthy. These participants were identified as stable because they did not show evidence of progressive difficulties in daily life on follow-up. Their cognitive testing, likewise, showed much greater stability than that of the other groups. It is potentially possible that these participants represent normal individuals who experience subtle declines in daily function but are not destined to develop AD over time. An alternative explanation is that these stable individuals have protective factors (e.g., genetic characteristics) that alter their rate of decline but that do not ultimately protect them from progression. A third possibility is that the stable participants represent a group of individuals who have always tended to function lower than the norm. We believe that this latter possibility is the least likely of the three because the educational attainment of the stable participants does not differ significantly from that of the other groups. Also, measures of GK do not suggest that this group is selectively impaired in comparison with the other groups, and the participants were identified on the basis of reports of progressive difficulty in daily life not seen among the controls. A further examination of these potential explanations, in combination with continued follow-up of the cohort, should help determine which of the alternatives is correct.

The longitudinal data studied here have features common to many data sets on aging, and the statistical methods we used may be useful in further research on the evolution of AD. For example, these data were not collected in a randomized controlled pretest–posttest design. As a result, independent causal impacts among variables are confounded and difficult to isolate. Also, there are large differences in initial cross-sectional ages and smaller differences in longitudinal time lags. Given these kinds of limits, these analyses were designed to isolate the temporal components of the changes within a person and a group (i.e., the leading indicators in time), and this is a critical first step in understanding change processes related to AD and representing possible ways to adjust for these processes.

In addition, we attempted to account for the nonrandom criteria by which participants were selected for retesting by including all longitudinal and cross-sectional data in the models. It must, however, still be acknowledged that potential confounds could still play a role, especially those related to key variables, such as the age selection criteria of the cohort (McArdle & Anderson, 1990; Miyazaki & Raudenbush, 2000).

The contemporary statistical modeling techniques used here represent only one example of the feasibility of age-based change models in the analysis of unbalanced longitudinal data. These models can be applied to other cognitive variables as well, and extensions using latent dynamic variables would be appropriate. Group differences based on demographics, prior experiences, or current illness can be studied using any of the same techniques. The dynamic results presented here may differ over these kinds of groupings of persons and also with regard to lead-lag dynamic models (McArdle et al., 2004).

References

Albert, M., Moss, M., Tanzi, R., & Jones, K. (2001). Preclinical prediction of AD using neuropsychological tests. *Journal of the International Neuropsychological Society, 7,* 631–639.

Atiya, M., Hyman, B. T., Albert, M., & Killiany, R. (2003). Structural magnetic resonance imaging in established and prodromal Alzheimer's disease: A review. *Alzheimer Disease and Associated Disorders, 17,* 177–195.

Backman, L., Small, B., & Fratiglioni, L. (2001). Stability of the preclinical phase episodic memory deficit in Alzheimer disease. *Brain, 124,* 96–102.

Baker, E. L., Letz, R., & Fidler, A. T. (1985). A computer administered neurobehavioral evaluation system for occupational and environmental epidemiology. *Journal of Occupational Medicine, 27,* 206–212.

Bennett, D., Wilson, R., Schneider, J., Evans, D. A., Beckett, L. A., Aggarwal, N. T., et al. (2002). Natural history of mild cognitive impairment in older persons. *Neurology, 59,* 198–205.

Benton, A. L., & Hamsher, K. (1976). *Multilingual Aphasia Examination.* Iowa City: University of Iowa Press.

Blacker, D., Haines, J. L., Rodes, L., Terwedow, H., Go, R. C., Harrell, L.E., et al. (1997). ApoE-4 and age at onset of Alzheimer's disease: The NIMH Genetics Initiative. *Neurology, 48,* 139–147.

Bondi, M., Monsch, A., Galasko, D., Butters, N., Salmon, D., & Delis, D. (1994). Preclinical cognitive markers of dementia of the Alzheimer type. *Neuropsychology, 8,* 374–384.

Chen, P., Ratcliff, G., Belle, S., Cauley, J., DeKosky, S., & Ganguli, M. (2000). Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. *Neurology, 55,* 1847–1853.

Chen, P., Ratcliff, G., Belle, S., Cauley, J., DeKosky, S., & Ganguli, M. (2001). Patterns of cognitive decline in presymptomatic Alzheimer disease: A prospective community study. *Archives of General Psychiatry, 58,* 853–858.

