Journal of Alzheimer's Disease 24 (2011) 125–136
DOI 10.3233/JAD-2010-100450
IOS Press

Differential Cognitive Deterioration in Dementia: A Two Year Longitudinal Study

Samuel R. Chamberlain[a,b,*], Andrew D. Blackwell[a,b], Pradeep J. Nathan[a,b,c], Geoff Hammond[a], Trevor W. Robbins[a,b], John R. Hodges[d], Albert Michael[a], James M. Semple[a], Edward T. Bullmore[a,b,c,1] and Barbara J. Sahakian[a,b,1]
[a]*Department of Psychiatry, University of Cambridge School of Clinical Medicine, Addenbrooke's Hospital, Cambridge, UK*
[b]*Behavioural and Clinical Neuroscience Institute (BCNI), University of Cambridge, Cambridge, UK*
[c]*GlaxoSmithKline, Clinical Unit Cambridge, Addenbrooke's Centre for Clinical Investigations, Cambridge, UK*
[d]*The Prince of Wales Medical Research Institute, Prince of Wales Hospital, Sydney, Australia*

Handling Associate Editor: Paulo Caramelli

Accepted 15 November 2010

Abstract. The ability to predict cognitive deterioration in patients with dementia holds valuable potential for clinical trials and early intervention. This study identified cognitive domains deteriorating differentially over time as well as baseline predictors of subsequent cognitive decline in patients referred to a memory clinic. Twenty-six subjects with Alzheimer's disease (AD) and 43 subjects with Subjective Memory Impairment (SMI) were entered into a longitudinal study in which cognitive function was assessed at baseline and at 8-monthly intervals for 2 years, using a range of well-validated measures. Thirty-seven patients with depression and 39 healthy controls were also longitudinally assessed. AD was associated with disproportionate deterioration over time on general measures of cognitive function, multiple measures of mnemonic processing, mental fluency (letter and category), and aspects of motor speed. SMI showed restricted relative cognitive deterioration on general measures of cognitive function, on a subset of memory measures, and on letter but not category fluency. Secondary analysis showed that earliest detectable ADAS-cog and MMSE decline in AD was at 16 months, while several specific neuropsychological indices were sensitive as early as 8 months (graded naming test, semantic naming, and the category/letter fluency tests). In combination, baseline/early changes in cognitive performance, alongside clinical measures, predicted 48% of disease progression over two years in memory impaired patients as a whole. These findings have implications for identifying patients likely to benefit from disease modifying agents, and for designing, powering, enriching, and implementing future clinical trials. Follow-up studies in independent populations are needed to validate predictive algorithms identified.

Keywords: Alzheimer's disease, cognition, dementia, enrichment, longitudinal, prediction, semantic

Supplementary data available online: http://www.j-alz.com/issues/24/vol24-1.html#supplementarydata01

INTRODUCTION

Alzheimer's disease (AD) is a chronic degenerative disorder affecting almost 26.6 million people today, and whose prevalence is expected to quadruple by 2050 [1]. Dementia is prevalent, particularly with increasing age [2], and is thought to contribute to 11.2% of years lived with disability by people over the age of 60 years [3]. In the UK, it is estimated that approximately 43% of prevalent cases require a high level of care, with costs amounting to approximately £5.2 billion (US$8.4billion) per annum [4, 5].

[1]Joint senior authors.

*Correspondence to: Samuel R. Chamberlain, Tel.: +44 1223 767040; Fax: +44 1223 336968; E-mail: srchamb@cam.ac.uk.

ISSN 1387-2877/11/$27.50 © 2011 – IOS Press and the authors. All rights reserved

The identification of biomarkers capable of detecting AD in the early stages, and of predicting the rate and nature of cognitive deterioration, would be of considerable social and economic importance with the prospect of effective treatments that retard the progression of, or ameliorate, this devastating disease [6]. For example, it is estimated that if interventions could delay both the disease onset and progression by even 1 year, there would be nearly 9.2 million fewer cases of AD in 2050 [1]. Although a definite diagnosis of AD has to be based on neuropathological features, its early detection and subsequent assessment depends on a combination of clinical, pathophysiological (i.e., amyloid PET scanning and cerebrospinal fluid levels of amyloid-β and tau), and neuropsychological criteria [6]. Objective biomarkers, such as neuropsychological measures, would be valuable in drug development and in clinical trials, for enabling recruitment of 'enriched' samples, and for determining appropriate sample sizes to adequately power novel studies [7].

Accepted criteria for a diagnosis of probable AD requires deficits in two or more areas of cognition, one of which must be memory, in addition to other criteria based on neuropsychological tests [8]. Neuropsychological tests, such as those in the Cambridge Neuropsychological Test Automated Battery (CANTAB), have been validated in patients with well-defined neurosurgical cortical excisions. The main elements of CANTAB have already been tested in probable (subsequently confirmed) cases of AD in a memory clinic setting, including longitudinal follow-up of some patients [9–12] and in a double blind, placebo controlled trial of the acetylcholinesterase inhibitor Tacrine [13]. The pattern of cognitive deficits shown on CANTAB fits quite well with the generally accepted pattern of neural progression for the disease. Thus, tests particularly sensitive to temporal lobe (including hippocampal) damage are, generally speaking, more sensitive to impairments in AD than tests shown to be sensitive to frontal lobe dysfunction [9, 10]. Normative data are available for many of these tests from a large ($n \approx 800$) sample of healthy community dwelling elderly volunteers and acceptable to good test-retest reliability has been confirmed.

Though AD and its prodrome have been linked extensively to neurocognitive dysfunction, less is known regarding whether specific domains deteriorate over time, and whether there is differential deterioration in people with these disorders, as compared to the healthy population or to patients with depression (which has itself been linked to cognitive problems [14]) and potentially deleterious effects of repeated mood episodes on neuropsychological function (particularly relating to hippocampal status) [15]. Depression often co-presents with dementia, making clinical and neuropsychological disentanglement complex in some cases [16].

The current paper reports results from a two year longitudinal study conducted in patients with AD, patients with subjective memory problems not meeting AD criteria (i.e., Subjective Memory Impairment, SMI), patients with depression, and healthy controls. The SMI group was recruited on the basis of self-reported problems with memory rather than objective memory deficits to provide a naturalistic sample for comparison purposes. As such, it was felt that findings with respect to any differential cognitive decline over time in this group would be highly clinical relevant, including in terms of proof-of-concept clinical trials. It should be noted that our definition of SMI differs from that of mild cognitive impairment (MCI). SMI refers to non-demented patients with memory complaint; MCI refers more selectively to non-demented patients with memory complaint, as well as normal activities of daily living, normal general cognitive function, and abnormal memory for age [17].

