Diagnostic differentiation of mild cognitive impairment due to Alzheimer's disease using a hippocampus-dependent test of spatial memory

Kuven K Moodley[1], Ludovico Minati[1,2], Valeria Elisa Contarino[3], Sara Prioni[4], Ruth Wood[1], Rebecca Cooper[5], Ludovico D'Incerti[3], FabrizioTagliavini[4], Dennis Chan[1]

[1]Brighton and Sussex Medical School, Brighton, UK
[2]U.O. Direzione Scientifica, Fondazione IRCCS Istituto Neurologico "Carlo Besta", Milano, Italy
[3]U.O Neuroradiologia, Fondazione IRCCS Istituto Neurologico "Carlo Besta", Milano, Italy
[4]U.O. Neuropatologia, Fondazione IRCCS Istituto Neurologico"Carlo Besta", Milano, Italy
[5]Brighton and Sussex University Hospitals NHS Trust, UK

Corresponding Author:
Dr Dennis Chan, Herchel Smith Building for Brain and Mind Sciences, University of Cambridge, Forvie Site, Robinson Way, Cambridge CB2 0SZ
Email: dc598@medschl.cam.ac.uk

Word count	Summary	331
	Main text	5736

Number of tables (main text)	4
Number of figures (main text)	5
Supplementary tables	4

Keywords: Alzheimer's disease, mild cognitive impairment, spatial memory, hippocampus, precuneus

Summary

The hippocampus is one of the earliest brain regions affected in Alzheimer's disease (AD) and tests of hippocampal function have the potential to detect Alzheimer's disease (AD) in its earliest stages. Given that the hippocampus is critically involved in allocentric spatial memory, this study applied a short test of spatial memory, the 4 Mountains Test (4MT), to determine whether test performance can differentiate MCI patients with and without CSF biomarker evidence of underlying AD and whether the test can distinguish patients with mild cognitive impairment (MCI) and mild AD dementia when applied in different cultural settings.

Healthy controls (HC), patients with MCI and mild AD dementia were recruited from study sites in UK and Italy. Study numbers were: HC (UK 20, Italy 10), MCI (UK 21, Italy 14), AD (UK 11, Italy 9). Nineteen UK MCI patients were grouped into CSF biomarker-positive (MCI+, n=10) and biomarker-negative (MCI-, n=9) subgroups. Behavioural data were correlated with hippocampal volume and cortical thickness of the precuneus and posterior cingulate gyrus.

Spatial memory was impaired in both UK and Italy MCI and AD patients. Test performance additionally differentiated between MCI+ and MCI- subgroups ($p = 0.001$). A 4MT score of ≤ 8/15 was associated with 100% sensitivity and 90% specificity for detection of early AD (MCI+ and mild AD dementia) in the UK population, and with 100% sensitivity and 50% specificity for detection of MCI and AD in the Italy sample. 4MT performance correlated with hippocampal volume in the UK population and cortical thickness of the precuneus in both study populations.

In conclusion, performance on a hippocampus-sensitive test of spatial memory differentiates MCI due to AD with high diagnostic sensitivity and specificity. The observation that similar diagnostic sensitivity was obtained in two separate study populations, allied to the scalability and usability of the test in community memory clinics, supports future application of the 4MT in the diagnosis of pre-dementia AD.

Introduction

Alzheimer's disease (AD) is the commonest cause of dementia and its management represents one of the highest priorities for health systems worldwide. It is now recognised that the AD pathological process begins many years before the onset of dementia and this is reflected in the replacement of the 1984 NINDS-ADRDA diagnostic criteria for AD by new criteria that encompass the concept of prodromal AD (Dubois *et al.,* 2010) and that of mild cognitive impairment (MCI) as a pre-dementia stage of AD (Albert *et al.,* 2011). However, MCI due to underlying AD may be clinically indistinguishable from MCI due to other causes, including non-neurodegenerative disorders such as anxiety. Furthermore, memory tests commonly used in clinical psychometric testing, such as the Rey Auditory Verbal Learning Test (RAVLT), the Logical Memory test of the Wechsler Memory Scale, or the various versions of the paired associate learning test (PAL) (Wechsler, 1945) lack diagnostic specificity for AD(Fowler *et al.,* 2002)

Differentiation of MCI due to AD has major prognostic implications. However, while testing for AD biomarkers, in the form of amyloid-PET scanning or CSF studies of amyloid and tau, has discriminatory value their usage in routine clinical diagnostic practice is limited by their invasive nature, high cost and restricted availability, all of which preclude their application to the wider population of patients with MCI that are diagnostic not in university hospitals but in community clinics.