Cnaan, A., Laird, N. M., & Slasor, P. (1997). Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Statistics in Medicine, 16,* 2349–2380.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied regression/correlational analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities* (pp. 409–422). Amsterdam: Elsevier Press.

Daly, E., Zaitchik, D., Copeland, M., Schmahmann, J., Gunther, J., & Albert, M. (2000). Predicting "conversion" to AD using standardized clinical information. *Archives of Neurology, 57,* 675–680.

Davis, H., & Rockwood, K. (2004). Conceptualization of mild cognitive impairment: A review. *International Journal of Geriatric Psychiatry, 19,* 313–319.

Delis, D., Kramer, J., Kaplan, E., & Ober, B. (1987). *The California Verbal Learning Test.* New York: Psychological Corp.

Diggle, P., & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics, 43,* 49–93.

Folstein, M., Folstein, S., & McHugh, P. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198.

Gibbons, D., & Hedeker, D. (1997). Random effects probit and logistic regression models for three-level data. *Biometrics, 53,* 1527–1537.

Gomez-Isla, T., Price, J. L., McKeel, D. W., Morris, J. C., Growdon, J. H., & Hyman, B. T. (1996). Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. *Journal of Neuroscience, 16,* 4491–4500.

Grober, E., & Buschke, H. (1987). Genuine memory deficits in dementia. *Developmental Neuropsychology, 3,* 13–36.

Growdon, J., Locascio, J., Corkin, S., Gomez-Isla, T., & Hyman, B. (1996). Apolipoprotein E genotype does not influence rates of cognitive decline in Alzheimer's disease. *Neurology, 47,* 444–448.

Hixson, J., & Vernier, D. (1990). Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with Hha I. *Journal of Lipid Research, 31,* 545–548.

Holtzman, D., Fagan, A., Mackey, B., Tenkova, T., Sartorius, L., Paul, S. M., et al. (2000). Apolipoprotein E facilitates neuritic and cerebrovascular plaque formation in an Alzheimer disease model. *Annals of Neurology, 47,* 739–747.

Howieson, D., Camicioli, R., Quinn, J., Silbert, L., Care, B., Moore, M., et al. (2003). Natural history of cognitive decline in the oldest old. *Neurology, 60,* 1489–1494.

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry, 140,* 566–572.

Jacobs, D., Sano, M., Dooneief, G., Marder, K., Bell, K., & Stern, Y. (1995). Neuropsychological detection and characterization of preclinical Alzheimer's disease. *Neurology, 45,* 957–962.

Kantarci, K., & Jack, C. R. (2003). Neuroimaging in Alzheimer disease: An evidence-based review. *Neuroimaging Clinics of North America, 13,* 197–209.

Kaplan, E., Goodglass, H., & Weintraub, S. (1982). *Boston Naming Test.* Philadelphia: Lea & Febiger.

Kluger, A., Ferris, S., Golomb, J., Mittelman, M., & Reisberg, B. (1999). Neuropsychological performance of decline in dementia in non-demented elderly. *Journal of Geriatric Psychiatry and Neurology, 12,* 168–179.

Knopman, D. S., & Ryberg, S. (1989). A verbal memory test with high predictive accuracy for dementia of the Alzheimer type. *Archives of Neurology, 46,* 141–145.

Lange, K., Bondi, M., Salmon, D., Galasko, D., Delis, D., Thomas, R., & Thal, L. (2002). Decline in verbal memory during preclinical Alzheimer's disease: Examination of the effect of APOE genotype. *Journal of the International Neuropsychology Society, 8,* 943–955.

Little, R. (1995). Modeling the dropout mechanism in repeated-measures studies. *Journal of the American Statistical Association, 90,* 1112–1121.

McArdle, J. J, & Anderson, E. (1990). Latent variable growth models for research on aging. In J. Birren & K. W. Schaie (Eds.), *The handbook of the psychology of aging* (pp. 21–43). New York: Plenum Press.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal multilevel structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology, 38,* 115–142.

McArdle, J. J., Hamagami, F., Jones, K., Jolesz, F., Kikinis, R., Spiro, A., & Albert, M. (2004). Structural modeling of dynamic changes in memory and brain structure using longitudinal data from the normative aging study. *Journal of Gerontology, 59,* P292–P304.

McArdle, J. J., & Nesselroade, J. (2003). Growth curve analyses in contemporary psychological research. In J. Schinka & W. Velicer (Eds.), *Comprehensive handbook of psychology: Research methods in psychology* (Vol. 2, pp. 447–480). New York: Pergamon Press.