In previous reports based on this sample, it was shown that baseline performance on a Paired Associates Learning (PAL) memory test distinguished patients with AD from patients with depression, and from healthy control subjects [11]. Moreover, the entry variables were capable of predicting development of formal dementia in recruited SMI subjects over two years [18]. Thus, risk factors associated with AD development were older age at study entry, worse baseline performance on a measure from the PAL test, and worse baseline performance on the Graded Naming Test. The current paper presents the full longitudinal data from this study, and uses growth curve modeling to examine for differential deterioration of specific cognitive measures over time in clinical groups of interest compared to healthy volunteers. We also identify composite measures capable of optimally predicting overall neurocognitive decline over the study period.

MATERIALS AND METHODS

This study was conducted at the Memory Clinic, Addenbrooke's Hospital, Cambridge, United Kingdom. Participants were aged <80 years and had no history of cerebrovascular events (including transient ischemic attacks or strokes), serious head injury which required surgical intervention, and did not

have vascular dementia, epilepsy, uncontrolled diabetes, schizophrenia, bipolar disorder or depression, extrapyramidal signs, hallucinations, active treatment for cancer, recent chronic treatment with benzodiazepines/neuroleptics/anticonvulsants or benzodiazepines at the time of testing. Verbal IQ was assessed using the National Adult Reading Test (NART) [19], and depressive mood with the Geriatric Depression Scale [20].

Patients with mild AD ($n = 26$), and patients with subjective memory complaints (SMI, $n = 43$), entered the study after diagnosis and preliminary clinical and neuropsychological screening. Together, the AD and SMI subjects comprised a sample of subjects presenting to memory clinic. AD was diagnosed according to the NINCDS-ADRDA criteria and tests based on them, including the Hachinski scale, Dementia Rating Scale, Judgment of Line Orientation, Complex Figure Test, Object Matching (Unusual Views) and digit span [8]. AD patients also received SPECT or CT scans which were used together with other clinical features to define possible exclusion criteria arising from non-AD pathology (e.g., lobar atrophy, multi-infarct dementia, Lewy body dementia) group. In the AD group, only mild AD patients not fulfilling DSM-IV criteria for depression were enrolled. In the SMI group, patients with subjective memory complaints who did not fulfil the NINCDS-ADRDA criteria for AD, and who were free from DSM-IV depressive disorders, were enrolled.

Thirty-seven subjects meeting DSM-IV criteria for Major Depressive Disorder and with MMSE scores >24 were recruited from psychiatric departments in Bury St. Edmunds and Cambridge ($n = 37$). In order to provide comparator data and to identify cognitive functions declining over time in normalcy, $n = 39$ healthy controls free from DSM-IV diagnoses and subjective memory complaints (with MMSE >24), were recruited from amongst spouses and friends of patients, and through an advertisement placed in the Alzheimer's Disease Society newsletter.

Neuropsychological assessment

Recruits were tested on a range of pen/paper and computerized tests at baseline and then at 8-monthly (±30 day) intervals for two years. Neuropsychological tests are listed separately below along with a brief description of each task and salient output variables. Please refer to the citations for details of validation and descriptions of the tasks.

Pen/paper tests

Mini Mental State Examination (MMSE)

This is a screening test for dementia. It gives a total score of 0–30 [21].

ADAS-cog

A scale for assessing cognitive symptoms in AD. The scores (maximum = 70) represent the number of errors [22].

Wechsler Logical Memory

This test examines free recall of 2 story passages after a 30 minute delay [23]. The maximum score, representing total items recalled is 50. Higher scores represent superior performance.

Baddeley doors test (doors recognition)

This is a test of 4-choice recognition of photographs of doors (2 lists of 12). The maximum score is 24, with higher scores representing better performance [24].

Graded Naming Test (GNT)

This test assesses object-naming ability [25]. Thirty different line drawings are displayed, one at a time. The subject must identify (i.e., name) the object depicted in each drawing. The scores represent the number of items correctly named (maximum = 30).

Semantic Naming Test

Subjects name line drawings, from 8 semantic categories (e.g., vehicles – sledge, train...) (maximum = 64) [26].

Warrington Short Recognition Memory Test (SRMT)

This is a 2-choice recognition test of visually presented words and photographs of male faces (list of 25 for each) [27]. The maximum scores are 25 for each sub-test (words, faces).

Category and letter fluency tests

The category fluency test examines the ability to spontaneously generate names of items from 3 categories, namely animals, fruit, and household items, each in 60 seconds [28, 29]. The total score represents total items named. Higher scores indicate superior performance. The letter fluency test is similar except that subjects must name items beginning with the letters F, A, and S.

Computerized tests

Delayed-Matching-To-Sample (DMS)

A 4-choice recognition test of abstract patterns sharing color or pattern with distracters (10 trials at each time delay). The key outcome variable is total correct out of a maximum of 10 at the hardest 12-second delay [9, 30].

Rapid Visual Information Processing (RVP)

RVP is a test of continuous performance and visual sustained attention [31]. A white box appears in the center of the computer screen, and digits from 2 to 9 appear in a pseudo-random order within the box at a rate of 100 digits per minute. Subjects are requested to detect target sequences of digits (i.e., 2-4-6, 3-5-7, 4-6-8) and to register responses using the press pad. The key outcome variables are RVP A′ (representing the ability to detect targets, range 0–1, with 1 representing perfect detection) and mean response latency.

Visuospatial Paired Associates Learning (PAL)

This test assesses visual memory and learning [9]. Boxes are displayed on the screen and are opened in a randomized order. One or more boxes contain a pattern. The patterns are then displayed in the middle of the screen, one at a time, and the subject must touch the box where the pattern was originally located. If the subject makes an error, the patterns are re-presented to remind the subject of their locations. The key outcome measure is the total number of errors made, which is corrected to take into account stages that were failed /not attempted. Secondary measures are the number of stages passed (maximum 8, with higher scores representing superior performance), and total numbers of errors made on the 6-level (hard) difficulty stage.

Intra/Extra Dimensional set shift (IED)

IED is a test of the ability to acquire an attentional set and show flexibility following negative feedback [32]. Two stimuli are presented on-screen on each trial, and the subject has to learn the relevant stimulus dimension. The key outcome measure is the number of stages successfully completed, out of a maximum of 9.

