Journal of Geriatric Psychiatry and Neurology

Learning and Generalization Tasks Predict Short-Term Cognitive Outcome in Nondemented Elderly

Catherine E. Myers, Alan Kluger, James Golomb, Mark A. Gluck and Steven Ferris

The online version of this article can be found at:

Published by:
SAGE Publications
http://www.sagepublications.com

Additional services and information for *Journal of Geriatric Psychiatry and Neurology* can be found at:

Email Alerts: http://jgp.sagepub.com/cgi/alerts

Subscriptions: http://jgp.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations (this article cites 24 articles hosted on the
SAGE Journals Online and HighWire Press platforms):
http://jgp.sagepub.com/cgi/content/refs/21/2/93

Learning and Generalization Tasks Predict Short-Term Cognitive Outcome in Nondemented Elderly

Catherine E. Myers, PhD, Alan Kluger, PhD,
James Golomb, MD, Mark A. Gluck, PhD, and Steven Ferris, PhD

This study examines whether behavioral measures obtained in nondemented elderly can predict cognitive status at 2-year follow-up. Prior studies have established that delayed paragraph recall can help predict short-term risk for decline to mild cognitive impairment and Alzheimer disease. It was examined whether prediction accuracy can be improved by adding a discrimination-and-generalization task that has previously been shown to be disrupted in nondemented elderly with hippocampal atrophy, a risk factor for Alzheimer disease. Fifty nondemented, medically healthy elderly patients received baseline clinical diagnosis and cognitive testing; 2 years later, patients received a follow-up clinical diagnosis of normal, mild cognitive impairment, or probable Alzheimer disease. In all, 2 baseline variables, delayed paragraph recall and generalization performance, were predictive of follow-up outcome with sensitivity of 81% and specificity of 91%—better than the classification accuracy based on either of these measures alone. These preliminary results suggest that these behavioral tasks may be useful tools in predicting short-term cognitive outcome in nondemented elderly.

Keywords: aging; cognition; mild cognitive impairment; Alzheimer disease; hippocampus; memory

An important issue in gerontology is the ability to predict short-term risk for cognitive impairment in elderly individuals before the onset of symptoms severe enough to warrant a diagnosis of probable Alzheimer disease (AD) or other forms of dementia. Clinically, this is important because existing AD medications act to slow but not reverse or prevent decline; early diagnosis would allow aggressive intervention to prolong high levels of cognitive function for as long as possible. Pharmacological research

would also benefit from early diagnosis methods, allowing better identification of at-risk individuals who might be available to participate in trials of new treatments to prevent or delay cognitive decline. Accurate diagnosis would also improve the quality of prognostic information given to patients and family.

Previous study has shown that delayed recall of short paragraphs may be a sensitive indicator of short-term risk for cognitive decline.[1] Nondemented elderly who perform poorly on delayed paragraph recall are at heightened risk to decline to AD within the next 3 to 4 years.[1] Delayed paragraph recall is impaired in individuals with mild cognitive impairment (MCI) relative to healthy elderly[2,3] and is also impaired in healthy elderly individuals with hippocampal atrophy,[4] a risk factor for developing dementia over the next several years.[5] Thus, delayed paragraph recall has promise as a simple, noninvasive tool for prediction of cognitive decline in nondemented elderly individuals.

One possible explanation for the disruption of delayed paragraph recall in individuals at short-term

93

risk of cognitive decline is its dependence on the hippocampus; indeed, delayed paragraph recall is greatly disrupted in amnesic patients with bilateral medial temporal lobe damage.[6] Radiological studies have demonstrated tissue loss in the hippocampal formation in patients with mild AD,[7] and progressive hippocampal dysfunction has been proposed as a possible neuroanatomical basis of AD.[8-10] Hippocampal atrophy can also occur in older adults with MCI, a presumably heterogeneous group that may include individuals with greater-than-average forgetfulness related to aging (age-associated memory impairment) and individuals with subclinical dementia.[11]

In this context, it is possible that other tasks, which are similarly disrupted in individuals with hippocampal dysfunction, might also have utility in predicting short-term cognitive outcome in nondemented elderly. We have considered one such task, a computer-based discrimination-and-generalization task, in which subjects learn a series of visual discriminations and are then challenged to generalize when stimulus information is altered.[12] In the previous study, we have shown that nondemented elderly individuals with mild-to-moderate hippocampal atrophy are selectively impaired on the generalization portion of this task compared with nonatrophied peers[12]; amnesic patients with bilateral medial temporal damage are similarly impaired on generalization.[13] This pattern of impairment in individuals with compromised hippocampal function is similar to that seen on delayed paragraph recall, which led us to ask whether the discrimination-and-generalization task might have similar power to predict incipient cognitive decline in nondemented elderly and whether the 2 tasks together might predict cognitive outcome with greater accuracy than either alone.

As an initial investigation of this question, we administered the paragraph recall and discrimination-and-generalization tasks to a group of 50 nondemented elderly individuals. At baseline, all were diagnosed as cognitively normal or MCI. At a follow-up session, 2 years later, these individuals were reassessed for cognitive status. We examined the degree to which performance on these 2 learning tasks at baseline was predictive of a 2-year cognitive outcome.

Methods

Subjects

A group of 53 nondemented, medically healthy, community-dwelling elderly individuals were initially recruited from among the pool of subjects participating in the ongoing gerontological research at the New York University Alzheimer Disease Center (NYU ADC). Individuals who apply to participate in research at the NYU ADC receive comprehensive medical, physical, neurologic, and psychiatric evaluations. This includes full behavioral assessment and cognitive testing in an outpatient setting; routine laboratory testing that comprised blood chemistry, serum B_{12} and folate levels, thyroid function, and urinalysis; and electrocardiograms and clinical brain scans (MRI or computed tomography).

Exclusions were made if any evaluations suggested the presence of a disease state that could affect brain functioning or cognition. Specifically, participants were excluded if there was clinical or brain scan evidence of infarction, inflammation, infection, or neoplastic disease or if there was any history of psychiatric or neurologic disorder, including affective disorder, Parkinson disease, normal pressure hydrocephalus, significant sensory impairment, peripheral neuropathy, or severe arthropathy. Also excluded were any participants with more than borderline hypertension (>160/90 mm Hg), with a Hamilton Depression Scale[14] score of 16 or greater or with a history of excessive alcohol intake.

