Neuropsychology
1993, Vol. 7, No. 3, 406-416

0894-4105/93/$3.00

Stability of Everyday Memory in Age-Associated Memory Impairment: A Longitudinal Study

James R. Youngjohn and Thomas H. Crook III

Everyday memory performance was examined longitudinally in 2 groups of Ss meeting the diagnostic criteria for age-associated memory impairment (AAMI). One group of 157 participants in a drug trial for reversing memory loss in AAMI was tested over multiple sessions. The other group of 75 persons did not participate in a drug trial and thus was tested only twice. Both groups were retested for longitudinal follow-up about 4 years after their initial session. Follow-up test performance remained fairly stable relative to initial performance in both groups. The drug study group showed large practice effects during the course of the drug studies, but these effects subsided after the drug studies' end. Implications regarding memory decline in the normal elderly and neuropsychological measurement issues are discussed.

It is widely recognized that many cognitive functions often decline in the later decades of adulthood, particularly learning and memory (Fozard, 1985; Poon, 1985; Youngjohn & Crook, in press). Although there have been numerous cross-sectional investigations confirming the deleterious effects of advancing age on memory and cognition, relatively few longitudinal studies have been completed. Noteworthy exceptions include the Baltimore longitudinal study (Alder, Adam, & Arenberg, 1990), the Duke longitudinal study (Siepler, McCarty, & Logue, 1982), and the Seattle longitudinal study (Cooney, Schaie, & Willis, 1988).

Longitudinal results have generally indicated that there are modest, gradual cognitive declines in older subjects, with the inflection point occurring somewhere around age 60 (Hertzog & Schaie, 1988). Mitrushina and Satz (1991) found that most subjects improved over three 1-year testing intervals, except their oldest subjects, who remained relatively stable. These results illustrate one of the hazards of conducting longitudinal research: practice effect artifact.

Until recently, there has been no generally accepted diagnostic classification for normal, healthy persons who experience declining memory and cognition in their later years. Kral (1962, 1966) introduced the term *benign senescent forgetfulness* to describe otherwise healthy elderly individuals who experience cognitive declines relative to their age peers. More recent studies have validated this concept (Larrabee, Levin, & High, 1986). However, this nosological framework fails to address the larger number of individuals experiencing memory loss associated with normal developmental processes.

A work group sponsored by the National Institute of Mental Health proposed diagnostic criteria for the classification of age-associated memory impairment (AAMI; Crook et al., 1986). This nosological category refers to persons more than 50 years of age who both complain of memory impairment in tasks of daily life and have objective memory test performance at least one standard deviation below the mean established for young adults.

A recent longitudinal epidemiological study (Lane & Snowdon, 1989) reported relatively high prevalence and incidence rates for AAMI (34.93% [*SE* = 4.54%] and 6.63% [*SE* = 9.41%], respectively, per annum), as compared with prevalence and incidence rates for probable Alzheimer's disease (13.01% [*SE* = 7.11%] and 3.06% [*SE* = 9.79%], respectively, per annum). The high incidence and prevalence rates for AAMI emphasize the need for further understanding of this area. Although the Lane and Snowdon (1989) study suggests that most subjects with AAMI do not progress to Alzheimer's disease, no large-scale longitudinal study has actually measured the cognitive functioning of AAMI patients over extended periods of time.

James R. Youngjohn and Thomas H. Crook III, Memory Assessment Clinics, Inc., Scottsdale, Arizona.

We would like to thank Lisa Burrows for her insightful comments on the article.

Correspondence concerning this article should be addressed to James R. Youngjohn, 7434 East Stetson Drive, Suite 250, Scottsdale, Arizona 85251.

The purpose of the present investigation was to better understand the progression of AAMI. It has been argued that AAMI is the result of cognitive decline seen in all elderly persons (Bamford & Caine, 1988; Blackford & LaRue, 1989; Smith et al., 1991). At present, the rate of cognitive decline in normal elderly people is unclear. Furthermore, the possibility that AAMI may actually be the result of very early Alzheimer's disease has not yet been conclusively ruled out (Youngjohn, Larrabee, & Crook, 1992a). Consequently, a longitudinal study of this population would be of considerable interest.

Longitudinal studies that rely on psychometric variables as outcome measures are complicated by the well-described phenomenon of practice effects (e.g., Youngjohn, Larrabee, & Crook, 1992b). Simply, performance on ability tests tends to improve on each subsequent administration because the subject has increasing familiarity with the material. Although practice effects in memory testing can be minimized by using alternate test forms of equivalent difficulty (Crook, Youngjohn, & Larrabee, 1992), they may not be eliminated entirely because the format of the instrument remains the same. Consequently, our study used two experimental groups, one in which practice effects were minimized and one that involved multiple administrations of parallel forms of the same tests.

Method

Subjects

Seven hundred seventy-two unpaid volunteers were recruited through newspaper and radio advertising. They were initially evaluated for participation in one of several trials, supervised by the U.S. Food and Drug Administration, of candidate pharmaceutical compounds for treating or reversing memory loss associated with normal aging. All subjects satisfied the diagnostic criteria for AAMI (Crook et al., 1986). Specifically, participants had to be more than 50 years old, report gradual decline of everyday memory function, have at least one objective memory test performance falling at least one standard deviation below the mean for young adults (Wechsler Memory Scale [WMS; Wechsler, 1945] Logical Memory, 6 or less; WMS Paired Associate Learning hard pairs, 6 or less; and Benton Visual Retention Test [BVRT; Benton, 1974] correct, 7 or less), show evidence of adequate intelligence as determined by a Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955) Vocabulary raw score of at least 32, and show absence of dementia as determined by a score of 24 or higher on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975). Exclusion criteria included self-report on a health history questionnaire of any medical, psychiatric, or neurologic disorder that could produce cognitive deterioration. All subjects who were actually enrolled in a drug study also received comprehensive medical and neurologic workups, including computed tomography (CT) scans of the brain.