McArdle, J. J., Prescott, C., Hamagami, F., & Horn, J. (1998). A contemporary method for developmental-genetic analyses of age changes in intellectual abilities. *Developmental Neuropsychology, 14,* 69–114.

McDonald, R. P. (1985). *Factor analysis and related methods.* Mahwah, NJ: Erlbaum.

McKhann, G., Drachman, D., Folstein, M. F., Katzman, R., Price, D., & Stadlan, E. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimers Disease. *Neurology, 34,* 939–944.

Miyazaki, Y., & Raudenbush, S. W. (2000). Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological Methods, 5,* 24–63.

Morris, J. C., Storandt, M., Miller, J., McKeel, D., Price, J., Rubin, E., & Berg, L. (2001). Mild cognitive impairment represents early stage Alzheimer's disease. *Archives of Neurology, 58,* 1705–1706.

Moss, M., Rosene, D., & Peters, A. (1988). Effects of aging on visual recognition memory in the rhesus monkey. *Neurobiology of Aging, 9,* 495–502.

Newmann, S., Warrington, E., Kennedy, A., & Rossor, M. (1994). The earliest cognitive change in a person with familial Alzheimer's disease: Presymptomatic neuropsychological features in a pedigree with familial Alzheimer's disease confirmed at necropsy. *Journal of Neurology, Neurosurgery, and Psychiatry, 57,* 967–972.

National Institute on Aging and Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer's Disease (1997). Consensus recommendations for the postmortem diagnosis of Alzheimer's disease. *Neurobiology of Aging, 18,* S1–S27.

Petersen, R. (2004). Mild cognitive impairment. *Journal of Internal Medicine, 256,* 183–194.

Petersen, R., Smith, G., Ivnik, R., Kokmen, E., & Tangalos, E. (1994). Memory function in very early Alzheimer's disease. *Neurology, 44,* 867–872.

Petersen, R., Smith, G., Waring, S., Ivnik, R., Tangalos, E., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology, 56,* 303–308.

Petersen, R., Thomas, R., Grundman, M., Bennett, D., Doody, R., Ferris, S., et al. (2005). Vitamin E and donepezil for the treatment of mild cognitive impairment. *New England Journal of Medicine, 352,* 2379–2388.

Petrides, M., & Milner, B. (1982). Deficits in subject-ordered tasks after frontal and temporal lobe lesions in man. *Neuropsychologia, 20,* 249–262.

Porteus, S. (1959). *The Maze Test and clinical psychology.* Palo Alto, CA: Pacific Books.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8,* 271–276.

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique [Psychological examination of cases of traumatic encephalopathy]. *Archives de Psychologie, 28,* 286–340.

Rubin, E., Storandt, M., Miller, J. P., Kinscherf, A., Grant, E., Morris, J., & Berg, L. (1998). A prospective study of cognitive function and onset of dementia in cognitively healthy elders. *Archives of Neurology, 55,* 395–401.

Saunders, A. M., Strittmatter, W. J., Schmechel, D., George-Hyslop, P. H., Pericak-Vance, M. A., Joo, S. H., et al. (1993). Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology, 43,* 1467–1472.

Small, B., Rosnik, C., Fratiglioni, L., & Backman, L. (2004). Apolipoprotein E and cognitive performance: A meta-analysis. *Psychology and Aging, 19,* 592–600.

Small, G., LaRue, A., Komo, S., Kaplan, A., & Mandelkern, M. (1995). Predictors of cognitive change in middle-aged and older adults with memory loss. *American Journal of Psychiatry, 152,* 1757–1764.

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences, USA, 93,* 13515–13522.

Storandt, M., Grant, E., Miller, J., & Morris, J. (2002). Rates of progression in mild cognitive impairment and early Alzheimer's disease. *Neurology, 59,* 1034–1041.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Tierney, M., Szalai, J., Snow, W., Fisher, R., Nores, A., Nadon, G., et al. (1996). Prediction of probable Alzheimer's disease in memory-impaired patients: A prospective longitudinal study. *Neurology, 46,* 661–665.

Tuokko, H., Vernon-Wilkinson, J., Weir, J., & Beattie, W. (1991). Cued

recall and early identification of dementia. *Journal of Clinical and Experimental Neuropsychology, 13,* 871–879.