All participants were diagnosed in an outpatient setting by NYU ADC clinicians and given a baseline Global Deterioration Scale (GDS) score.[15] This is a 7-point scale with ratings of 1 (no memory deficit), 2 (subjective complaints of memory deficit with no objective evidence), 3 (MCI), 4 (mild dementia), and so on through 7 (severe dementia). For inclusion in this study, participants were required to have a baseline GDS score of 3 or lower. The GDS scores were assigned following a semistructured interview conducted by trained physicians with subjects and with knowledgeable informants such as family members and were based on the subject's overall level of cognitive and functional status in accordance with published procedures.[15]

Participants were also given a follow-up GDS rating 2 years (22-26 months) after baseline assessment. At this point, it was determined that 3 of the participants had developed unrelated medical conditions during the intervening period (mild vascular disease, vitamin B deficiency, or normal pressure hydrocephalus). These participants' data were excluded post hoc, leaving a final sample size of 50. This included 28 women and 22 men, with a mean age of 67.46 years (standard deviation [SD] = 8.5) and mean education level of 15.42 years (SD = 2.25).

Informed consent was obtained from all subjects prior to initiation of baseline testing; subjects were volunteers and received no payment for participation in this study.

Baseline and Follow-Up Diagnosis

For analysis, we divided this sample in 2 ways: based on baseline diagnosis and again based on follow-up diagnosis (independent of baseline status).

At baseline, 39 subjects were classified as cognitively normal (GDS stage 1 or 2) and formed the Normal-b group; the remaining 11 subjects were classified as MCI or MCI (GDS stage 3) and formed the Impaired-b group. The Normal-b group included 24 women and 15 men, with a mean age of 66.6 years (SD = 8.0). The Impaired-b group included 4 women and 7 men, with a mean age of 70.5 years (SD = 10.0). There was no significant difference between Normal-b and Impaired-b groups in baseline age (analysis of variance [ANOVA], $F_{(1,48)}$ = 1.79; P > .05), education level (ANOVA, $F_{(1,48)}$ = 0.30; P > .05), or in proportion of men and women (Yates-corrected χ^2 = 1.30, df = 1; P > .05).

At a 2-year follow-up, of the 39 individuals who had been classified as cognitively normal at baseline, 8 had declined to MCI (GDS stage 3) and one had declined to probable AD (GDS stage 4). Of the 11 individuals who had been classified as MCI at baseline, 6 were again diagnosed as MCI; one had declined to probable AD, and 4 were now diagnosed as cognitively normal (GDS stage 2). The 34 individuals classified as cognitively normal at follow-up (regardless of baseline status) formed the Normal-f group. Due to the relatively low number of probable AD cases at follow-up, the individuals classified as MCI and AD at follow-up were combined into a single cognitively impaired group (Impaired-f) for statistical analysis. The Normal-f group included 22 women and 12 men, with a mean age of 67.2 years (SD = 9.0); the Impaired-f group included 6 women and 8 men, with a mean age of 68.0 years (SD = 7.6). Both the individuals classified as probable AD were men (ages 69 and 83). The Impaired-f and Normal-f groups did not differ in their baseline age (ANOVA, $F_{(1,48)}$ = 0.07; P > .05), education level ($F_{(1,48)}$ = 0.25; P > .05), or in proportion of men and women (Yates-corrected χ^2 = 2.26, df = 1; P > .05).

Procedure

Baseline Neuropsychological Testing

At the time of baseline diagnosis, all subjects received neuropsychological tests, including the Logical Memory (LM) subtest of the Wechsler Memory Scale-Revised (WMS-R)[16] including both immediate (LMI) and delay (LMII) portions, the digit span (DIG) and digit symbol substitution (DSST) subtests of the Wechsler Adult Intelligence Scale-Revised (WAIS-R),[17] and the mini-mental state (MMS) test.[18]

In particular, the LM test involves reading a short paragraph aloud to the subject; the subject is then asked to repeat the paragraph immediately from memory, and receives one point for each story element correctly recalled (maximum 25). Immediate recall summed over 2 paragraphs produces an LMI score. Following a 20-minute filled delay, subjects are then asked to recall each story again from memory. Delayed recall summed over the 2 paragraphs produces an LMII score.

Discrimination-and-Generalization Task

Also at the time of baseline diagnosis, subjects were administered the discrimination-and-generalization task of Myers et al.[12] In brief, this is a short (20 minute) computerized learning task, administered on a Macintosh i-book or comparable computer. Phase 1 is a discrimination learning task. On each trial, the subject sees a pair of objects, one of which has arbitrarily been designated as correct. The subject learns through trial-and-error to pick the correct object and enters a response by pressing one of the 2 keys labeled "LEFT" and "RIGHT" to choose the object that appears on the left or the right of the screen (Figure 1A). The chosen object is then raised and, if the subject's choice is correct, a smiley face is revealed underneath.