Of the original 772 patients who satisfied the criteria for AAMI, 331 elected to participate in one of a dozen drug trials investigating treatment effects. The remaining 441 subjects declined to participate because of personal choice, incompatible medications, and so forth. For this study, we attempted to recontact the original 772 subjects by letter and telephone.

Two hundred thirty-two subjects (82 men and 150 women) agreed to come in, yielding an overall longitudinal study participation rate of 30%. The ages of these participants on initial evaluation ranged from 50 to 77 (M = 62.08, SD = 6.53). Years of education ranged from 6 to 26 (M = 15.41, SD = 2.69). The mean Affective Rating Scale score (Yesavage et al., 1983) was in the nondepressed range (6.36, SD = 4.87), with 90% of the participants obtaining a score of 10 or less.

We divided our participants into two subgroups. The first group consisted of 75 persons (27 men and 48 women) who were not enrolled in a drug study after their initial assessment. The longitudinal study response rate for these non–drug study participants was 17%. These subjects were tested only two times (initial and long-term follow-up). Their ages at the beginning of the investigation ranged from 51 to 74 (M = 61.93, SD = 6.05). Years of education ranged from 12 to 22 (M = 15.82, SD = 2.37).

The second group consisted of 157 individuals (55 men and 102 women) who were actually enrolled in a study and were enlisted in either the drug or placebo condition. The response rate of these drug study participants was 47%. Because all studies required periodic monitoring of treatment effects, these participants had been administered different forms of the same everyday memory tests across multiple trials (mean number of test administrations = 7.16, SD = 3.67). These subjects ranged in age from 50 to 77 (M = 62.15, SD = 6.77), and years of education ranged from 6 to 26 (M = 15.21, SD = 2.81).

Memory Assessment Clinics (MAC) has investigated 12 candidate pharmaceutical agents to date. Only one of these agents, a phospholipid compound (Crook et al., 1991), has shown a positive treatment effect when compared with the placebo control group. Given the lack of a positive drug effect, we combined the drug and placebo control groups for the purposes of the present longitudinal investigation.

Apparatus

A number of computer-simulated everyday memory tests have been developed at MAC to measure ecologically meaningful treatment outcome. They are administered in a standardized manner with a 19-in. (48.26-cm) color monitor with a touch screen, interfaced with a computer equipped with a

20-megabyte hard-disk drive, a laser-disk player, and customized computer graphics hardware. Participants do not use the computer keyboard; rather, they respond verbally, with the touch screen or with a touch-tone or rotary telephone. The tester is present throughout the session and sits behind and generally out of view of the participant. Test administration and scoring are controlled by the tester with a separate computer monitor.

Procedure

All subjects were tested in a controlled, clinical environment. Each subsequent administration of a test involved a different parallel form, when available. There are now eight parallel forms of each of the everyday memory tests (Crook et al., 1992). The participants in the present study were contacted a number of years after the initial evaluation or the completion of the drug study and were asked to return for follow-up testing. In the case of those who did not participate in drug studies, the average test–retest interval was 4.6 years (M = 1,683 days, SD = 292, range = 1,035 days to 2,130 days).

Those subjects who participated in drug studies were enrolled in either the drug or placebo condition, as noted earlier, and returned at frequent intervals for retesting so that their progress could be monitored for the duration of the study. The mean length of time that elapsed between their initial evaluation and their last drug study visit was slightly less than a year (M = 348 days, SD = 283). The drug study participants were contacted for the long-term follow-up assessment several years after the study had been completed and they had discontinued the experimental drug or placebo, or both. The average time that elapsed between the end of the drug study and the long-term follow-up examination was 3 years (M = 1,085 days, SD = 461, range = 58 days to 1,996 days).

In the following paragraphs, we describe the dependent variables, including the MAC computer-simulated everyday memory tests and the traditional memory tests.

Name–Face Association. In this test, subjects are presented with live recordings of individuals introducing themselves by common first names. Recall is assessed by showing the same individuals in a different order, stating the name of the city in which they reside, and asking the subject to provide the name of each person. This study used two learning trials in which 14 name–face pairs were presented and recall was assessed. The total number of name–face pairs recalled was summed over both learning trials. The test has been shown to be a sensitive measure of age-related memory decline (Crook & West, 1990).

First–Last Names. This is a verbal learning test using the WMS paired associates paradigm. Subjects are presented with a series of six first–last name pairs, followed by a presentation of the last name only; the task of subjects is to recall the first name. There were two learning trials in this investigation. The First–Last Names test is sensitive to age effects and loads highly on a verbal memory factor (Youngjohn, Larrabee, & Crook, 1991).

Grocery List Selective Reminding Test (GLSRT). This test follows the standard selective reminding paradigm (Buschke, 1973), using common grocery-list items as the stimuli to be recalled. On the first trial, subjects are requested to read aloud a list of 15 grocery items that appear on the video screen. After the recall attempt, the words that the subjects have not recalled briefly reappear on the screen; subjects must then attempt to recall the entire list, having been selectively reminded of the words that they omitted on the previous trial. Our dependent variable in this study was the total number of words recalled on the first two learning trials. The GLSRT has been demonstrated to be negatively related to age and to load highly on a verbal memory factor (Youngjohn et al., 1991).