An alternative strategy for identification of MCI due to AD involves the use of a theory-driven approach based on the knowledge that the hippocampus and related medial temporal lobe structures are affected from the earliest stages of AD. Evidence that the hippocampus is critically involved in spatial memory originates from the initial demonstration of place-related firing activity of hippocampal neurons in freely moving rats (O'Keefe and Dostrovsky, 1971), which led to the "cognitive map" theory of hippocampal function (O'Keefe and Nadel, 1978). Subsequent work has shown that the human hippocampus is also involved in spatial cognition, as part of a network of brain regions including the precuneus and posterior cingulate gyrus (Maguire *et al*., 1998; Burgess *et al*., 2001; Iaria *et al*., 2007).

The 4 Mountains Test (4MT) is a brief behavioural test of spatial memory designed to reflect hippocampal function (Hartley *et al*., 2007). Landscapes containing four mountains in differing configurations are computer-generated, with subsequent presentation of same- and rotated-view landscapes permitting assessment of allocentric (viewpoint-independent) spatial perception and memory.

The initial design and testing of the 4MT paradigm encompassed testing of spatial and nonspatial perception and memory, the latter two conditions involving alterations of conditions such as lighting levels and vegetation colour. In an initial study patients with focal hippocampal damage exhibited impairment of spatial memory but preservation of spatial perception, nonspatial perception and nonspatial memory (Hartley *et al*., 2007). Given the early involvement of the hippocampus in AD and its relative sparing in non-AD dementias, and the clinical importance of differential diagnosis of dementia, two subsequent studies applied the 4MT to groups of patients with dementia. Bird *et al*. (2010) showed that performance on the 4MT spatial memory test was impaired in patients with early AD and discriminated these patients from those with frontotemporal dementia (FTD), representing a non-AD dementia (Bird *et al*., 2010; Pengas *et al*., 2010). Pengas *et al.* (2010) employed the 4MT as one of several behavioural tasks of spatial memory, with similar observations that spatial memory testing differentiated between AD and clinical subtypes of FTD.

The current study builds on the results from these previous patient studies and critically the focus shifts from differential diagnosis of established dementia to that of MCI. In view of the substantial neuropathological evidence of hippocampal damage in AD occurring prior to the onset of dementia (Braak and Braak, 1991; Arriagada *et al*., 1992; Price *et al*., 2009) the central study hypothesis was that spatial memory performance (as tested using the 4MT) is impaired in prodromal AD, manifest as MCI. Since proof of hypothesis would have significant implications for the use of this test in clinical diagnostic practice, and in view of the importance attached worldwide to the detection of AD prior to the onset of dementia, several study objectives were defined. The primary objective was to determine whether testing of spatial memory would differentiate MCI due to AD from MCI without biomarker evidence of underlying AD, using CSF AD biomarker profiles to distinguish the two groups of MCI patients. The second objective was to determine whether performance on the 4MT correlated with structural measurements of brain regions involved in spatial processing, specifically the hippocampus, precuneus and posterior cingulate gyrus. The third objective was to evaluate the discriminative ability of the 4MT when applied to patients recruited from countries with different clinical diagnostic practices. To this effect patients with MCI and early AD were recruited in two parallel studies involving memory clinics in the UK and Italy, with patient data in each instance compared locally with those from matched control subjects.

METHODS AND MATERIALS

Subjects

Patients with MCI and mild AD dementia were recruited at two sites. At both sites, MCI and AD were diagnosed respectively according to Petersen (Petersen, 2004) and McKhann criteria (McKhann *et al*., 2011). For patients with AD dementia, mild dementia was determined by Mini-Mental State Examination (MMSE) (Folstein *et al*., 1975) scores > 22 and a Clinical Dementia Rating (CDR) (Morris, 1993) of one, the latter score representing the overall CDR, rather than CDR - Sum of Boxes, as a global measure of the severity of cognitive impairment.
All MCI patients had MMSE scores ≥ 26 (UK) or ≥ 25 (Italy) with CDR ≤ 0.5. MCI was diagnosed in clinic on the basis of a history of change in cognitive performance from baseline, corroborated by an informant, with objective evidence of cognitive decline, in the absence of dementia and presence of largely intact functional activities. For UK-based patients objective cognitive decline was established using either the Addenbrooke's Cognitive Examination – Revised (Mioshi et al. 2006) or the Queen Square Screening Test for Cognitive Deficits (© EK Warrington 2003) plus MMSE. UK MCI patients were not stratified further into amnestic/non-amnestic/multi-domain subsets. For Italy-based patients, objective cognitive decline was established using the MMSE and short story recall test (Novelli *et al*., 1986; Caffarra *et al*., 2002)] and only patients with amnestic MCI (single- or multi-domain) were recruited for the study, in line with the clinical practice at this Italian site of stratifying MCI patients to assess risk of conversion to AD, patients with amnestic MCI being considered to have a higher risk of converting to AD compared to patients with non-amnestic MCI (Petersen 2004).