Verbeke, G., Molenberghs, G., Krickeberg, K., & Fienberg, S. (2000). *Linear mixed models for longitudinal data.* New York: Springer Verlag.

Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19,* 87–95.

Wechsler, D. (1988). *The Wechsler Adult Intelligence Scale—Revised.* New York: Psychological Corporation.

Wilson, R., Beckett, L., Barnes, L., Schneider, J., Bach, J., Evans, D., & Bennett, D. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging, 17,* 179–193.

Wilson, R., Schneider, J., Barnes, L., Beckett, L., Aggarwal, N., Cochran, E., et al. (2002). The apolipoprotein E epsilon 4 allele and decline in different cognitive systems during a 6-year period. *Archives of Neurology, 59,* 1154–1160.

Appendix

Technical Notes

Note 1: First-Order Growth Models

One common form of a growth model is based on a trajectory over time written as:

$$Y[t]_n = y_{0,n} + B[t].y_{1,n} + e[t]_n.$$

In this model y_0 (in lower case) represents unobserved or latent scores representing an individual's initial level; $B[t]$ represents a set of basis coefficients describing the function of the timing of the observations (e.g., $B[t] = 0$ or 1 or $B[t] = Age[t]$); y_1, termed the *latent slopes*, represents unobserved latent scores (in lower case) for the individual change over time; and $e[t]$ represents unobserved but independent errors of measurement at each time. In this latent growth model, the age at any particular time is not considered as a random variable; instead, the changes over age in the scores are considered as the random variables (y_1), and the average group changes are described by the fixed parameters $B[t]$.

Note 2: Second-Order Models

In the typical model included in the article, we also fit a second-order model as a set of regressions:

$$y_{0,n} = \mu_0 + \beta_0 . X_n + e_{0,n} \text{ and } y_{1,n} = \mu_1 + \beta_1 . X_n + e_{1,n},$$

where the latent initial levels (y_0, the expected scores at age 70) and slopes (y_1, the expected changes per decade) are impacted by other variables (X = education, diagnostic group, *APOE* group, etc.) coded in the usual way (see Cohen et al, 2003). The random part of this second-order model also includes residual variables (e_0 e_0), which we assume have variances (σ_o^2 and σ_1^2) and a correlation ($\rho_{01} = .46$). From these parameters, one can calculate other informative statistics about individual differences between persons ($\eta^2 = \sigma_o^2/(\sigma_o^2 + \sigma_e^2)$, as well as goodness of fit (χ^2 tests).

Note 3: Computer Programs

The parameters of latent growth curve models can be estimated using many available mixed-effects computer packages (e.g., SAS MIXED and NLMIXED; Verbeke, Molenberghs, Krickeberg, & Fienberg, 2000) and structural equation programs. This approach to growth curve analysis offers advanced techniques for dealing with the problem of unbalanced and nonrandomly incomplete data. In computational terms, the available information for any participant on any data point (i.e., any variable measured at any occasion) is used to build up maximum likelihood estimates (MLE) using a numerical routine that optimizes the model parameters with respect to any available data. These MLE are based on fitting structural models to the raw score information for each person on each variable at each time (e.g., McArdle et al., 2002). The goodness of fit of each model presented in the article was assessed using classical statistical principles about the model likelihood (f_{MLE}). This approach can be used when the data are incomplete because of attrition and other factors and not missing at random (MAR). In most models presented in the article, we used the MAR assumption to deal with incomplete longitudinal records, but these assumptions were tested whenever possible (e.g., Little, 1995).

Note 4: Dealing With Incomplete Data

The mixed-effect MLE approach allows one to deal with nonrandom attrition by including all the longitudinal and cross-sectional data and introduce tests of these "pattern-mixture" assumptions whenever possible (e.g., age-selectivity; see Gibbons & Hedeker, 1991).

The strategic use of these MAR assumptions and pattern-mixture considerations allows one to include more models over the full age range analyzed in the article. For this age-basis model to be viable, one needs to presume the untestable MAR assumptions apply to this age dimension. This implies that the score measured on a person at each age gives one some indication of his or her likely scores at the ages not measured and that persons measured at specific ages represent the age-based scores for anyone. This MAR assumption has become an incomplete data design problem that includes assumptions about age changes in different cohorts (e.g., McArdle & Anderson, 1990; Miyazaki & Raudenbush, 2000).

Received February 28, 2006
Revision received August 14, 2006
Accepted September 18, 2006 ■