Telephone Dialing (with interference). This task is a variation of the standard digit repetition paradigm. Participants are presented with a 10-digit (long distance) telephone number on the monitor screen and asked to read it aloud. The number then disappears from the screen and they are instructed to dial the number on a touch-tone phone interfaced with the computer. After the subjects have completed dialing, they hear a busy signal. The instruction "Please Redial" appears on the video screen, and they are asked to redial. Credit is given for each digit dialed in the correct position, regardless of errors made elsewhere in the sequence. West and Crook (1990) demonstrated that older subjects are particularly sensitive to the interference condition of this test.

Recognition of Faces—Delayed, Nonmatching to Sample. Subjects are presented with facial photographs on the touch-screen monitor, with a new face being added to the array until there are 25 faces. Subjects are required to identify the new face added by touching it on the monitor. In the present study, we examined the number of correctly completed trials before the first error was committed. Previous factor-analytic studies have demonstrated this test to be a relatively robust measure of visual memory (Larrabee & Crook, 1989). An earlier version was shown to discriminate between normal and demented elderly subjects (Flicker, Ferris, Crook, & Bartus, 1987).

Misplaced objects. This is a test of object location recall. Subjects are required (using the touch screen) to place 20 common objects into the schematic representation of a 12-room house (maximum of 2 objects per room). Delayed recall is tested at 40 min by requesting subjects to press the room in which they previously placed each of the objects. A second attempt is allowed if they miss on the first attempt, to closely simulate the common activity of looking for a lost article. Our dependent variable in the present investigation was the total number of objects correctly located after both attempts. Previous studies have suggested that the misplaced objects test is a measure of verbal–visual associative memory (Crook, Youngjohn, & Larrabee, 1990). It is structurally similar to a paradigm used in animal model pharmacological

research (Bartus, Dean, & Beer, 1983; Bartus, Fleming, & Johnson, 1978; Bartus & Johnson, 1976).

Divided Attention Recall. This test combines two tasks, a variation on a standard reaction time paradigm in which the task of driving a car and responding appropriately to changing traffic signals is simulated and a variation of the WMS Logical Memory subtest. Subjects are instructed to change brake and accelerator pedals appropriately and as quickly as possible in response to traffic light changes.

Subjects listen to simulated radio weather and traffic reports while performing the simulated driving task. After each broadcast is presented, immediate recall is assessed in a manner similar to that of the WMS Logical Memory subtest; the score is the number of ideas recalled, averaged between the two stories. Larrabee and Crook (1989) reported that performances on this test, the WMS Logical Memory subtest, and the WAIS Vocabulary subscale all loaded on a verbal memory and intelligence factor.

Traditional tests. The following traditional instruments were administered in the standard format: the WMS Logical Memory and Paired Associate Learning subtests (Wechsler, 1945), the WAIS Vocabulary subtest (Wechsler, 1955), and the BVRT (Benton, 1974).

Results

Long-Term Follow-Up Responders Versus Nonresponders

Because of the relatively high rate of attrition in our study, we compared participants who did not come in for long-term follow-up with those who did to assess the potential for selective attrition bias. The ages of the responders (M = 62.08 years, SD = 6.53) and nonresponders (M = 61.36 years, SD = 6.86) were not significantly different ($p > .05$). There was a slightly higher ratio of women in the responding sample (65%) as opposed to the nonresponding group (56%; $p < .05$). Education level was slightly higher in the nonresponding group (M = 16.09 years, SD = 2.95) than the responding group (M = 15.41 years, SD = 2.69; $p < .05$).

To determine whether the effects of selective attrition were significantly biasing our results, we compared initial test performances in the responding and nonresponding groups. A multivariate analysis of variance (MANOVA) examining the seven everyday memory test variables and the five traditional memory test variables in the two groups was not significant ($p > .05$). Although the MANOVA did not reveal significant overall differences, independent univariate analyses of variance (ANOVAs) suggested that the participants who came back for long-term follow-up performed better than those who did not on two of the seven everyday memory tests (GLSRT and telephone dialing; $p < .05$) and three of the five traditional measures (WMS Logical Memory, BVRT correct, and WAIS Vocabulary). Initial performances on name–face association, First–Last Names, Recognition of Faces—Delayed, Nonmatching to Sample, Misplaced Objects, Divided Attention Recall, WMS Paired Associate Learning, and BVRT errors did not differ significantly between the two groups ($p > .05$).

Long-Term Follow-Up in Non–Drug Study Participants

In our investigation of longitudinal effects on those participants who came in for long-term follow-up, we first looked at the subgroup in which practice effects were minimized (i.e., those subjects who did not participate in drug studies). Table 1 presents mean performances on the initial and long-term follow-up sessions. We used two statistical techniques to analyze our data: repeated measures MANOVA/ANOVA and multiple regression analysis.