In all, 8 such object pairs are trained; within each pair, the objects differ in color or shape but not both. In all, 4 of the pairs contain 2 objects of the same color, such as a green mushroom shape versus a green frame shape (illustrated in black and white in Figure 1A), with the mushroom shape designated as the correct object. In this case, shape can be used to determine the correct answer (left or right); color is redundant and therefore irrelevant with respect to determining the correct answer. The other 4 object pairs involve 2 objects of the same shape, such as a
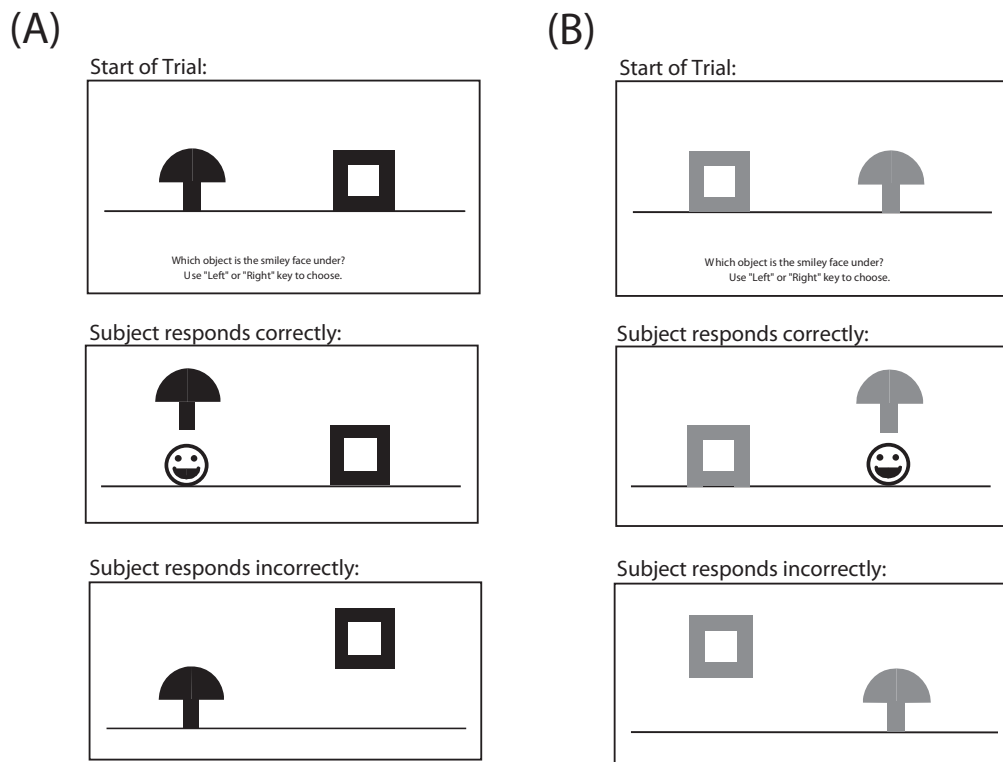
Figure 1. Screen events in the discrimination-and-generalization task. A, During each phase 1 (discrimination learning) trial, a pair of objects appeared that differed in shape or color but not both. The subject would then choose one object by pressing a labeled key. The chosen object was raised and, if the subject's response was correct, a smiley face appeared underneath. B, In phase 2 (generalization testing), object pairs were altered so that the previously irrelevant dimension was novel but the previously relevant dimension remained the same.

red octagon versus a yellow octagon. In these cases, color can be used to determine the correct answer; shape is redundant and therefore irrelevant with respect to determining the correct answer.

The 8 object pairs are trained concurrently, in blocks of 16 trials that contain 2 presentations of each pair—once with the correct object appearing on the left and once on the right. Trial order within a block is randomized. Phase 1 continues until the subject makes 16 consecutive correct responses or to a maximum of 96 trials.

Next, phase 2 begins without warning to the subject. Phase 2 is a generalization test. This phase is identical to phase 1 except that, in each pair, the redundant or irrelevant feature is altered. Thus, as Figure 1B shows, one pair would still consist of a mushroom shape versus a frame shape but the color would be altered (eg, gray). In pairs where color is the predictive feature, color would remain unaltered, but the irrelevant shape feature would change (eg, from red octagon vs yellow octagon to red arrow vs

yellow arrow). Again, phase 2 continues until the subject makes 16 consecutive correct responses or to a maximum of 48 trials. The entire task, including phases 1 and 2, takes approximately 20 minutes to complete.

For each subject, performance is scored as total number of errors during phase 1 discrimination learning (ERR1) and phase 2 generalization (ERR2). Note that low ERR1 and ERR2 scores therefore indicate good performance.

Data Analysis

The primary outcome measure was follow-up diagnosis (Normal-f or Impaired-f). On the basis of this follow-up diagnosis, we conducted a stepwise discriminant function analysis with predictors of sex and baseline age, GDS, LMI, LMII, DSST, DIG, MMS, ERR1, and ERR2 to determine which of these baseline measures could be used to predict the follow-up diagnosis.

Table 1.   Group Means at Baseline for Subjects Given Baseline Diagnosis of
Cognitively Normal (Normal-b) or Cognitively Impaired (Impaired-b)[a]

Diagnosis at Baseline	N	Age	MMS	LMI	LMII	DIG	DSST	ERR1	ERR2
Normal-b	39	66.6 (8.0)	29.0[b] (1.1)	28.0[b] (6.6)	23.6[b] (6.7)	12.5 (2.1)	53.5 (11.2)	17.9[b] (11.8)	9.7 (12.8)
Impaired-b	11	70.5 (10.0)	28.2 (1.3)	22.9 (9.0)	17.4 (11.3)	11.8 (2.2)	45.7 (15.6)	28.7 (16.6)	16.5 (14.9)

NOTE: MMS = mini-mental state test score; LMI = Wechsler Memory Scale–Revised logical memory immediate score; LMII = Wechsler Memory Scale–Revised logical memory delay score; DIG = Wechsler Adult Intelligence Scale–Revised digit span (forward + backward); DSST = Wechsler Adult Intelligence Scale–Revised digit symbol substitution score; ERR1 = total errors on phase 1 of the discrimination-and-generalization task; ERR2 = total errors on phase 2 of the discrimination-and-generalization task.
a. Standard deviations in parentheses.
b. Significant differences between Normal-b and Impaired-b groups ($P < .05$).

In addition, we examined whether performance on the paragraph recall and discrimination-and-generalization tests differed significantly among individuals given baseline diagnoses of cognitively normal (Normal-b) or cognitively impaired (Impaired-b), that is, whether cognitive status at time of testing was predictive of task performance and whether performance on these 2 tasks was correlated within individuals.