The exclusion criteria include i) evidence of significant vascular lesion load on imaging, ii) Hachinski Ischaemic Score > 4 (Moroney et al. 1997). "Significant vascular lesion load" was defined as the presence of cortical infarcts, extensive and/or confluent white matter hyperintensities (WMH) and WMH > 10mm diameter. These guidelines for determining the diagnostic significance of WMH are comparable to those listed in the positioning paper by Wardlaw et al. (2013) on neuroimaging standards for research into small vessel disease and also to the imaging criteria outlined in the Sachdev et al. (2014) recommendations for the diagnosis of vascular cognitive disorder.

UK-based patients were recruited from the Cognitive Disorders Clinic, Hurstwood Park Neurological Centre, Hayward Heath, West Sussex, and consisted of 21 MCI and 11 AD patients; as part of their diagnostic workup 19 MCI patients testing for CSF AD biomarkers using the CSF collection protocol and ELISA assay kits (Innotest, Innogenetics, Ghent, Belgium) as outlined in the CSF substudy of the Alzheimer's Disease Neuroimaging Initiative (Shaw *et al*., 2009). Patients were divided into MCI biomarker-positive (MCI+), indicative of prodromal AD in line with the criteria of Dubois et al., (2010) and Albert et al., (2011), and MCI biomarker-negative (MCI-) groups on the basis of abnormal CSF β-amyloid$_{1-42}$ and tau levels according to updated normal ranges for these indices (Mulder *et al*., 2010). These were as follows: CSF β-amyloid$_{1-42}$>550 pg/mL, CSF tau <375 pg/mL, tau:amyloid ratio <0.8. MCI+ patients were classified accordingly on the basis of CSF β-amyloid$_{1-42}$< 550 pg/mL *and* CSF tau > 375 pg/mL, with tau:amyloid ratio > 0.8. For the MCI-group, all but three patients had CSF amyloid and tau levels in the normal range, with correspondingly normal CSF tau:amyloid ratios; the remaining three patients had high tau levels but had normal range amyloid levels (>900 pg/mL in all cases) and additional had a history of stable cognitive function over a minimum 24 month follow-up period and as such were diagnosed clinically as having stable MCI.

Italy-based patients were recruited from the specialist dementia clinic at the Istituto Neurologico "Carlo Besta", Milan, and consisted of 14 MCI and 9 AD patients. MCI patients recruited at this site did not undergo amyloid-PET or CSF biomarker testing as part of their diagnostic workup, reflecting the clinical practice at this site. One patient was not included in the final analyses due to lack of engagement with the task.

At both sites all patients underwent clinical and laboratory assessments (including blood tests for thyroid function, vitamin B12 status) to exclude potentially treatable causes of cognitive decline. All subjects observed to have a significant vascular lesion load on initial brain scanning were excluded from the study.

At both study sites age-matched healthy control (HC) subjects without a history of cognitive impairment were recruited (UK n=20, Italy n=10). Separate approval was obtained from the UK Research Ethics Committee South East Coast - Brighton and Sussex (references 10/H1107/23 and 13/LO/0277) and from the human subjects ethics committee of the Fondazione IRCCS Istituto Neurologico "Carlo Besta" (reference fMRI-AD). At both sites the study was performed in accordance with the Declaration of Helsinki. All participants gave written informed consent.

Demographics

There were no significant differences between HC, MCI and AD groups in terms of age, gender and years of education when the UK and Italy cohorts were compared separately (Table 1). A comparison of the UK and Italy control subjects revealed that the latter were significantly older (UK: age 62.6±6.2 years, Italy age 71.3±4.0 years, $t(28)=4.1$, $p<0.001$).