The repeated measures MANOVA comparing initial everyday memory and traditional test performances with long-term follow-up performances was significant, Hotelling's T = 7.12, approximate $F(7, 12) = 4.15$, $p < .05$. A series of univariate repeated measures ANOVAs was then conducted. Table 2 presents F values, degrees of freedom, significance levels, and effect sizes (η^2). Inspection of Table 2 shows that long-term follow-up performance significantly differed from initial performance on only four of seven everyday memory tests (i.e., Name–Face Association, GLSRT, Telephone Dialing, and Divided Attention Recall) at the .05 level. Marginally significant differences ($p < .06$) were noted for two traditional measures, WMS Logical Memory and BVRT errors. Effect sizes were generally modest. Furthermore, these differences did not necessarily conform to the expected pattern of decline over time. Inspection of Table 1 shows that test performance actually improved on long-term follow-up for both Name–Face Association and BVRT errors.

Initial test performance and amount of elapsed time were used to predict long-term follow-up everyday memory and traditional test performance in the non–drug study participants in a series of multiple regression analyses. Initial performance was entered into the equation first, and quantity of elapsed time was entered

second. Table 3 presents squared multiple correlations for initial everyday and traditional memory test performance and the change in those values for elapsed time, as well as significance levels.

As expected, subjects' initial performance on a measure was a powerful predictor of their subsequent performance on that same test at follow-up. Initial test performance was a significant predictor of follow-up performance on five of seven everyday memory tests; the only exceptions were Recognition of Faces—Delayed, Nonmatching to Sample and Misplaced Objects. Proportions of variance in follow-up performance accounted for by initial performance in those equations that were significant at the .05 level ranged from a high of 47% for Name–Face Association to a low of 13% for Divided Attention Recall.

In contrast, the amount of elapsed time between the initial and follow-up sessions was a very poor predictor of follow-up performance. The addition of elapsed time resulted in a significant increase in predictive accuracy at the .05 level on only three measures: the GLSRT, on which it accounted for an additional 6% of performance variance; WMS Logical Memory, on which elapsed time accounted for an additional 3% of performance variance; and WAIS Vocabulary, on which elapsed time accounted for 2% of additional variance.

Table 1 shows very slight decreases in performance on long-term follow-up for GLSRT, telephone dialing, WMS Logical Memory, and WMS Paired Associate Learning. A slight increase in performance was noted on WAIS Vocabulary. Negative associations between follow-up performance and elapsing time were confirmed by negative standardized beta weights for telephone dialing ($\beta = -.244$), WMS Logical Memory ($\beta = -.196$), and WMS Paired Associate Learning ($\beta = -.111$). However, increasing elapsed time was actually associated with improved long-term follow-up performance, as demonstrated by positive beta weights for GLSRT ($\beta = .239$) and WAIS Vocabulary ($\beta = .152$).

Drug Study Participants

We examined the drug study participants in the second phase of the study. The effects of both time and practice were considered in this stage of the investigation. Table 4 presents means for initial test performances, performances at the end of the study, and performances on long-term follow-up. Three repeated measures MANOVAs were conducted: (a) Initial performances were compared with performances at the end of the drug study, (b) performances at the end of the drug study were compared with long-term follow-up performances, and (c) initial performances were compared with long-term follow-up performances.

We first compared initial test performances with performances at the end of the drug study, which involved multiple administrations of different forms of the same instruments. The repeated measures MANOVA was significant, Hotelling's $T = 1.44$, approximate $F(6, 51) = 12.27$, $p < .001$.

Table 1
Initial and Long-Term Follow-Up Performance on Everyday Memory and Traditional Memory Tests in Non–Drug Study Participants

Test	*n*	Visit 1		Follow-up visit	
		M	*SD*	*M*	*SD*
Name–Face Association	72	9.47	5.23	11.00	5.12
First–Last Names	50	4.16	2.22	4.12	2.07
Grocery List Selective Reminding Test	49	30.61	6.20	29.48	5.58
Telephone Dialing	45	4.59	1.84	3.80	2.24
Recognition of Faces—Delayed, Nonmatching to Sample	45	12.60	6.38	12.55	5.97
Misplaced Objects	49	14.12	2.67	14.44	2.52
Divided Attention Recall	45	20.16	2.94	19.21	3.74
Wechsler Memory Scale					
Logical Memory	71	8.84	2.79	8.22	2.22
Paired Associate Learning	73	14.52	3.18	14.23	3.61
Benton Visual Retention Test					
Correct	73	6.48	1.43	6.66	1.85
Error	73	5.45	2.57	4.73	3.02
WAIS Vocabulary	78	64.32	9.80	65.06	9.20

Note. WAIS = Wechsler Adult Intelligence Scale.

Table 2
Repeated Measures Analysis of Variance Summary Data for Non–Drug Study Participants

Test	F	df	α	η^2
Name–Face Association	10.76	1, 71	.002*	.13
First–Last Names	0.25	1, 49	.662	.01
Grocery List Selective Reminding Test	4.10	1, 49	.048*	.08
Telephone Dialing	7.08	1, 43	.011*	.14
Recognition of Faces—Delayed, Nonmatching to Sample	0.75	1, 45	.392	.02
Misplaced Objects	0.00	1, 58	.967	.00
Divided Attention Recall	5.24	1, 40	.027*	.12
Wechsler Memory Scale				
Logical Memory	3.73	1, 70	.058	.05
Paired Associate Learning	0.49	1, 72	.488	.01
Benton Visual Retention Test				
Correct	0.63	1, 72	.432	.01
Errors	3.67	1, 72	.059	.05
WAIS Vocabulary	1.66	1, 69	.203	.02

Note. WAIS = Wechsler Adult Intelligence Scale.
* $p < .05$.