Results

Baseline Testing

Table 1 shows group means at baseline for participants given a baseline diagnosis of cognitively normal (Normal-b) or cognitively impaired (Impaired-b). The Normal-b group significantly outperformed the Impaired-b group on MMS (ANOVA, $F_{(1,48)} = 4.54$; $P = .038$), LMI ($F_{(1,48)} = 4.30$; $P = .044$), and LMII ($F_{(1,48)} = 5.33$; $P = .025$) but not on DIG ($F_{(1,48)} = 0.85$; $P > .05$) or DSST ($F_{(1,48)} = 3.45$; $P > .05$). Unsurprisingly, there was a strong within-subjects correlation between LMI and LMII scores (Pearson r = 0.915; $P < .001$). Therefore, we also computed a % recall measure as LMII divided by LMI; the Normal-b group average (75.2%, SD = 18.9) and Impaired-b group average (67.8%, SD = 29.0) were significantly different (ANOVA, $F_{(1,48)} = 5.64$; $P = .022$).

On phase 1 of the discrimination-and-generalization task, the Normal-b group outperformed the Impaired-b group in terms of ERR1 score (ANOVA, $F_{(1,45)} = 4.23$; $P = .045$) with no effect of sex ($F_{(1,45)} = 1.19$; $P > .05$) or age ($F_{(1,45)} = 0.05$; $P > .05$) and with no interactions ($P > .05$). Overall, subjects performed better on phase 2 (generalization) than phase 1, reflected in significantly lower ERR2 than ERR1 scores (paired-samples t test, $t_{49} = 4.40$; $P < .001$). However, performance on phase 2 did not differ between groups (ANOVA, $F_{(1,45)} = 1.53$; $P > .05$) with no effects of age or sex and with no interactions (all $P > .05$). Phase 2 performance depends to some extent on phase 1 performance (Pearson r = 0.425; $P = .002$). Accordingly, we also computed a retention measure, defined as ERR2 divided by ERR1; using this measure, the Normal-b group average (63.4%, SD = 53.2) and the Impaired-b group average (69.6%, SD = 1.17) did not differ significantly (ANOVA, $F_{(1,48)} = 0.03$; $P > .05$).

The 2 measures which had previously been shown to be sensitive indices of hippocampal function, LMII and ERR2, showed a significant negative within-subject correlation (Pearson r = −0.472; $P = .001$); the negative correlation reflects the fact that good performance is reflected in a numerically high LMII score and a numerically low ERR2 score.

Follow-up Diagnosis

Table 2 shows the mean baseline measurements for subjects who received follow-up diagnoses of cognitively normal (Normal-f) or cognitively impaired (Impaired-f). Table 2 also shows scores for individuals within the Impaired-f group who were diagnosed as MCI or probable AD. The Impaired-f and Normal-f groups did not differ in baseline MMS ($F_{(1,48)} = 0.01$; $P > .05$), DIG ($F_{(1,48)} = 1.13$; $P > .05$), or DDST ($F_{(1,48)} = 2.51$; $P > .05$). However, baseline scores on paragraph recall were higher for the Normal-f group than for the Impaired-f group (Figure 2). A 2-factor ANOVA with between-subject variable of group and within-subject variable of immediate versus delay recall (LMI vs LMII) confirmed a significant effect of group ($F_{(1,48)} = 21.48$; $P < .001$) and a within-subject difference on immediate versus delay recall

Table 2.   Mean (and Standard Deviation) of Baseline Neuropsychological and Behavioral Measurements for Subjects Diagnosed as Cognitively Normal (Normal-f) or Cognitively Impaired (Impaired-f) at a 2-Year Follow-up

Diagnosis at Follow-up	N	Age	MMS	LMI	LMII	DIG	DSST	ERR1	ERR2
Normal-f	34	67.2 (9.0)	28.9 (1.1)	29.8[a] (6.4)	25.4[a] (6.1)	12.6 (2.3)	53.7 (12.0)	16.3[a] (11.1)	5.3[a] (7.5)
Impaired-f	16	68.0 (7.6)	28.8 (1.4)	21.3 (6.4)	15.5 (8.1)	11.9 (1.7)	47.8 (13.1)	28.6 (15.0)	23.9 (14.6)
Mild cognitive impairment	14	66.8 (6.9)	29.1 (1.2)	23.0 (4.8)	17.5 (6.5)	12.0 (1.8)	50.3 (9.6)	26.4 (14.6)	23.8 (14.6)
Alzheimer disease	2	76.0 (9.9)	27.0 (1.4)	9.5 (0.7)	1.5 (0.7)	11.0 (1.4)	30.0 (25.5)	44.5 (5.0)	24.5 (20.5)

NOTE: MMS = mini-mental state test score; LMI = Wechsler Memory Scale–Revised logical memory immediate score; LMII = Wechsler Memory Scale–Revised logical memory delay score; DIG = Wechsler Adult Intelligence Scale–Revised digit span (forward + backward); DSST = Wechsler Adult Intelligence Scale–Revised digit symbol substitution score; ERR1 = total errors on phase 1 of the discrimination-and-generalization task; ERR2 = total errors on phase 2 of the discrimination-and-generalization task.
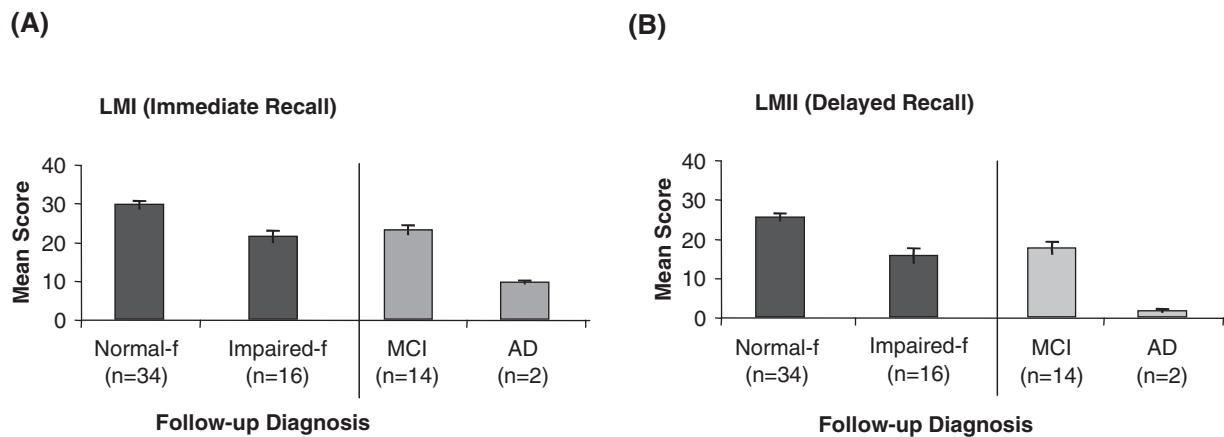a. Significant differences between Normal-f and Impaired-f groups ($P < .05$).