No significant differences in these demographics were noted on comparison of the UK MCI+ and MCI- subgroups, who were also matched for disease duration (Table 2).

Insert Table 1 around here

Behavioural studies

General neuropsychological assessment

UK and Italy subjects underwent a battery of neuropsychological tests. The following domains were tested, with the tests used in parentheses (where not otherwise specified, these were common to both centres): 1) Episodic memory [UK: Rey Auditory Verbal Learning Test, RAVLT (Rey 1941, Van der Elst *et al.*, 2005); Italy: short story recall and Rey figure recall (Spinnler and Tognoni 1987; Caffarra *et al*., 2002)], 2) Attention and Executive function [Trail Making Test A and B (Reitan 1958; Giavagnoli et al., 1996); Italy only: FAB (Appollonio et al., 2005)], 3) Executive function [Lexical and semantic fluency (Benton et al., 1994; Novelli et al., 1986; Tombaugh et al., 1999)], 4) Working memory [Digit span (Blackburn and Benton, 1957; Spinnler and Tognoni 1987); Italy only: Corsi block-tapping task (Spinnler and Tognoni 1987)], 5) Higher visual processing [Object decision (James and Warrington, 1991)], 6) Premorbid IQ [UK only: National Adult Reading Test (NART) estimated IQ (Nelson & Willison 1991)].

With regard to assessment of episodic memory, differences in the tests employed at the two study sites reflected local differences in clinical practice. Other differences in test administration between sites related to scoring of the Trail making B test, with UK and Italy participants who were unable to complete the test given scores of 300 and 600 respectively in reflection of local practice, and administration of the semantic fluency test, with two noun categories administered in the UK while an additional

category was assessed in Italy). Again in keeping with local clinical diagnostic practice, UK neuropsychometric data are presented as raw scores whereas Italy data are corrected for age and education, with the exception of the VOSP data which are presented as raw scores due to the lack of an age/education correction algorithm for this particular test.

In the UK, the test battery outlined above was not undertaken in three HC subjects, who opted out of these tests, and two MCI patients (one with no CSF biomarker data, one CSF biomarker negative) who declined testing due to anxiety. The NART was not performed in one MCI patient who did not undergo CSF studies.

The 4 Mountains Test

A fuller description of the 4MT is provided in the article by Hartley *et al.* (2007).

In order to maintain consistency with the terminology used in previous work involving application of the 4MT (Hartley *et al*., 2007; Bird *et al*., 2010; Pengas *et al*., 2010), the allocentric spatial perception and memory subtests of the 4MT are referred to as place perception (PP) and place memory (PM) tests, and abbreviated accordingly in the tables and figures. Testing was preceded by the presentation of three training slides to aid familiarisation with the task. In brief, the task involves presentation of computer-generated landscapes containing four mountains with a semicircular mountain range in the background. Participants are shown a sample landscape along with a panel of four landscapes, consisting of the original landscape seen from another viewpoint and three foils. For the perceptual task this panel is presented at the same time as the target image, whereas for the memory task the panel is presented after a 2 s delay. The three foils for each test item were generated from the target image as follows. For the “spatial” foil the position of one mountain is shifted but the order of the four mountains around the centre of the image is preserved. For the “ordinal” foil the ordering of the mountains about the origin is altered by exchanging the location of two or more of the mountains. In the “elemental” foil the shape and/or size of one mountain is changed, whereas spatial layout is preserved. The design and usage of these foils helped to ensure that generation of an allocentric representation of the presented image would be required to distinguish the target image from the foils while maintaining local visual similarities between target and foils.

In the PP task a maximum of 30 s is given for a forced choice match-to-sample. In the PM task a 4MT landscape is shown for 10 s, followed by a 2 s interval during which the landscape is removed, with subsequent presentation of the original

landscape seen from another viewpoint and three foils (Figure 1), with a maximum of 30s given for the forced choice delayed match-to-sample.
Non-spatial features (e.g. lighting level, extent of vegetation cover) varied between presentation and testing in order to ensure that correct matching to sample could not be made on the basis of non-spatial aspects of the task.
All landscapes are shown in printed form on A4 sized pages within a ring-bound booklet such that, for the PP task, the target image is shown on one page and the four match-to-sample choices simultaneously on a separate page. For the PM task, after presentation of the target image, participants are shown a blank white page for 2s before being presented with the four match-to-sample choices on a subsequent page. Total test duration was approximately 20 minutes.