Table 5 presents the results from the repeated measured ANOVAs. Inspection of Table 5 reveals that significant improvements had occurred on all seven of the everyday memory tests by the completion of the drug study (at the .05 level). Effect sizes ranged from a high of 44% of variance accounted for by practice in name–face association to a low of 2% for first–last names.

We further examined the influence of practice effects by regressing initial performances against test performances at the end of the drug study, entering the number of test administrations occurring between the initial evaluation and last drug study session, and, finally, entering the amount of intervening time between the initial and last drug study session. Table 6 presents squares multiple correlations for initial test performance, and the change in those values for number of administrations and elapsed time, as well as significance levels for each stage of the analyses.

As expected, subjects' initial performance on a measure was a powerful predictor of their performance on a different form of the same test at the end of a drug study. Initial performance was significantly associated with performance at the end of the drug study, at the .05 level, on all seven everyday memory tests. Squared multiple correlations ranged from a high of .39 for telephone dialing to a low of .05 for recognition of faces—delayed nonmatching-to-sample (see Table 6).

Adding the number of visits to the equation resulted in significant increases in predictive accuracy for three of the seven everyday memory tests at the .05 level (Name–Face Association, GLSRT, and Divided Attention Recall). Additional proportions of performance variance accounted for by adding the number of visits were 9% for Name–Face Association, 6% for Divided Attention Recall, and 4% for GLSRT (see Table 6). Table 4 demonstrates improvements in performance from the initial visit to the last drug study visit on all everyday memory tests, suggesting that increased exposure to the tests resulted in improved performance, as predicted. Beta weights for number of visits had positive values across all seven everyday memory tests, further demonstrating that increased exposure resulted in improving test performance.

Adding the amount of elapsed time to the equation predicting performance at the end of the drug study resulted in significant improvement in predictive accuracy on only one everyday memory test (Name–Face Association) at the .05 level (see Table 6). Interestingly, a positive beta weight demonstrated that increases in elapsing time were actually associated with improvements in performance on this measure (β = .144). This finding probably reflects additional influence of practice effects not picked up by the number of visits. However, the additional variance in Name–Face Association performance accounted for by elapsing time was minimal (ΔR^2 change = .02).

Table 3
Regression Analyses for Non–Drug Study Participants

	First administration		Elapsed time	
Test	R^2	α	ΔR^2	α
Name–Face Association	.47	.001*	.01	.265
First–Last Names	.19	.002*	.00	.710
Grocery List Selective Reminding Test	.34	.001*	.06	.040*
Telephone Dialing	.32	.001*	.06	.054
Recognition of Faces—Delayed, Nonmatching to Sample	.00	.800	.03	.221
Misplaced Objects	.03	.210	.00	.683
Divided Attention Recall	.13	.021*	.01	.584
Wechsler Memory Scale				
Logical Memory	.25	.001*	.03	.003*
Paired Associate Learning total	.16	.001*	.01	.078
Benton Visual Retention Test				
Correct	.12	.001*	.01	.162
Errors	.18	.001*	.01	.411
WAIS Vocabulary	.73	.001*	.02	.001*

Note. Initial test performance was entered first as a predictor, and amount of elapsed time between initial and follow-up sessions was entered second. WAIS = Wechsler Adult Intelligence Scale.
* $p < .05$.

Table 4
Everyday Memory Test Performance in Drug Study Participants

Test	*n*	Visit 1		Last drug study visit		Follow-up	
		M	*SD*	*M*	*SD*	*M*	*SD*
Name–Face association	155	9.68	4.47	14.17	5.61	13.19	5.08
First–Last Names	102	4.04	2.12	4.59	2.31	3.82	2.18
Grocery List Selective Reminding Test	131	29.95	5.54	32.23	5.85	29.07	6.31
Telephone Dialing	119	4.04	2.26	4.82	2.21	4.05	2.15
Recognition of Faces—Delayed, Nonmatching to Sample	130	10.87	5.17	13.72	6.40	14.64	6.18
Misplaced Objects	106	14.58	2.91	16.24	2.74	14.31	2.56
Divided Attention Recall	126	15.63	6.37	19.36	7.24	14.14	6.41

The effects of elapsed time between the final drug study administration and the long-term follow-up session were then considered. The repeated measures MANOVA was significant, Hotelling's $T = 1.25$, approximate $F(6, 56) = 11.67$, $p < .001$.

The results from repeated measures ANOVAs comparing performances at the end of the drug study with performances on long-term follow-up are presented in Table 7. Inspection of Table 7 demonstrates that significant declines in performance between the end of the drug study and the long-term follow-up evaluation occurred on all but one of the everyday memory tests (Recognition of Faces—Delayed, Nonmatching to Sample) at the .05 level. Effect sizes ranged from a high of .41 for Divided Attention Recall to a low of .05 for Name–Face Association.

Test performance at the end of the drug study and amount of elapsed time were used to predict long-term follow-up test performance in the drug study participants in an additional series of multiple regression analyses. As expected, the final drug study test score was a significant predictor of long-term follow-up performance on all seven everyday memory tests at the .05 level. As shown in Table 8, R^2 ranged from a high of .49 for Name–Face Association to a low of .12 for Recognition of Faces—Delayed, Nonmatching to Sample.