Figure 2.   Performance at baseline on (A) immediate (LMI) and (B) delayed (LMII) paragraph recall, as a function of follow-up diagnosis of cognitively normal (Normal-f) or cognitively impaired (Impaired-f), which can be further subdivided into individuals classified as mild cognitive impairment or probable Alzheimer disease. MCI indicates mild cognitive impairment; AD = Alzheimer disease; LMI = Wechsler Memory Scale–Revised logical memory immediate score; LMII = Wechsler Memory Scale–Revised logical memory delay score.

(LMII < LMI, $F_{(1,48)} = 101.78$; $P < .001$) but no significant interaction ($F_{(1,48)} = 2.97$; $P > .05$).

The Normal-f and Impaired-f groups also differed in performance on the discrimination-and-generalization test (Figure 3). A 2-factor ANOVA with between-subject variable of group and within-subject variable of discrimination versus generalization (ERR1 vs ERR2) confirmed a significant effect of group ($F_{(1,48)} = 33.33$; $P < .001$) and a within-subject effect of discrimination versus generalization (ERR2 < ERR1, $F_{(1,48)} = 13.15$; $P = .001$) but no interaction ($F_{(1,48)} = 2.10$; $P > .05$). Figure 4 shows these data.

Predicting 2-Year Outcome Based on Baseline Measurements

Discriminant function analysis was conducted to determine whether 2-year outcome could be predicted based on measurements taken at baseline. A stepwise analysis was conducted on measures of sex and on baseline age, GDS, MMS, DIG, DSST, LMI, LMII, ERR1, and ERR2, with a factor of 2-year outcome (Normal-f vs Impaired-f) and with α for adding and removing set to .05.

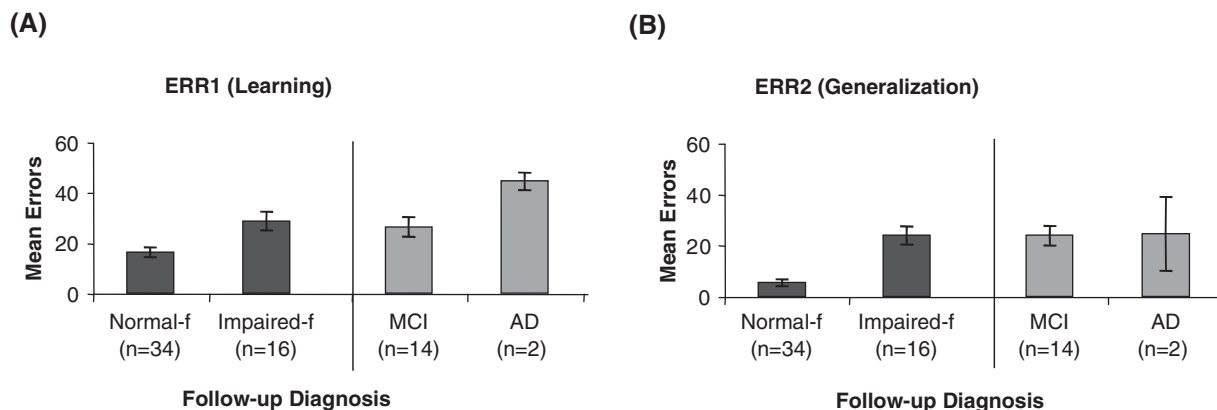The discriminant analysis identified 2 variables as predictive: LMII (F-to-remove 8.68, tolerance

(A)

(B)

Figure 3. Performance at baseline on (A) discrimination (ERR1) and (B) generalization (ERR2) task, as a function of follow-up diagnosis of cognitively normal (Normal-f) or cognitively impaired (Impaired-f), which can be further subdivided into individuals classified as MCI or probable AD. ERR1 indicates total errors on phase 1 of the discrimination-and-generalization task; ERR2 = total errors on phase 2 of the discrimination-and-generalization task; MCI = mild cognitive impairment; AD = Alzheimer disease.
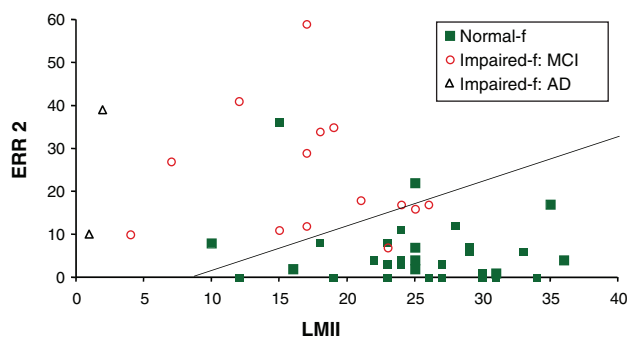

Figure 4. Scatter plot of individual cases, as a function of baseline LMII and ERR2 scores, with cases coded according to follow-up diagnosis of Normal-f or Impaired-f (which includes cases classified as MCI or probable Alzheimer disease). A boundary line is shown that approximates the discriminant function (discussed below) separating Normal-f from Impaired-f cases; the line is drawn so that the 3 Normal-f and 3 Impaired-f cases that are misclassified by the discriminant function fall on the wrong side of this boundary line. LMII indicates Wechsler Memory Scale–Revised logical memory delay score; ERR2 = total errors on phase 2 of the discrimination-and-generalization task; MCI = mild cognitive impairment; AD = Alzheimer disease.