In the Bird *et al*. (2010) and Pengas *et al*. (2010) studies 15 PP and 15 PM scenes were presented, reflecting the fact that a common aim of both studies related to the comparison of spatial perception and spatial memory performance. In contrast Hartley et al. (2012) omitted the PP subtest altogether, in light of the study aim of determining an association between 4MT PM performance and hippocampal volume. In this study, some of the UK patients were initially tested on the 4MT protocol including a 15 item PP; the remaining 16 UK participants (5 controls, 10 MCI, 1 AD) recruited subsequently and the Italy-based participants were tested on a shortened 6 item PP in view of the lack of study hypothesis associated with the PP component of the 4MT and in order to shorten test duration. Use of 15 or 6 item PP had no effect on the associated PM scores in the same individuals (eg MCI patients tested with 15 PP items, PM scores 7.3±3.6; MCI patients tested with 6 PP items, PM scores 7.8±1.9. P=0.7).

Insert Figure 1 around here

Volumetric MRI studies

UK. UK-based subjects underwent MRI on a 1.5T scanner (Avanto, Siemens AG, Erlangen DE) at the Clinical Imaging Sciences Centre, Brighton and Sussex Medical School. Two AD patients were unable to tolerate MRI scanning. T1-weighted 3D volumetric MRI data were acquired by means of a magnetization-prepared rapid-acquisition gradient-echo (MPRAGE) sequence, having 1x1x1mm^3 voxel size,

TI=600ms, TE=4 ms, TR=1160ms. Due to logistical reasons, 4 patients from MCI+ group could not have a scan, structural correlations are therefore reported for the remaining 17 cases.

Italy. Italy-based subjects underwent MRI on a 3.0 T scanner (Achieva TX, Philips Medical Systems NV, Best NL). T1-weighted 3D volumetric data were acquired by means of a turbo field-echo sequence, having 1x1x1mm^3 voxel size, TI=1223 ms, TE=4.6 ms, TR=9.9 ms. All Italy subjects completed scanning but quantitative data could not be extracted from one MCI subject due to severe motion artefact, structural correlations are therefore reported for the remaining 22 cases.

Cortical thickness was measuring using the FreeSurfer workflow (Massachusetts General Hospital, Harvard University, Boston MA, USA), which, as detailed elsewhere (Fischl, 2012), involves iterative reconstruction of the white-gray matter interface and pial surface, and subsequent labelling with non-linear morphing to a probabilistic brain atlas. The Desikan probabilistic brain atlas was used (Desikan *et al*., 2006), with the posterior cingulate gyrus and precuneus chosen as regions of interest (ROIs) for quantitative analysis. All segmentations were manually cleaned up and refined by a specialized operator blinded to disease status.

Hippocampal volumes were measured using the FSL/FIRST tool (FMRIB, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Oxford, UK) (Patenaude et al. 2011). This additional segmentation was undertaken in view of the current lack of consensus regarding the different semi-automated tools used to analyse hippocampal volume (Morey et al., 2009; Sánchez-Benavides et al., 2010; Mulder et al, 2014); on our images, preliminary expert evaluation revealed that FreeSurfer tended to include parts of other medial temporal structures such as the parahippocampal gyrus and amygdala.

These ROIs were specifically chosen to encompass regions that are affected early in AD and represent components of the brain network considered to underpin memory functions (Greicius *et al*., 2003; Buckner *et al*., 2008).

Statistical analysis

For HC, MCI and AD group data, ANOVA was used to determine between-group differences for the demographic and neuropsychometric data. Post-hoc pair-wise comparisons were undertaken using the conservative Scheffé test, which controls for family wise error rate across all planned contrasts. Two-tailed t-tests were used for comparison of MCI+ and MCI- subgroups; the alpha threshold was Bonferroni adjusted for multiple comparisons (twelve comparisons inclusive of general and 4MT

psychometric measures; alpha = 0.004). To allow comparison between patients who were tested with 15 and 6 PP scenes, the former PP scores were rescaled to scores in the range of 0-6.

Place memory scores were correlated with total (i.e. combined left and right) hippocampal volumes. Additional correlations were undertaken with total cortical thickness of the precuneus and posterior cingulate gyrus, in view of previous studies that have suggested a role for these brain regions in spatial memory (Pengas *et al*., 2012). Age was inserted as covariate for both analyses, having excluded effects of sex. Total intracranial volume from Freesurfer segmentation was included as an additional covariate in hippocampal volume-place memory correlation analyses.