However, the addition of the elapsed time between the end of the drug study and the long-term follow-up evaluation significantly added to the predictive accuracy of the equation for only one everyday memory test, Divided Attention Recall, at the .05 level; on this test, elapsed time accounted for an additional 6% of the variance in performance (see Table 8). The negative beta weight ($\beta = -.238$) confirms that increasing elapsed time was associated with declining recall performance on the Divided Attention task. The relative lack of association between the amount of elapsed time between the end of the drug study and the long-term follow-up session suggests that the significant declines in performance noted earlier were the result of the attenuation of practice effects rather than an actual longitudinal decline of cognitive ability.

In a final repeated measures MANOVA, we compared the drug study participants' initial test performances with their final test performances at long-term follow-up. The repeated measures MANOVA was again significant, Hotelling's $T = .81$, approximate $F(6, 87) = 11.78$, $p < .001$.

The summary of the univariate repeated measures ANOVAs presented in Table 9 reveals significant improvements on long-term follow-up on two everyday memory measures, Name–Face Association and Recognition of Faces—Delayed, Nonmatching to Sample; the effects of both intervening time and experience accounted for 38% of the variance in performance on the former test and 19% of the variance in performance on the latter test. A significant decline was noted on one

Table 5
Repeated Measures Analysis of Variance Summary Data Comparing Initial Scores With Scores at the End of the Drug Study

Test	*F*	*df*	α	η^2
Name–Face Association	122.91	1, 154	.001*	.44
First–Last Names	5.51	1, 95	.021*	.02
Grocery List Selective Reminding Test	21.63	1, 115	.001*	.16
Telephone Dialing	21.99	1, 118	.001*	.16
Recognition of Faces—Delayed, Nonmatching to Sample	20.36	1, 121	.001*	.14
Misplaced Objects	33.45	1, 87	.001*	.28
Divided Attention Recall	50.29	1, 113	.001*	.31

* $p < .05$.

Table 6
Regression Analyses of Drug Study Participants Predicting Final Drug Study Test Performance

Test	First visit		No. of visits		Elapsed time	
	R^2	α	ΔR^2	α	ΔR^2	α
Name–Face Association	.24	.001*	.09	.001*	.02	.045*
First–Last Names	.07	.010*	.01	.630	.00	.526
Grocery List Selective Reminding Test	.35	.001*	.04	.005*	.00	.875
Telephone Dialing	.39	.001*	.01	.093	.01	.112
Recognition of Faces—Delayed, Nonmatching to Sample	.05	.012*	.00	.439	.01	.198
Misplaced Objects	.25	.001*	.01	.409	.00	.668
Divided Attention Recall	.25	.001*	.06	.003*	.00	.517

Note. Visit scores were entered first, number of visits during the drug study was entered second, and elapsed time of the study was entered third.
* $p < .05$.

everyday memory test, the GLSRT, but the effect size was minimal ($\eta^2 = .04$).

Discussion

The most notable finding of the present investigation is the relative lack of decline over time in the everyday memory performance in either of our experimental groups. Although we found a number of significant longitudinal declines in performance, the proportions of performance variance accounted for by elapsing time were generally small. Furthermore, we occasionally found significant longitudinal effects that were in the opposite direction of our predictions (i.e., increasing elapsed time between assessments was actually associated with slightly improving, rather than declining, test performance).

These results present strong evidence that the condition of AAMI is not rapidly progressive but, rather, remains relatively stable. Consequently, AAMI does not appear to be caused by a progressive neuropathologic process such as very early Alzheimer's disease, because our subjects showed minimal deterioration on long-term follow-up. Rather, AAMI is more likely to be associated with relatively subtle memory and cognitive declines in the elderly that occur over decades instead of months or years.

Table 7
Repeated Measures Analysis of Variance Summary Data Comparing Scores at the End of the Drug Study With Scores at Long-Term Follow-Up

Test	F	df	α	η^2
Name–Face Association	8.65	1, 156	.004*	.05
First–Last Names	14.37	1, 99	.001*	.13
Grocery List Selective Reminding Test	48.66	1, 115	.001*	.30
Telephone Dialing	19.28	1, 120	.001*	.14
Recognition of Faces—Delayed, Nonmatching to Sample	1.15	1, 122	.286	.01
Misplaced Objects	40.02	1, 87	.001*	.32
Divided Attention Recall	75.89	1, 109	.001*	.41

* $p < .05$.

This lack of longitudinal progression in AAMI suggests that in terms of memory and cognitive test performance, AAMI is essentially equivalent to the effects of normal aging. A number of investigators have taken the position that the only factor that differentiates normal aging from AAMI is the patient's complaint that his or her memory is not what it used to be (Bamford & Caine, 1988; Blackford & LaRue, 1989; Smith et al., 1991). Our results tend to support this position. Of course, although cognitive and memory declines associated with aging may be normal, no one would argue that they are desirable, and few would fault elderly persons for complaining of them.