One Touch Stockings of Cambridge test (OTS)

This test examines the ability to plan a sequence of ball movements on-screen in order to obtain a goal arrangement predetermined by the computer [33]. Executive planning is assessed in terms of the number of attempts taken to select the correct number of moves needed at the 5-level difficult of the task. Higher scores indicate worse performance.

Choice Reaction Time (5-choice reaction time, CRT)

The CRT assesses speed of response to visual stimuli appearing on-screen [13], in one of five locations. The two outcome variables are accuracy (total hits, maximum 25) and mean response latency.

Pattern Recognition Memory (PRM)

This is a test of visual pattern recognition memory in a 2-choice forced discrimination paradigm [9]. The subject is presented with two series of 12 visual patterns, presented one at a time. In the recognition phase, the subject is required to choose between a pattern they have already seen and a novel pattern. The key outcome measure is the percentage of correct patterns chosen. The key outcome measure is the percentage of patterns correctly chosen.

Spatial Recognition Memory (SRM)

This is a 2-choice recognition test of locations of white boxes on a computer screen (4 lists of 5) [9]. The key outcome variable is the percentage of correct choices, with higher scores reflecting superior performance.

Statistical methodology

Data were analyzed using the Statistical Package for the Social Sciences (SPSS) version 15. Differences in baseline characteristics between the groups were assessed by one-way analyses of variance (ANOVAs) with follow-up Least Significant Difference (LSD) tests as appropriate. Changes in cognition over time in the four groups of interest were explored by means of random effects regression, including a group by time interaction effect. In order to compare the rate of change across conditions over the 24 months, a random effect linear model was implemented using the MIXED procedure within SPSS. A linear growth model with a random effect of individual and slope was developed for each of the outcomes, using the group condition as a baseline factor in addition to a predictor of slope. Multiple comparisons were minimized by focusing only on primary measures from each task of interest. Slope coefficients (group by time) interactions were interpreted and compared across conditions to assess differential rates of change. Advantages of this approach are that it allows for the inclusion of

individuals with partially missing data and that it is natively suited for the assessment of slope as opposed to mean change (as in an ANOVA). We also conducted a secondary analysis to compare whether performance deteriorated between baseline and follow-up time points in each separate group, using within-subject *t*-tests (these results are reported in the Supplementary data, available online: http://www.j-alz.com/issues/24/vol24-1.html#supplementarydata01). Significance was defined as $p < 0.05$ uncorrected. Where data were missing from the raw datasheets, no imputation for missing data was made in SPSS unless otherwise explicitly indicated – therefore, the current analysis should be considered 'per protocol'.

Multiple regression analyses were then conducted in order to identify those combinations of cognitive, clinical, and demographic variables that optimally predicted ADAS-cog deterioration over 24 months (including early change scores, i.e., the calculated difference in performance from baseline to the time point in question). Since the aim was to identify 'best possible' predictive algorithms, both primary and secondary measures from cognitive tasks were included in this analysis. Analyses were conducted first in the combined AD-SMI group and then in the AD and SMI groups separately, using a step-wise approach.

Finally, a supplementary principal components analysis was used to explore the factor structure of baseline variables. Underlying factors were identified by means of maximal 25 iterations, with varimax rotation. The number of factors for a given model was decided on the basis of screen plot inspection. Factor analysis was deemed valid where the subject-to-variable ratio was 3 : 1 or greater, and where the variable-to-factor ratio was 3 or greater [34]. Minimal sample size requirements necessitated that the principal components analysis was restricted to the AD-SMI combined group. Measures loading heavily on the identified factors were defined as exhibiting a relationship of $r \geq 0.70$.

RESULTS

Baseline characteristics of the sample

As can be seen in Table 1, the study groups exhibited mean ages in the order of 60–70 years. Depressed patients were significantly younger than the three other groups ($p < 0.01$), which did not differ from each other for age (all $p > 0.05$). As expected, the depressed group showed significantly higher Geriatric Depression Scale (GDS) scores than the three other groups (all $p < 0.01$). The AD and SMI groups also showed significantly higher GDS scores than healthy volunteers (both $p < 0.01$) but the mean scores were beneath threshold for clinically significant depression (mild depression 10+; severe depression 20+). The groups did not differ significantly on gender ratios [Chi-square ($df = 3$) = 1.746, $p > 0.10$]. Of the AD patients, 4 were receiving Donepezil at study entry, and 2 Rivastigmine. Of the SMI patients, 1 was receiving Donepezil at study entry. Eleven SMI patients (26%) converted to AD during the study period.

Table 1
Baseline demographic characteristics (mean, SD) of each study group (AD = Alzheimer's disease, SMI = Subjective Memory Impairment, Depressed patients, Controls)

	AD	SMI	Depression	Controls
N	26	43	37	39
Age (yrs)	68.5 (8.2)	64.9 (9.1)	60.4 (8.3)	64.5 (8.7)
M : F	09 : 17	20 : 23	17 : 20	14 : 25
MMSE	21 (3)	27.8 (2.4)	28.2 (1.6)	29.2 (1.2)
ADAS-cog	22.2 (5.8)	11.2 (5.9)	9.3 (3.9)	6.7 (2.4)
NART-IQ	108.4 (10.1)	118.4 (7.2)	111.2 (9.5)	119.2 (7.7)
GDS	7.5 (4.7)	9.1 (5.4)	22 (5.7)	4.1 (2.7)

MMSE – Mini Mental State Examination; ADAS-cog – Alzheimer's Disease Assessment Scale, cognitive; NART-IQ – National Adult Reading Test IQ; GDS – Geriatric Depression Scale.

In terms of broad intellectual functioning, mean IQ scores were all well above 100 for all the groups, with AD patients exhibiting significantly lower IQs than the SMI and healthy volunteer groups (both $p < 0.01$). The depressed patients showed IQs comparable to the AD group ($p > 0.10$), i.e., lower IQ scores than the SMI and healthy control groups (both $p < 0.01$). On MMSE, AD patients had lower scores than all three other groups (all $p < 0.01$) as expected. SMI patients had lower MMSE scores than healthy volunteers ($p < 0.01$) but did not differ from the depressed patients ($p > 0.10$). With respect to baseline ADAS-cog, AD patients showed higher scores than all three other groups (all $p < 0.01$); and the SMI and depressed groups showed higher scores than the healthy volunteers (both $p < 0.05$).