All correlations were undertaken using patient data only, with exclusion of control data in order to avoid biasing the correlations as a result of the strong group differences. A univariate general linear model was used to assess differences between HC, MCI and AD quantitative MRI group data and between UK MCI +ve and MCI -ve subgroup data, age inserted as a covariate for comparisons involving the PCG and precuneus, age and TIV inserted as a covariates for hippocampal comparisons; Bonferroni adjusted for multiple comparisons; alpha = 0.02).

RESULTS

Behavioural studies

General neuropsychometric assessment

Consistent with their diagnostic classification, MCI patients were impaired on tests of delayed recall and executive function whereas AD patients were impaired in all tested cognitive domains. The full breakdown of test scores, with the UK and Italy study populations tabulated separately, are provided in the Supplementary Tables 1 and 2.

There were no significant differences in the test scores obtained by the MCI- and MCI+ patients (Table 2).

Insert Table 2 around here

4 Mountains Test

Place perception

For the UK study population, ANOVA revealed group differences in test performance ($p < 0.001$, Table 3). Pairwise comparisons, corrected for multiple comparisons, revealed that the difference in scores was significant between HC and AD groups ($p < 0.001$) and between MCI and AD groups ($p = 0.001$). Further analyses revealed a significant difference between the MCI biomarker negative and AD groups only ($p < 0.001$). No significant differences were observed between the MCI subgroups or between MCI subgroups and HC.

For the Italy study population, ANOVA revealed group differences in test performance ($p = 0.02$, Table 3). Pairwise comparisons, corrected for multiple comparisons, revealed a significant difference in PP performance between HC and AD groups ($p = 0.02$).

Place memory

For the UK study population, ANOVA revealed group differences in test performance ($p < 0.001$, Table 3). Pairwise comparisons, corrected for multiple comparisons, revealed that the difference in scores was significant between HC and MCI groups ($p < 0.001$), between HC and AD groups ($p < 0.001$) and between MCI and AD groups ($p = 0.004$). Further analyses revealed significant differences in PM scores for the following pairwise group comparisons: HC vs MCI+ ($p < 0.001$), HC vs AD ($p < 0.001$), MCI- vs MCI+ ($p = 0.002$), MCI- vs AD ($p < 0.001$). By comparison no significance difference in PM test scores was observed for the comparison between HC and MCI- ($p = 0.3$) or between MCI+ and AD ($p = 0.6$).

For the Italy study population, ANOVA revealed group differences in test performance ($p < 0.001$, Table 3). Corrected pairwise comparisons revealed that significant differences in scores were observed for HC vs MCI ($p = 0.002$) and HC vs AD ($p < 0.001$) but not for MCI vs AD ($p = 0.5$).

Figure 2 shows the individual PM scores and the differences in score between participant groups, with data from UK and Italy participants represented separately.

Insert Table 3 around here

Insert Table 4 around here

Insert Figure 2 around here

Receiver operating characteristic (ROC) curves

The discriminative ability of 4MT place memory testing is illustrated by use of Receiver Operating Characteristics (ROC) curves (Figure 3). For the UK population test performance is associated with an area under the curve of 0.98 (differentiating MCI+ and AD patients from controls and MCI- patients) and of 0.90 for the Italy population (differentiating MCI and AD patients from controls). PM scores of 8 or below were associated with 100% sensitivity and specificities of 90% and 50% for the UK and Italy study populations respectively.

Insert Figure 3 around here

Correlations between spatial memory performance and quantitative MRI data

Partial correlations undertaken for the UK patient population (MCI and AD), corrected for age and total intracranial volume (hippocampal volume only), revealed, after averaging between left and right hemisphere, significant associations between 4MT PM score, hippocampal volume ($r=0.42$, $p= 0.03$, not surviving the corrected alpha threshold of 0.02) and cortical thickness of the precuneus ($r=0.55$, $p = 0.003$) but not the posterior cingulate gyrus ($r=0.19$, $p = 0.4$). For the Italy patient population (MCI and AD) a significant association was observed with cortical thickness of the precuneus($r=0.58$, $p = 0.006$) but not with hippocampal volume ($r=0.09$, $p = 0.7$) or with cortical thickness of the posterior cingulate gyrus ($r=0.34$, $p = 0.1$). Scatterplot representations of the correlations between behavioural data and measures of hippocampal volume and cortical thickness of the precuneus and posterior cingulate gyrus are provided in Figure 4.