We attribute the significant improvements that occurred over the course of the drug studies to practice effects. Although the reader might be tempted to speculate that these improvements were actually the result of drug treatment effects, we have seen positive effects relative to placebo controls in only one compound of the dozen that we have examined (Crook et al., 1991). Our finding of significant practice effects on everyday memory tests across multiple administrations extends results of previous investigations (Youngjohn et al., 1992b) in which everyday memory test performance improved on a single readministration. Notably, strong practice effects were observed in spite of the use of multiple parallel test forms. The use of parallel test forms of equivalent difficulty has been proposed as a

Table 8
Regression Analyses Examining Predictive Influence of Last Drug Study Visit and Adding Amount of Elapsed Time to Predict Long-Term Follow-Up Performance

Test	Last drug study visit		Elapsed time	
	R^2	α	ΔR^2	α
Name–Face Association	.49	.001*	.00	.900
First–Last Names	.20	.002*	.00	.710
Grocery List Selective Reminding Test	.42	.001*	.00	.431
Telephone Dialing	.24	.001*	.00	.643
Recognition of Faces—Delayed, Nonmatching to Sample	.12	.001*	.00	.645
Misplaced Objects	.21	.001*	.01	.437
Divided Attention Recall	.24	.001*	.06	.004

* $p < .05$.

means of attenuating practice effects when performing serial evaluations over time (Crook et al., 1992). Serial evaluations are required to assess the treatment efficacy of pharmacologic therapies and rehabilitation programs, as well as the rate of disease progression, in diverse conditions such as Alzheimer's disease, head injury, stroke, and so forth.

The results of the present investigation clearly demonstrate that learning occurs with repeated exposure to a procedure of a particular memory test, even when the actual stimuli to be remembered change. Consequently, the use of parallel forms of equivalent difficulty does not appear to be sufficient to eliminate practice effects entirely. Similar results have also been reported by Goldstein (1991), who suggested that double-blind placebo control groups continue to be essential when conducting pharmaceutical trials.

It is notable that when our drug study ended and the practice stopped, subjects tended to revert to their initial levels of performance on most, but not all, measures on long-term follow-up evaluation. Chelune (1991) has suggested that there may be a critical period of elapsed time in neuropsychological testing after which the size of practice effects can be expected to be minimized. Our results support this hypothesis. It would be of considerable interest and clinical utility to determine the specific amount of time that must pass after which a test could be readministered without concern regarding practice effect artifact. Our results suggest that, for many measures, it is possible to retest after a period of 3 years and be fairly confident that practice effects will be minimal. Future investigators are encouraged to determine whether briefer test–retest intervals with minimal practice effects are possible.

There are several potential weaknesses in our study. For example, our participants were highly educated and may not be representative of the general population. In addition, the literature has demonstrated a consistent selective attrition bias in which participants who remain in longitudinal studies tend to have higher initial levels of cognitive functioning than those who do not (Mitrushina & Satz, 1991). Furthermore, those subjects who drop out during the course of longitudinal studies tend to be declining at faster rates than those who do not (Cooney et al., 1988; Siepler et al., 1982).

Although initial differences between the long-term follow-up responders and nonresponders in our investigation appear to have been small, it is certainly possible that the participants who did not come in for retesting had greater longitudinal declines in their cognitive function than those who did come in. The long-term follow-up response rate of the subjects who originally had declined to participate in one of the drug studies was low (17%), raising the chances of selective attrition bias. On the other hand, the response rate for the drug study participants was much better (47%), increasing confidence in the representativeness of this particular group. The increased response rate of the drug study participants perhaps reflects their greater involvement with and commitment to research. Of course, the ideal longitudinal study would entail follow-up testing of all subjects (in which the attrition rate would be minimized).

Finally, our study was limited by relatively brief longitudinal follow-up intervals. It would be of considerable interest to measure the cognitive function and self-report of normal adults over much greater time spans. Ideally, initial measurement would occur prior to the onset of AAMI, perhaps when subjects are in

Table 9
Repeated Measures Analysis of Variance Summary Data Comprising Initial Scores With Scores at Long-Term Follow-Up in Drug Study Participants

Test	F	df	α	η^2
Name–Face Association	94.64	1, 154	.001*	.38
First–Last Names	1.05	1, 101	.309	.01
Grocery List Selective Reminding Test	4.71	1, 129	.032*	.04
Telephone Dialing	0.00	1, 117	.961	.00
Recognition of Faces—Delayed, Nonmatching to Sample	29.81	1, 125	.001*	.19
Misplaced Objects	0.07	1, 93	.789	.00
Divided Attention Recall	3.30	1, 118	.072	.03

* $p < .05$.

their 40s, with periodic assessment of the incidence and prevalence of AAMI occurring throughout the remaining natural life span.

References

Alder, A. G., Adam, J., & Arenberg, D. (1990). Individual-differences assessment of the relationship between change in and initial level of adult cognitive functioning. *Psychology and Aging, 5,* 560–568.

Bamford, K. A., & Caine, E. D. (1988). Does "benign senescent forgetfulness" exist? *Clinics in Geriatric Medicine, 4,* 897–915.

Bartus, R. T., Dean, R. L., & Beer, B. (1983). An evaluation of drugs for improving memory in aged monkeys: Implications for clinical trials in humans. *Psychopharmacology Bulletin, 19,* 168–184.

Bartus, R. T., Fleming, D., & Johnson, H. R. (1978). Aging in the Rhesus monkey: Effects on short-term memory. *Journal of Gerontology, 33,* 858–871.

Bartus, R. T., & Johnson, H. R. (1976). Short-term memory in the Rhesus monkey: Disruption from the anticholinergic scopolamine. *Pharmacology, Biochemistry and Behavioral Science, 5,* 39–46.

Benton, A. L. (1974). *The revised Visual Retention Test* (4th ed.). New York: Psychological Corporation.

Blackford, R. C., & LaRue, A. (1989). Criteria for diagnosing age-associated memory impairment: Proposed improvements from the field. *Developmental Neuropsychology, 5,* 295–306.

Buschke, H. (1973). Selective reminding for analysis of memory and learning. *Journal of Verbal Learning and Verbal Behavior, 12,* 543–550.