Growth curve modeling

Figures 1 and 2 indicate mean performances for each group for pen/paper and computerized tests respectively (see Supplementary Table 1 and Supplementary Figures 1 and 2 for full data, including subgroups of SMI who did and did not convert]. Results of the growth curve analysis are presented in Table 2, which indicates mean trajectory differences between each clinical group and the healthy controls (left columns)

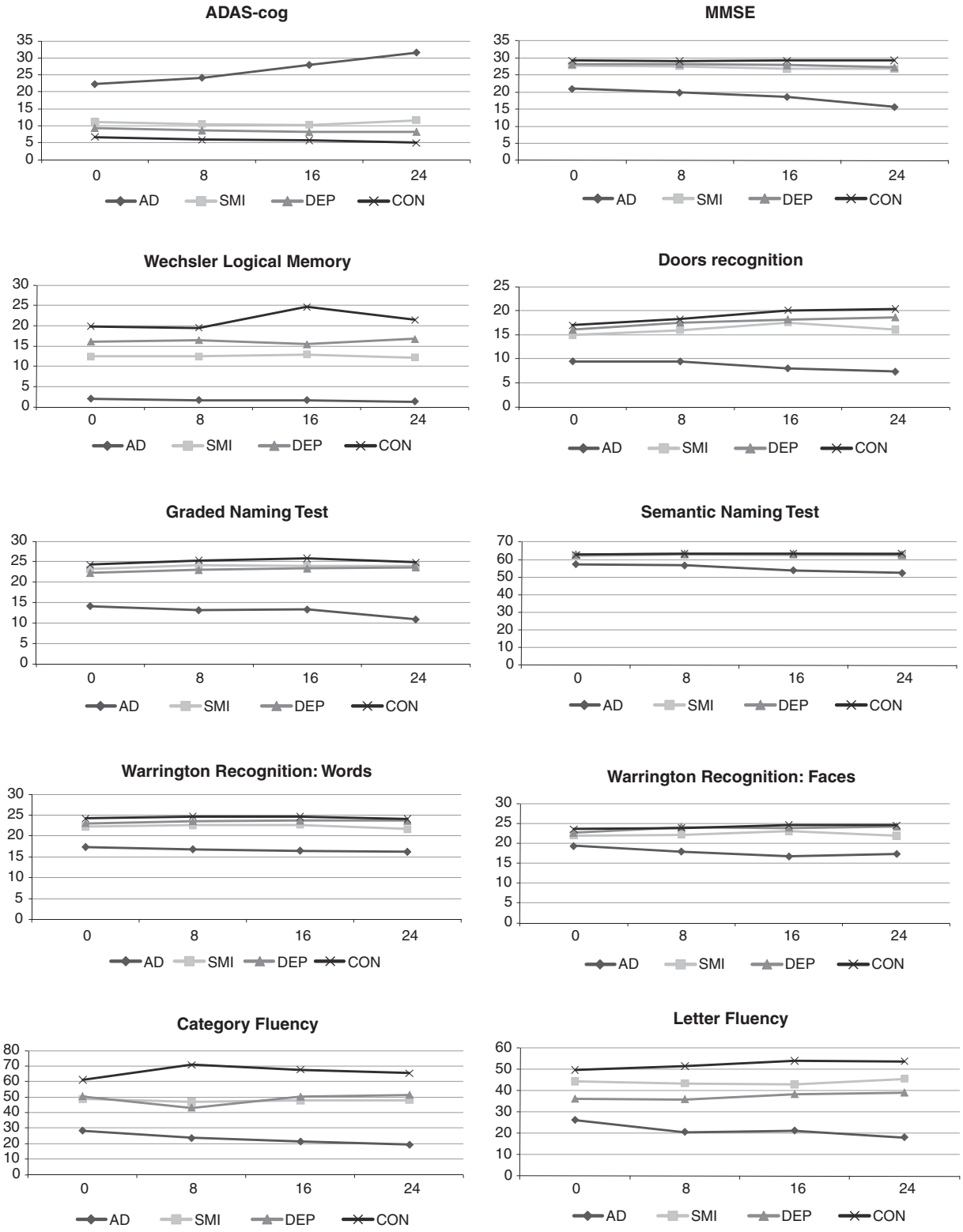

Fig. 1. Performance on pen/paper measures over time in each group. X-axis corresponds to months since baseline, y-axis to performance measure.