0.979) and ERR2 (F-to-remove 19.14, tolerance 0.979). No other variables added significant predictive accuracy. A group classification function was then calculated (coefficients: LMII = 0.081, ERR2 = −0.74, constant = −0.975). This classification

function correctly predicted class membership for 13 of 16 Impaired-f cases and 31 of 34 Normal-f cases for an overall predictive accuracy of 88% (Table 3).

By contrast, a function based on LMII alone correctly classified 11 of 16 Impaired-f cases and 28 of 34 Normal-f cases for an overall predictive accuracy of 78%. A function based on ERR2 alone correctly classified 11 of 16 Impaired-f cases and 31 of 34 Normal-f cases for an overall predictive accuracy of 84%. Thus, a classification function based on both LMII and ERR2 provides greater predictive accuracy than either measure alone.

The 6 cases that were misdiagnosed by the joint classification function were all individuals who had been diagnosed as cognitively normal at baseline; 3 were still diagnosed as cognitively normal at a 2-year follow-up, and 3 had declined to MCI. Table 4 shows the age and neuropsychological and behavioral measures for these cases compared with the Normal-f and Impaired-f group means with these 6 cases excluded. In particular, at baseline testing, cases 1542, 303, and 100 showed relatively strong LMII performance (within 1 SD of Normal group mean) and poor generalization performance (2 of 3 cases show ERR2 greater than 1 SD from the Normal-f group mean). Conversely, cases 110, 122, 160 showed relatively poor LMII and generalization performance compared with the Normal-f group means.

Table 3. Predictive Variables Identified by Discriminant Function Analysis

	Sensitivity (%)	Specificity (%)	Overall Accuracy (%)
LMII and ERR2			
(combined)	81.3	91.0	88.0
LMII (alone)	69.0	82.0	78.0
ERR2 (alone)	69.0	91.0	84.0

NOTE: LMII = Wechsler Memory Scale–Revised logical memory delay score; ERR2 = total errors on phase 2 of the discrimination-and-generalization task.

Summary and Discussion

The central question being examined here was whether a short battery of neuropsychological tests, administered to nondemented, medically healthy elderly patients, can predict cognitive outcome at a 2-year follow-up. The results suggested that 2 measures in particular, delayed paragraph recall and the generalization component of a discrimination-and-generalization test, could predict with 88% accuracy which individuals would present at follow-up as cognitively normal or with cognitive impairment (MCI or probable AD). This article builds on earlier studies documenting that delayed paragraph recall can predict cognitive decline in nondemented elderly.[1] However, in this study, prediction accuracy was improved when both delayed paragraph recall and generalization were considered, relative to either measure alone.

Characteristics of the Sample

In the current data, we found a 2-year conversion rate of 2.6% from cognitively unimpaired to AD and of 9.1% from MCI to AD. These conversion rates are broadly consistent with other studies finding that individuals with MCI tend to progress to probable AD at a rate of approximately 10% to 15% per year compared with a conversion rate of 1% to 3% per year among the cognitively unimpaired elderly.[3,19-24] We also had a relatively high rate of MCI at baseline (11 of 50 or 22%) and of conversion from normal to MCI over the 2-year study period (8 of 39 or 21%). These rates are somewhat higher than might be expected from medically healthy elderly individuals in the broader population and reflect the characteristics of the volunteers who present themselves at the NYU ADC—many of whom do so in response to subjective concerns about memory decline. Therefore,

the current results should be interpreted with caution, both on account of the small sample size and of the characteristics of this sample (see also Limitations of the Current Study).

Most strikingly, in the current sample, we found that 4 cases diagnosed as MCI at baseline were subsequently reclassified as cognitively normal at follow-up (4 of 11, or 36%). This highlights the difficulty in assessing cognitive status based on a single evaluation as an individual's performance may fluctuate due to factors not directly related to cognitive decline (eg, attention or motivation) and thus highlights the need for studies that repeat testing several times during a longer follow-up interval (see also Limitations of the Current Study).

At the same time, in most cases, our behavioral results are consistent with those of the prior studies. For example, at baseline, our Impaired-b group performed significantly worse than the Normal-b group on several measures that involved learning and memory components, including MMS, LMI, LMII, and ERR1, but not on other tests that do not critically involve recent memory, including DIG and DSST. This difference in LMII between Normal-b and Impaired-b groups is consistent with prior studies showing that delayed paragraph recall is a neuropsychological measure, which is particularly likely to distinguish subjects with MCI from healthy elderly.[2,3] The Normal-b and Impaired-b groups did not differ significantly on ERR2, which is consistent with our prior finding that cognitive status (cognitively normal vs MCI) did not significantly affect generalization performance on this task.[12]

Predicting 2-Year Outcome Based on Baseline Performance

Dividing our sample according to 2-year follow-up, both paragraph recall (LMI and LMII) and behavioral task performance (ERR1, ERR2) were significantly different between the Normal-f and Impaired-f groups; within the Impaired-f group, the AD subgroup performed especially poorly on these measures. However, only 2 measures—LMII and ERR2—contributed significantly to predicting 2-year outcome. When LMII was considered separately, it alone could provide 78% prediction accuracy consistent with prior studies implicating delayed paragraph recall as a simple, noninvasive procedure that can provide a high degree of accuracy in assessing risk for cognitive decline.[1] However,

Table 4. Age and Neuropsychological and Behavioral Measurements at Baseline for the 6 Subjects Whose Follow-up Outcome Differed From That Predicted by the Discriminant Function[a]

Case	Baseline Age	MMS	LMI	LMII	DIG	DSST	ERR1	ERR2	Actual Outcomes	Predicted Outcome
1542	64	30	30	25	10	60	33	16	Impaired-f	Normal
303	66	28	26	23	13	57	23	7	Impaired-f	Normal
95-100	71	30	25	26	15	58	13	17	Impaired-f	Normal
97-119	74	30	14	10	13	53	15	8	Normal-f	Impaired
97-122	75	27	18	15	12	39	44	36	Normal-f	Impaired
98-160	63	30	31	25	9	52	20	22	Normal-f	Impaired
Normal-f group mean	66.9 (9.2)	28.8 (1.1)	30.3 (5.6)	26.2 (5.4)	12.7 (2.3)	54.2 (12.3)	15.4 (10.4)	3.6 (4.2)		
Impaired-f group mean	68.22 (8.3)	28.7 (1.4)	20.0 (6.4)	13.0 (7.5)	11.7 (1.5)	45.0 (13.4)	29.9 (15.9)	26.3 (15.0)		