Chelune, G. J. (1991, August). *Impact of confidence intervals and baserate data on clinical interpretation of change scores.* Paper presented at the 99th Annual Convention of the American Psychological Association, San Francisco, CA.

Cooney, T. M., Schaie, K. W., & Willis, S. L. (1988). The relationship between prior functioning on personality and cognitive dimensions and subject attrition in longitudinal research. *Journal of Gerontology, 43,* P12–P17.

Crook, T. H., Bartus, R. T., Ferris, S. H., Whitehouse, P., Cohen, G. D., & Gershon, S. (1986). Age-associated memory impairment: Proposed diagnostic criteria and measures of clinical change—Report of a National Institute of Mental Health workgroup. *Developmental Neuropsychology, 2,* 261–276.

Crook, T. H., Tinklenberg, J., Yesavage, J., Petrie, W., Nunzi, M. G., & Massari, D. C. (1991). Effects of phosphatidylserine in age-associated memory impairment. *Neurology, 41,* 644–649.

Crook, T. H., & West, R. L. (1990). Name recall performance across the adult life-span. *British Journal of Psychology, 81,* 335–349.

Crook, T. H., Youngjohn, J. R., & Larrabee, G. J. (1990). The Misplaced Objects Test: A measure of everyday visual memory. *Journal of Clinical and Experimental Neuropsychology, 12,* 819–833.

Crook, T. H., Youngjohn, J. R., & Larrabee, G. J. (1992). Multiple equivalent forms in a computerized, everyday memory battery. *Archives of Clinical Neuropsychology, 7,* 221–232.

Flicker, C., Ferris, S. H., Crook, T., & Bartus, R. T. (1987). A visual recognition memory test for the assessment of cognitive function in aging and dementia. *Experimental Aging Research, 13,* 127–132.

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

Fozard, J. L. (1985). Psychology of aging-normal and pathological age differences in memory. In J. C. Brocklehurst (Ed.), *Textbook of geriatric medicine and gerontology* (pp. 122–144). Edinburgh, Scotland: Churchill/Livingstone.

Goldstein, G. (1991, August). *Practice effect phenomena in a national hypertension study.* Paper presented at the 99th Annual Convention of the American Psychological Association, San Francisco, CA.

Hertzog, C., & Schaie, K. W. (1988). Stability and change in adult intelligence: Simultaneous analysis of longitudinal means and covariance structures. *Psychology and Aging, 3,* 122–130.

Kral, V. A. (1962). Senescent forgetfulness: Benign and malignant. *Journal of the Canadian Medical Association, 86,* 257–260.

Kral, V. A. (1966). Memory loss in the aged. *Diseases of the Nervous System, 27,* 51–54.

Lane, F., & Snowdon, J. (1989). Memory and dementia: A longitudinal survey of suburban elderly. In P. Lovibond & P. Wilson (Eds.), *Clinical and abnormal psychology* (pp. 365–376). New York: Elsevier Science.

Larrabee, G. J., & Crook, T. H. (1989). Dimensions of everyday memory in age-associated memory impairment. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 92–97.

Larrabee, G. J., Levin, H. S., & High, W. M. (1986). Senescent forgetfulness: A quantitative study. *Developmental Neuropsychology, 2,* 373–385.

Mitrushina, M., & Satz, P. (1991). Changes in cognitive functioning associated with normal aging. *Archives of Clinical Neuropsychology, 6,* 49–60.

Poon, L. W. (1985). Differences in human memory with aging: Nature, causes, and clinical implications. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 427–462). New York: Van Nostrand Reinhold.

Siepler, I. C., McCarty, S. M., & Logue, P. E. (1982). Wechsler Memory Scale scores, selective attrition, and distance from death. *Journal of Gerontology, 37,* 176–181.

Smith, G. E., Ivnik, R. J., Petersen, R. C., Malec, J. F., Kokmen, E., & Tangalos, E. (1991). Age-associated memory impairment diagnoses: Problems of reliability, concerns for terminology. *Psychology and Aging, 6,* 551–558.

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

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale* [manual]. New York: Psychological Corporation.

West, R. L., & Crook, T. H. (1990). Age differences in everyday memory: Laboratory analogues of telephone number recall. *Psychology and Aging, 5,* 520–529.

Yesavage, J., Brink, T., Rose, T., Lum, O., Huang, O., Adey, V., & Leirer, V. (1983). Development and validation of a geriatric depression scale: A preliminary report. *Journal of Psychiatric Research, 17,* 37–49.

Youngjohn, J. R., & Crook, T. H. (in press). Learning, forgetting, and retrieval of everyday material across the adult life span. *Journal of Clinical and Experimental Neuropsychology.*

Youngjohn, J. R., Larrabee, G. J., & Crook, T. H. (1991). First–Last Names and the Grocery List Selective Reminding Tests: Two computerized measures of everyday verbal learning. *Archives of Clinical Neuropsychology, 6,* 287–300.

Youngjohn, J. R., Larrabee, G. J., & Crook, T. H. (1992a). Discriminating age-associated memory impairment and Alzheimer's disease. *Psychological Assessment, 4,* 54–59.

Youngjohn, J. R., Larrabee, G. J., & Crook, T. H. (1992b). Test–retest reliability of computerized, everyday memory measures and traditional memory tests. *The Clinical Neuropsychologist, 6,* 276–286.

Received August 24, 1992
Revision received December 14, 1992
Accepted December 15, 1992 ■