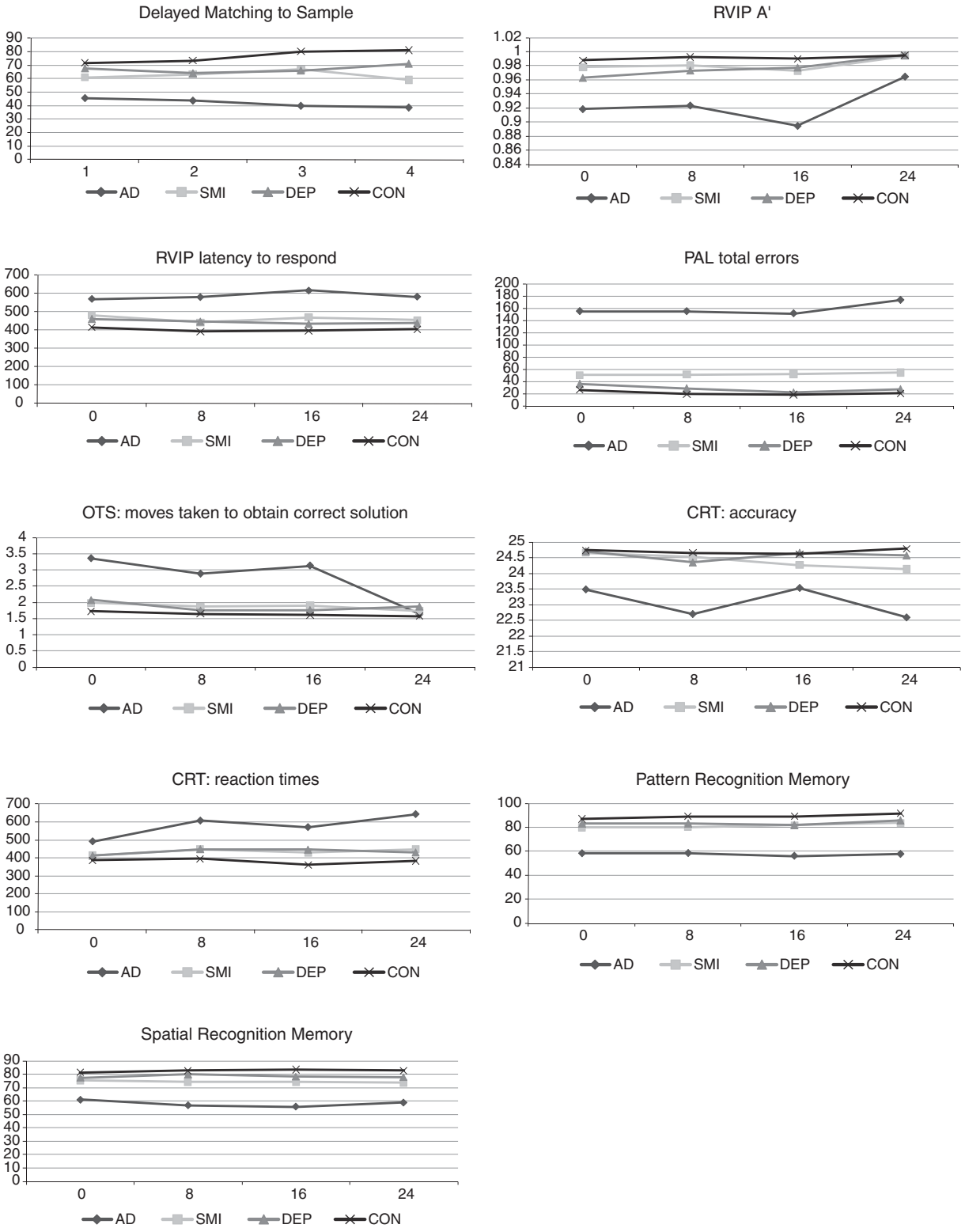

Fig. 2. Performance on computerised measures over time in each group. X-axis corresponds to months since baseline, y-axis to performance measure.

Table 2
Summary of growth curve analysis

Task	Deviation in slope (condition by time) coefficient (Est (SE))*				Significant comparisons of slope (numbers refer to groups; plain font $p<0.05$, bold font $p<0.01$)
	Alzheimer's	SMI	Depressed	Controls	
Global cognitive function					
ADAS-cog	3.92 (0.42)	0.88 (0.34)	0.17 (0.35)	–	1v2, 1v3, 1v4, 2v3, 2v4
MMSE	−1.84 (0.26)	−0.48 (0.21)	−0.32 (0.22)	–	1v2, 1v3, 1v4, 2v4
Recognition					
Warrington SRMT words	−0.35 (0.28)	−0.19 (0.23)	0.26 (0.24)	–	No sig. differences
Warrington SRMT faces	−1.13 (0.28)	−0.34 (0.23)	0.14 (0.24)	–	1v2, 1v3, 1v4, 2v3
PRM	−1.82 (0.97)	−0.45 (0.79)	−0.68 (0.83)	–	No sig. differences
SRM	−0.95 (1.10)	−1.18 (0.91)	−0.54 (0.95)	–	No sig. differences
Doors recognition	−1.93 (0.31)	−0.78 (0.25)	−0.36 (0.26)	–	1v2, 1v3, 1v4, 2v4
DMS (correct, 12 s delay)	−6.16 (1.93)	−4.12 (1.56)	−2.30 (1.66)	–	1v4, 2v4
Cued/free recall					
Logical memory (30 min)	−1.20 (0.47)	−1.29 (0.39)	−0.92 (0.41)	–	1v4, 2v4, 3v4
PAL (stages passed)	−0.10 (0.08)	−0.05 (0.07)	0.16 (0.07)	–	1v3, 2v3, 3v4
PAL (6-pattern errors)	1.04 (0.90)	0.24 (0.74)	0.83 (0.78)	–	1v3
PAL (total errors)	6.24 (1.95)	4.45 (1.59)	−1.60 (1.68)	–	1v2, 1v3, 1v4, 2v3, 2v4
Semantic naming/fluency					
Graded naming test	−1.06 (0.28)	−0.01 (0.23)	0.29 (0.24)	–	1v2, 1v3, 1v4
Semantic naming	−2.40 (0.28)	−0.14 (0.23)	−0.28 (0.24)	–	1v2, 1v3, 1v4
Category fluency	−4.42 (1.98)	−1.39 (1.63)	−0.40 (1.71)	–	1v3, 1v4
Divided attention					
5-choice RTI (accuracy)	−0.27 (0.16)	−0.19 (0.13)	−0.02 (0.13)	–	No sig. differences
5-choice RTI (latency, m sec)	−0.25 (0.16)	−0.17 (0.13)	0.02 (0.13)	–	#
Sustained attention					
RVP (A′)	0.002 (0.004)	0.002 (0.003)	0.008 (0.003)	–	2v3, 3v4
RVP (latency, m sec)	30.6 (12.5)	−1.9 (9.9)	−5.6 (10.4)	–	1v2, 1v3, 1v4
Executive function					
Letter fluency	−4.37 (0.9)	−1.59 (7.6)	−0.64 (0.79)	–	1v2, 1v3, 1v4, 2v4
IED (stages completed)	−0.19 (0.19)	−0.22 (0.16)	0.03 (0.16)	–	#
OTS (attempts, 5 moves)	−0.26 (0.15)	<−0.001	−0.02 (0.07)	–	No sig. differences

Negative residual variance in the model estimates, indicating possible model over fit.

and significant pairwise differential rates of change between groups (right column). Findings are discussed in more detail below.

Pen/paper tests

Slope of change differed significantly between AD and all other groups on the ADAS-cog and MMSE. This was attributable to deterioration in AD versus all other groups over time. Patients with SMI exhibited significantly different gradients on both measures compared to healthy volunteers over time, again due to relative deterioration over time in patients. Depressed patients did not differ significantly from healthy volunteers on ADAS-cog and MMSE rates of change. SMI differed on slope for ADAS-cog but not MMSE versus depressed subjects.