NOTE: MMS = mini-mental state test score; LMI = Wechsler Memory Scale–Revised logical memory immediate score; LMII = Wechsler Memory Scale–Revised logical memory delay score; DIG = Wechsler Adult Intelligence Scale–Revised digit span (forward + backward); DSST = Wechsler Adult Intelligence Scale–Revised digit symbol substitution score; ERR1 = total errors on phase 1 of the discrimination-and-generalization task; ERR2 = total errors on phase 2 of the discrimination-and-generalization task.

a. For comparison, group means (and standard deviation) are shown for the Normal-f and impaired-f groups, with these cases excluded.

in the current study, prediction accuracy could be improved to 88% by considering a short computerized learning and generalization test in addition to delayed paragraph recall.

In addition to replicating and extending these results, an important direction for future development will be attempting to improve prediction accuracy still further. In the current study, of 50 participants, 6 cases were misclassified by the discriminant function. As shown in Table 4, this included 3 participants who performed relatively well on LMII and ERR2 at baseline and were accordingly classified with a predicted outcome of Normal but who in fact were diagnosed as MCI at a 2-year follow-up. It remains an open question whether these 3 individuals—and, indeed, the rest of our MCI group—will go on to develop AD, or whether their cognitive impairments do not reflect prodromal dementia. Conversely, the discriminant function incorrectly classified 3 individuals as Impaired (based on poor LMII and ERR2 scores) who were in fact determined to be cognitively normal at follow-up. Again, it is conceivable that these individuals are experiencing prodromal AD at a stage where deficits are seen on these tasks but not yet in other clinical assessment. Only long-term tracking of the subjects can provide the answer.

The discriminant function did correctly classify 4 individuals who were diagnosed as MCI at baseline but who were rediagnosed as cognitively normal at the 2-year follow-up. All 4 of these individuals scored as well as or better than the Normal-f group mean on both LMII and ERR2, suggesting that the baseline clinical diagnosis might have been improved by including information relating to subjects' performance on these 2 tests.

The Hippocampal Connection

One reason why we wished to explore the paragraph recall and discrimination-and-generalization tasks as predictors of cognitive decline involves the hippocampus. As mentioned above, patients with mild AD show hippocampal volume reductions, and progressive hippocampal dysfunction has been proposed as a possible neuroanatomical basis of AD. Elderly individuals with MCI who show hippocampal atrophy on MRI are at heightened statistical risk for developing dementia relative to nonatrophied peers.[5]

To the extent that hippocampal atrophy may result in disruption or reduction of normal hippocampal function, nondemented individuals with hippocampal atrophy due to prodromal AD might show abnormal performance on tests of hippocampal function, before abnormalities begin to show up on other, more general tests of memory and cognition. Delayed paragraph recall is one such test, which is known to be dependent on the medial temporal lobes and to be disrupted in amnesic patients with bilateral medial temporal lobe damage,[6] as well as in nondemented

elderly individuals with hippocampal atrophy.[4] These findings are consistent with the widely held view that the hippocampus and related medial temporal lobe structures are important for the formation of new declarative (consciously accessible and easily verbalizable) memories.[25]

The generalization portion of our discrimination-and-generalization test is another test that is disrupted in amnesic patients with bilateral medial temporal lobe damage[13] and in nondemented elderly individuals with hippocampal atrophy[12]. These findings are consistent with a view that, in addition to being important for declarative memory formation, the hippocampus and related medial temporal lobe structures are important for new learning, which supports subsequent generalization or flexible use of information when conditions alter.[26-29] Our findings are also consistent with the idea that hippocampal system pathology occurs early in AD and that MCI may often reflect a prodromal stage of AD, when pathology has not yet accrued to the point of causing behavioral symptomatology sufficient for a diagnosis of dementia.

However, it is important to note that not all individuals diagnosed as MCI will inevitably develop AD nor will they all necessarily have hippocampal atrophy. In addition, it is conceivable that hippocampal atrophy could exist in some subjects due to nonprogressive causes or due to diseases other than AD. Thus, although our current results are consistent with the observed correlation between hippocampal atrophy and AD risk, we cannot exclude the possibility that our behavioral data reflect some other factors.

Limitations of the Current Study

Several limitations of the current study have been alluded to above; the major limitations of the current study are related to its scope, namely, this study involved a small number of individuals who were reassessed only once and a relatively short time after baseline testing. Clearly, larger-scale studies, with repeated assessments over a longer follow-up interval would provide more information about cognitive decline. This would also help elucidate the situation with those individuals who were given baseline diagnoses of MCI but were reclassified as cognitively normal at follow-up. On the basis of information from 2 time points alone, it is impossible to be sure whether one or the other diagnosis may have been incorrect or whether the baseline diagnosis of MCI was accurate but represented a short-term condition, resulting in transient memory impairments

rather than a progressive cognitive decline. Additionally, it would be of great interest to examine hippocampal volumes in subjects both at baseline and at follow-up and to attempt to correlate volumetric reductions with behavioral declines.

Another limitation of our study is related to the fact that we administered the paragraph recall and discrimination-and-generalization tests only at the time of baseline diagnosis. It would be preferable to administer these tasks at multiple sessions to allow the possibility of tracking cognitive decline over time. In fact, the presence in our current sample of several cases who were classified at cognitively impaired at baseline, but later reclassified as cognitively normal at follow-up, highlights the risk of overreliance on measurements taken at a single time point. Multiple testing sessions would reduce this risk. Studies are underway to determine the validity of repeat testing with the discrimination-and-generalization task to confirm, for example, whether healthy young adults administered this task at yearly intervals (with new object pairs each time) show a consistent level of performance on both the learning and the generalization components. If so, then it may be possible to use the discrimination-and-generalization test to track cognitive decline in older individuals, to identify a time point at which performance shows decrements, and therefore to predict incipient cognitive decline.