On the Wechsler Logical Memory test, gradient of change differed significantly from healthy volunteers for AD, SMI, and depressed patients. The three clinical groups did not show a differential rate of change compared to each other. While the healthy controls numerically improved on logical memory over time, the other groups tended to remain static or to deteriorate. For Baddeley doors recognition, AD cases differed significantly from all other groups on slope of change over time. SMI differed significantly from healthy controls on this measure, while depressed patients did not differ significantly from healthy participants. While door recognition numerically decreased over time in AD, other groups showed increased scores, but to a lesser degree in SMI compared to depressed patients and controls. For the GNT, AD cases showed significantly different slope of change versus all other groups, due to deterioration over time (other groups were relatively stable or improved numerically). The same pattern of results was found on the Semantic Naming test. For the Warrington Short Recognition Memory Test, an interesting dissociation was found between performance on word and face stimuli. For word stimuli, there were no differential rates of change over time between any groups. For face stimuli, AD cases differed significantly from all other groups on gradient of change. SMI differed significantly from depressed cases but not from controls. While word recognition was relatively stable over time in all groups, face recognition numerically declined over time in the AD patients and to a lesser degree in the SMI

patients, while depressed patients and controls were stable or numerically increased their performance over time. On category fluency, AD differed significantly from healthy controls and depressed patients on rates of change. SMI and depressed patients did not differ from healthy controls on the slope of change. While category fluency reduced over time in AD, performance was relatively stable in the other groups overall. For letter fluency, AD differed significantly on slope versus all other groups. SMI patients differed significantly from healthy controls but not from depressed patients. Depressed cases did not differ significantly from the healthy controls on this measure. Patients with AD deteriorated over time while other groups were stable or improved performance somewhat.

Computerized tests

On the DMS test measure, slopes of change in AD and SMI groups differed significantly from the healthy volunteers. Depressed patients did not differ significantly from healthy recruits on this measure. AD and SMI declined overall, numerically, while depressed patients and controls were stable or improved. For the RVP test of sustained attention A′ measure, slope of change differed significantly between SMI and depressed groups; and between depressed and healthy control groups. This was due to relatively steeper improvement over time in depressed cases versus SMI and healthy control groups. For RVIP latency to respond, slope of change differed significantly between AD and all other groups, who did not differ from each other on this measure. AD cases showed steeper lengthening of reaction times over time than the other groups. On PAL total errors, patients with AD differed significantly from all other groups on rate of change. Patients with SMI differed significantly from depressed patients and healthy recruits, who did not differ significantly from each other on this measure. Findings were attributable to increased errors over time in AD and SMI with relative stability or reductions in errors over time in the depressed patients and in the controls. On the OTS test of executive planning, rates of change in performance over time did not differ significantly between any of the groups; it can be seen graphically nonetheless that AD patients initially required numerically more moves to obtain correct solutions, and that their performance normalized to that of the other groups over time. Other groups were likely at ceiling. No differential change over time was found for CRT accuracy, PRM, and SRM performance. On the latency measure from CRT, there was negative residual variance in the model, i.e., group comparisons on slope were not reported. Descriptively, AD patients were increasingly numerically slower on the task over time, while other groups were relatively stable.

Baseline predictors of subsequent ADAS-cog deterioration

Combined AD-SMI group

A model was identified that accounted for approximately 50% of the variance in two year ADAS-cog deterioration in the AD-SMI group [$F(3,33) = 10.148$, $p < 0.001$; R square = 0.480, adjusted R square = 0.433]. Two year ADAS-cog decline was predicted by: 23.886 − (1.481 × Baseline PAL stages passed) − (1.265 × Baseline IED stages passed) + (0.425 × Early change ADAS-cog score) (see Supplementary data for breakdown of beta coefficients and other measures for each predictive variable).

AD group

No models approached useful predictive significance (the most significant model accounted only for 16% of the variance).

SMI group

A model was identified that accounted for approximately 60% of the variance in two year ADAS-cog deterioration [$F(5,37) = 10.929$, $p < 0.001$; R square = 0.596, adjusted R square = 0.542]. Two-year ADAS-cog deterioration was predicted by: 2.630 + (3.221 × sex [1 = male, 2 = female]) + (0.182 × PAL 6-stage errors early change score) + (0.679 × ADAS-cog early change score) − (0.087 × Baseline spatial recognition memory) − (75.955 × Early change RVP A′ score).

DISCUSSION

This study evaluated cognitive function over two years in AD, SMI, depressed patients and healthy controls, using a comprehensive neuropsychological battery of tests. We identified selective differential rates of change in performance between study groups using growth curve modeling on a subset of the cognitive domains assessed. Further, we identified a combination of clinical and cognitive measures (at baseline, and at 8-months) that successfully predicted approximately 50% of the variance in two year ADAS-cog deterioration in patients referred to memory clinic (combined AD-SMI sample); and 60% of the variance when the SMI patients were considered alone.

Differential and selective cognitive change over time

The selection of appropriate tests and sample sizes for interventional trials in patients with AD and prodromal memory problems necessitates knowledge of cognitive measures that are differentially sensitive to the deterioration over time and disease progression. The findings from this study can be used to formulate power calculations in such future interventional studies, and to determine whether change in an individual's performance is outside what would ordinarily be expected. In this study, selective deterioration across multiple cognitive domains occurred in AD and in SMI groups relative to healthy volunteers, and in most cases relative to people with depression too. This is important since it is conceivable that some of the cognitive deterioration reported in previous AD literature could have been attributable to deleterious effects of depressive episodes on brain function rather than effects of the pathophysiology of AD itself.

AD was characterized by disproportionate deterioration over time on general measures of cognitive function (ADAS-cog, and MMSE), multiple measures of mnemonic processing (Warrington face recognition, doors recognition, delayed matching to sample, Logical memory, PAL, GNT, semantic naming), and aspects of motor speed (RVIP latency to respond). Interestingly, disproportionate deterioration over time was not seen for more abstract pattern recognition memory or spatial recognition memory, suggesting that difficulties may be restricted to aspects of memory that are challenging (e.g., PAL), or related to the types of stimuli likely to be experienced in day-to-day life. Aspects of mental fluency also deteriorated disproportionately in AD (category and letter fluency), while executive function (OTS and RVIP sustained attention) did not deteriorate disproportionately versus controls.

SMI cases showed a more restricted pattern of relative cognitive deterioration versus healthy volunteers over time than AD. Specifically, they showed disproportionate deterioration on general measures of cognitive function (ADAS-cog, MMSE), on a subset of memory measures (doors recognition, delayed matching to sample, logical memory), and on letter but not category fluency.

In regard to measures deteriorating earliest, secondary analysis (see Supplement online Figure 3) indicated that the earliest detectable ADAS-cog and MMSE decline in AD (in terms of significant change compared to group baseline) was at 16 months while several specific neuropsychological indices were sensitive to decline as early as 8 months. Tests found to be sensitive to decline at 8-months were the graded naming test (GNT), semantic naming, and the category/letter fluency tests. These findings are preliminary given that this was a secondary analysis.