Another important limitation of this study is the sample itself. Subjects were recruited from an extremely healthy group of seniors from which individuals with any of the various diseases and disorders that are common in the elderly were excluded. These strict exclusion criteria resulted in a "clean" sample for study but meant that our sample was not necessarily representative of the elderly population as a whole. Future studies should attempt to replicate this study in more representative samples, with fewer exclusion criteria, to document the degree to which these findings apply to the general population. Such studies are planned and will include larger-scale samples, longer tracking periods, and repeat testing on the discrimination-and-generalization and paragraph recall tests.

Acknowledgments

gratefully acknowledge Alok Vedyas, Valerie Hutchison, Geoffrey Schnirman, and Suditi Gupta. Preliminary results were presented as a poster at the 2005 Society for Neuroscience Annual Meeting; November 12-16, 2005; San Diego, CA.

References

1. Kluger A, Ferris S, Golomb J, et al. Neurophysiological prediction of decline to dementia in nondemented elderly. *J Geriatr Psychiatry Neurol.* 1999;12:168-179.
2. Kluger A, Gianutsos J, Golomb J, et al. Patterns of motor impairment in normal aging, mild cognitive decline, and early Alzheimer's disease. *J Gerontol Psychol Sci Soc Sci.* 1997;52:P28-P39.
3. Grundman M, Petersen R, Ferris S, et al. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Arch Neurol.* 2004;61:59-66.
4. Golomb J, de Leon M, Kluger A, et al. Hippocampal atrophy in normal aging: an association with recent memory impairment. *Arch Neurol.* 1993;50:967-973.
5. de Leon M, Golomb J, George A, et al. The radiologic prediction of Alzheimer disease: the atrophic hippocampal formation. *Am J Neuroradiol.* 1993;10:3531-3542.
6. Squire L, Zola-Morgan S. The medial temporal lobe memory system. *Science.* 1991;253:1380-1386.
7. de Leon M, George A, Stylopoulos L, et al. Early marker for Alzheimer's disease: the atrophic hippocampus. *Lancet.* 1989;2:672-673.
8. Ball M. Neuronal loss, neurofibrillary tangles and granulovacuolar degeneration in the hippocampus with ageing and dementia. A quantitative study. *Acta Neuropathol (Berl).* 1977;37:111-118.
9. Hyman B, Van Horsen G, Damasio A, Barnes C. Alzheimer's disease: cell-specific pathology isolates the hippocampal formation. *Science.* 1984;225:1168-1170.
10. Moscovitch M, Winocur G. The neuropsychology of memory and aging. In: Craik F, Salthouse T, eds. *The Handbook of Aging and Cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates; 1992:315-372.
11. Ferris S, Kluger A. Commentary on age-related memory impairment, age-related cognitive decline and mild cognitive impairment. *Aging Neuropsychol Cogn.* 1996;3:148-153.
12. Myers C, Kluger A, Golomb J, et al. Hippocampal atrophy disrupts transfer generalization in non-demented elderly. *J Geriatr Psychiatry Neurol.* 2002;15:82-90.
13. Myers C, Bryant D, DeLuca J, Gluck M. Dissociating basal forebrain and medial temporal amnesic syndromes: insights from classical conditioning. *Integrative Physiol Behav Sci.* 2002;37:85-102.
14. Hamilton M. Development of a rating scale for primary depression illness. *Br J Soc Clin Psychol.* 1967;6:278-296.
15. Reisberg B, Ferris S, de Leon M, Crook T. The global deterioration scale for assessment of primary degenerative dementia. *Am J Psychiatry.* 1982;139:1136-1139.
16. Wechsler D. *WMS-R Wechsler Memory Scale-Revised Manual.* New York, NY: Psychological Corporation, Harcourt Brace Jovanovich; 1987.
17. Wechsler D. *WAIS-R Wechsler Adult Intelligence Scale-Revised Manual.* New York, NY: Psychological Corporation; 1981.
18. Folstein M, Folstein S, McHugh P. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12:189-198.
19. Petersen R, Thomas R, Grundman M, et al. Vitamin E and donepezil for the treatment of mild cognitive impairment. *N Engl J Med.* 2005;352:2379-2388.
20. Devanand D, Folz M, Gorlyn M, Moeller J, Stern Y. Questionable dementia: clinical course and predictors of outcome. *J Am Geriatr Soc.* 1997;45:321-328.
21. Grundman M, Petersen RC, Morris JC, et al. Rate of dementia of Alzheimer type (DAT) in subjects with mild cognitive impairment: the Alzheimer's disease cooperative study [abstract]. *Neurology.* 1996;46:A403.
22. Morris J, Heyman A, Mohs R, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neurophysiological assessment of Alzheimer's disease. *Neurology.* 1989;39:1159-1165.
23. Reisberg B, Ferris S, Shulman E, et al. Longitudinal course of normal aging and progressive dementia of the Alzheimer's type: a prospective study of 106 subjects over a 3.6 year mean interval. *Prog Neuropsychopharmacol Biol Psychiatry.* 1986;10:571-578.
24. Flicker C, Ferris S, Reisberg B. A two-year longitudinal study of cognitive function in normal aging and Alzheimer's disease. *J Geriatry Neuropsychol Neurol.* 1993;8:84-96.
25. Squire L. *Memory and Brain.* New York, NY: Oxford University Press; 1987.
26. Eichenbaum H, Stewart C, Morris R. Hippocampal representation in place learning. *J Neurosci.* 1990;10:3531-3542.
27. Gluck M, Myers C. Hippocampal mediation of stimulus representation: a computational theory. *Hippocampus.* 1993;3:491-516.
28. Gluck M, Myers C, Nicolle M, Johnson S. Computational models of the hippocampal region: implications for the prediction of risk for Alzheimer's disease in non-demented elderly. *Curr Alzheimer Res.* 2006;3:247-257.
29. Schacter D. Multiple forms of memory in humans and animals. In: Weinberger N, McGaugh J, Lynch G, eds. *Memory Systems of the Brain: Animal and Human Cognitive Processes.* New York, NY: Guildford Press; 1985: 351-379.