In a longitudinal study assessing cognition every 6-months over two years, profound deterioration on PAL was evident from 6-months onwards in the group of participants with Questionable Dementia (QD) [35]. In QD recruits with PAL deterioration, all converted to probable dementia by the study end point. In our study, we found some deterioration on PAL measures over time in SMI participants who did convert (mean increase of approximately 30 PAL total-errors over two years), but to a numerically lesser degree than that seen in this prior work (mean increase of approximately 100 PAL total-errors in the prior study, over a similar time frame). This discrepancy could stem from differences in methodologies (e.g., recruitment and inclusion criteria) between the two studies.

It is important to consider several potential limitations. Data were analyzed per protocol, so attrition of cases may have led to an underestimation of true cognitive decline occurring in the clinical samples. Dropout rates were 10–30% in the current study, highest in the AD group. In a previous two year observational cohort study of patients newly referred to dementia clinics, dropout was 55% overall [36]. In a two year cohort study of patients with AD, dropout was ∼40% [37]. Clinical trials in MCI have reported all cause dropout rates of ∼30–40% across similar timeframes [38]. Therefore, dropout rates reported here compare favorably with other studies. We did not assess specific reasons for dropout. Previous data suggest important factors in the context of AD include refusal, death, institutionalization, and loss to follow-up [37]. Dropout may have introduced bias, for example, those recruits most likely to show cognitive deterioration may have been the most likely to drop out from the study.

Another potential limitation is that the conversion rate in SMI was relatively low (26%), suggesting that some subjects in this group comprised 'worried well', i.e., persons with apparent subjective memory complaints that were not associated with gross underlying pathology. Decline in this group may have been limited by selecting subjects at enrolment with subjective as opposed to objective memory impairment, without corroboration by an informant. This would have diminished the ability to detect deterioration associated with the prodromal stages of disease, in the SMI group as a whole. Consequentially, the SMI group was somewhat

heterogeneous, and these findings may not generalize to other studies using different recruitment methods (e.g., objective rather than subjective memory impairment criteria at entry for those with memory problems not fulfilling full criteria for AD). The inclusion of a depression control group was useful for showing relative stability of cognition over time in this group; however, depressed subjects showed lower IQ than the AD subjects; this may reflect bias attributable to the different recruitment sources used for these two groups. Lastly, six AD patients took cholinesterase inhibitors during the study; the current study was neither designed nor powered to evaluate effects of these medications on cognition.

Predicting two year decline in ADAS-cog

This study sought to identify potential predictors of ADAS-cog deterioration in AD and SMI subjects. The analyses in the AD-SMI group are likely to be more representative of those reporting to memory clinic than assessment of either subgroup alone. Combined group analysis also provides greater power to detect predictors.

The ability to predict the likelihood of AD and SMI subjects showing deterioration on ADAS-cog over time (and/or the magnitude of such deterioration), on the basis of baseline and early (8-month) changes in performance, would be extremely valuable for sample enrichment. Best fit models of 24-month ADAS-cog deterioration were generated by entering all baseline and early change scores into multiple regression analyses. When considering the combined AD-SMI group, 50% of the variance in ADAS-cog deterioration was successfully accounted for by a model comprising baseline performance on PAL (stages passed) and the IED test (stages passed), along with 8-month deterioration scores for ADAS-cog. The best-fit model for the AD group alone was somewhat less efficient, accounting for only 16% of the variance. In the SMI subjects, 60% of the variance in ADAS-cog deterioration was accounted for by a model including gender (females, worse outcome), PAL (6-stage errors), baseline spatial recognition memory, and early change scores for ADAS-cog and RVP A prime. The current findings may not generalize to other subject samples/studies. There are potential limitations of such modeling, e.g., overfit. Given the pilot nature of this study, trials of these algorithms in separate populations will be required in order to validate them. Also, the inclusion of early change scores implies the need to study subjects for a time before formally enrolling them into a clinical trial; this represents a potential pragmatic limitation.

Now that these salient variables have been identified, future interventional studies could hone in on specific baseline/early displayed features of study recruits likely to maximize the ability to show benefits of novel agents. Future studies could seek to validate the formulae in independent memory clinic samples, to see if it is capable of successfully predicting ADAS-cog deterioration on the basis of baseline and early change scores.

Summary

This paper comprises a rich dataset from a two-year longitudinal study in patients with memory dysfunction, along with depressive and healthy control groups. It represents a valuable source of information for guiding the selection of appropriately sensitive and reliable cognitive measures in future AD/SMI interventional studies; for powering of such studies; for defining what constitutes unexpectedly good or poor outcomes in individuals; and, potentially, for enriching samples.

ACKNOWLEDGMENTS

The authors thank Dr E. Marshall and Mrs C. Wightwick for assistance with document preparation and statistics. This study was co-funded by the Medical Research Council and GlaxoSmithKline. Cambridge Cognition markets the CANTAB battery. Dr S.R. Chamberlain consults for Cambridge Cognition, P1Vital, and Shire; and has received honoraria from Cambridge Cognition for work on this project. Dr A.D. Blackwell is an employee of Cambridge Cognition. Prof. E.T. Bullmore and Prof. P.J. Nathan are employees of GlaxoSmithKline. Prof. T.W. Robbins and B.J. Sahakian consult for Cambridge Cognition. The Behavioural and Clinical Neuroscience Institute is jointly funded by the Medical Research Council and Wellcome Trust.

Authors' disclosures available online (http://www.j-alz.com/disclosures/view.php?id=679).

REFERENCES

[1] Brookmeyer R, Johnsona E, Ziegler-Grahamb K, Arrighic HM (2007) Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement* **3**, 186-191.

[2] Plassman BL, Langa KM, Fisher GG, Heeringa SG, Weir DR, Ofstedal MB, Burke JR, Hurd MD, Potter GG, Rodgers WL, Steffens DC, Willis RJ, Wallace RB (2007) Prevalence

of dementia in the United States: the aging, demographics, and memory study. *Neuroepidemiology* **29**, 125-132.

[3] WHO (2003) *World Heath Report: Shaping the Future*, WHO, Geneva.

[4] Comas-Herrera A, Wittenberg R, Pickard I, Knapp M (2005) Cognitive impairment in older people: its implications for future demand for services and costs. *PSSRU Discussion Paper*, 1728.

[5] Comas-Herrera A, Wittenberg R, Pickard L, Knapp M (2007) Cognitive impairment in older people: future demand for long-term care services and the associated costs. *Int J Geriatr Psychiatry* **22**, 1037-1045.

[6] Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, Delacourte A, Frisoni G, Fox NC, Galasko D, Gauthier S, Hampel H, Jicha GA, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Sarazin M, de Souza LC, Stern Y, Visser PJ, Scheltens P (2010) Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol* **9**, 1118-1127.

[7] Frank R, Hargreaves R (2003) Clinical biomarkers in drug discovery and development. *Nat Rev Drug Discov* **2**, 566-580.

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

[9] Sahakian BJ, Morris RG, Evenden JL, Heald A, Levy R, Philpot M, Robbins TW (1988) A comparative study of visuospatial memory and learning in Alzheimer-type dementia and Parkinson's disease. *Brain* **111**(Pt 3), 695-718.

[10] Sahakian BJ, Downes JJ, Eagger S, Evenden JL, Levy R, Philpot MP, Roberts AC, Robbins TW (1990) Sparing of attentional relative to mnemonic function in a subgroup of patients with dementia of the Alzheimer type. *Neuropsychologia* **28**, 1197-1213.

[11] Swainson R, Hodges JR, Galton CJ, Semple J, Michael A, Dunn BD, Iddon JL, Robbins TW, Sahakian BJ (2001) Early detection and differential diagnosis of Alzheimer's disease and depression with neuropsychological tasks. *Dement Geriatr Cogn Disord* **12**, 265-280.

[12] Sahgal A, Galloway PH, McKeith IG, Lloyd S, Cook JH, Ferrier IN, Edwardson JA (1992) Matching-to-sample deficits in patients with senile dementias of the Alzheimer and Lewy body types. *Arch Neurol* **49**, 1043-1046.

[13] Sahakian BJ, Owen AM, Morant NJ, Eagger SA, Boddington S, Crayton L, Crockford HA, Crooks M, Hill K, Levy R (1993) Further analysis of the cognitive effects of tetrahydroaminoacridine (THA) in Alzheimer's disease: assessment of attentional and mnemonic function using CANTAB. *Psychopharmacology (Berl)* **110**, 395-401.

[14] Chamberlain SR, Sahakian BJ (2006) The neuropsychology of mood disorders. *Curr Psychiatry Rep* **8**, 458-463.

[15] Sheline YI, Gado MH, Kraemer HC (2003) Untreated depression and hippocampal volume loss. *Am J Psychiatry* **160**, 1516-1518.

[16] Steffens DC (2008) Separating mood disturbance from mild cognitive impairment in geriatric depression. *Int Rev Psychiatry* **20**, 374-381.

[17] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999) Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* **56**, 303-308.

[18] Blackwell AD, Sahakian BJ, Vesey R, Semple JM, Robbins TW, Hodges JR (2004) Detecting dementia: novel neuropsychological markers of preclinical Alzheimer's disease. *Dement Geriatr Cogn Disord* **17**, 42-48.

[19] Nelson HE (1982) *National Adult Reading Test (NART): Test Manual*, NFER-Nelson, Windsor.

[20] Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO (1982) Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* **17**, 37-49.

[21] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198.

[22] Rosen WG, Mohs RC, Davis KL (1984) A new rating scale for Alzheimer's disease. *Am J Psychiatry* **141**, 1356-1364.

[23] Weschler D (1987) *Weschler Memory Scale – Revised*, Psychological Corporation, San Antonio.

[24] Baddeley A, Emslie H, Nimmo-Smith I (1994) *Doors and People*, Thames Valley Test Company, Bury St Edmunds.

[25] McKenna P, Warrington EK (1983) *Graded Naming Test*, NFER Nelson, Windsor.

[26] Bozeat S, Lambon Ralph MA, Patterson K, Garrard P, Hodges JR (2000) Non-verbal semantic impairment in semantic dementia. *Neuropsychologia* **38**, 1207-1215.

[27] Warrington EK (1996) *Cambden Memory Tests*, Psychology Press, Hove.

[28] Newcombe F (1969) *Missile Wounds of the Brain*, Oxford University Press, London.

[29] Hodges JR, Patterson K (1995) Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications. *Neuropsychologia* **33**, 441-459.

[30] Hodges JR, Erzinclioglu S, Patterson K (2006) Evolution of cognitive deficits and conversion to dementia in patients with mild cognitive impairment: a very-long-term follow-up study. *Dement Geriatr Cogn Disord* **21**, 380-391.

[31] Sahakian B, Jones G, Levy R, Gray J, Warburton D (1989) The effects of nicotine on attention, information processing, and short-term memory in patients with dementia of the Alzheimer type. *Br J Psychiatry* **154**, 797-800.

[32] Downes JJ, Roberts AC, Sahakian BJ, Evenden JL, Morris RG, Robbins TW (1989) Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: evidence for a specific attentional dysfunction. *Neuropsychologia* **27**, 1329-1343.

[33] Owen AM, Sahakian BJ, Hodges JR, Summers BA, Polkey CE, Robbins T (1995) Dopamine-dependent frontostriatal planning deficits in early Parkinson's disease. *Neuropsychology* **9**, 126-140.

[34] Cattell RB (1978) *The Scientific Use of Factor Analysis*, Plenum, New York.

[35] Fowler KS, Saling MM, Conway EL, Semple JM, Louis WJ (2002) Paired associate performance in the early detection of DAT. *J Int Neuropsychol Soc* **8**, 58-71.

[36] Jacova C, Hsiung GY, Feldman HH (2006) Dropouts and refusals in observational studies: lessons for prevention trials. *Neurology* **67**, S17-S20.

[37] Coley N, Gardette V, Toulza O, Gillette-Guyonnet S, Cantet C, Nourhashemi F, Andrieu S, Grand A, Vellas B (2008) Predictive factors of attrition in a cohort of Alzheimer disease patients. The REAL. FR study. *Neuroepidemiology* **31**, 69-79.

[38] Diniz BS, Pinto JA Jr, Gonzaga ML, Guimaraes FM, Gattaz WF, Forlenza OV (2009) To treat or not to treat? A meta-analysis of the use of cholinesterase inhibitors in mild cognitive impairment for delaying progression to Alzheimer's disease. *Eur Arch Psychiatry Clin Neurosci* **259**, 248-256.