Considering each hemisphere in isolation, no statistically significant effects were obtained.

Vertex level comparisons, applied to the pooled UK and Italy study data, revealed an association between PM and cortical thickness of the precuneus, lateral parietal and supramarginal regions (Figure 5).

Insert Figure 4 around here

Insert Figure 5 around here

Discussion

Spatial memory performance, as evaluated using the brief 4 Mountains Test (4MT), is impaired in patients presenting with mild cognitive impairment (MCI) with biomarker evidence of underlying Alzheimer's disease (AD). Performance on the 4MT place memory testing (PM) was significantly impaired in patients with MCI and AD compared to age- and gender-matched control subjects, in keeping with previously published results (Bird *et al*., 2010; Pengas *et al*., 2010). When the UK MCI patients were grouped according to CSF AD biomarker status, 4MT PM scores differed significantly between MCI patients with and without CSF biomarker evidence of AD, in stiking contrast to the lack of difference between groups in terms of demographics, symptom duration, premorbid IQ and performance on general neuropsychometric testing. Of particular note was the observation that MCI subgroups did not differ in the scores obtained on the Rey Auditory Verbal Learning Test and the Trail Making Test part B, given that both tests that have been shown to have high diagnostic sensitivity for early AD and as such are widely used in clinical and research practice (Chapman *et al*., 2011; Ewers *et al*., 2012; Gainotti *et al*., 2014).

In order to explore the diagnostic ability of the 4MT in different clinical and cultural settings testing was undertaken in parallel on patients recruited from two memory clinics in the UK and in Italy, with patient data at each site compared with data obtained from age- and gender-matched control subjects with no history of cognitive impairment. For both UK and Italy study populations the PM scores obtained from both MCI and AD patient groups were significantly lower that the corresponding control scores; in both study populations no significant difference was observed

between MCI and AD scores, in keeping with previous findings (Bird *et al*., 2010). Comparison of data from the two study sites revealed that control PM scores were lower in the Italy control subjects but similar for the two AD groups. When comparison is made across the Italy and UK MCI patients, with the latter grouped according to CSF biomarker status, the mean PM score of the Italy patients was identical to that of the UK MCI biomarker positive subgroup (5.8) but significantly different to that of the UK MCI biomarker negative subgroup (mean score 9.6). This may reflect the selection at the Italy site of patients with a diagnosis of amnestic MCI who are considered more likely to have underlying AD, and as such are more in keeping with the UK MCI biomarker positive subgroup.

The ability of 4MT PM testing to detect the presence of disease is illustrated by the determination of test sensitivity and specificity and calculation of the area under the ROC curve (AUC). For the UK population, the AUC was 0.98 for patients with early AD (i.e. MCI+ and mild AD dementia), with a score of 8 or below yielding100% sensitivity and 90% specificity for the detection of early AD. For the patients recruited from an Italy memory clinic, in whom CSF AD biomarker testing was not undertaken as part of the clinical diagnostic process, the AUC was 0.9, with a score of 8 or below yielding 100% sensitivity and 50% specificity for detection of a clinically defined disease state (MCI and AD). Further scrutiny of these latter data indicates that the lower specificity of the 4MT in the Italy population results primarily from the lower scores obtained by the control subjects (UK HC 11.1 ± 2.1; Italy HC 9.0 ± 2.3; $t(28) = 2.52$; $p = 0.02$). When the two study sites are considered together, a PM score of 8 or below was associated with a positive predictive value between 82-88% and a negative predictive value of 100%. These findings indicate that the 4MT has high diagnostic sensitivity for early AD in different countries, while the diagnostic specificity associated with a PM score of 8/15 varied according to clinical practice, with higher specificity observed when clinical diagnosis was supplemented by testing for AD biomarkers.

Previous work (Hartley *et al*., 2007, Bird *et al*., 2010) has shown that there is relative preservation of place perception in the context of hippocampal damage and that 4MT PM, but not PP, scores differentiated AD from other dementias. As such the decision was made *a priori* to shorten the PP test battery from 15 to 6 items in order to reduce the test time, in light of the potential future application of the 4MT as a test for use in clinical practice. A comparison of the data acquired from subjects tested with either 15 to 6 PP items revealed no effect on the subsequent PM scores and overall the PP

scores did not discriminate between controls and MCI patients in either UK or Italy study populations, or between MCI+ and MCI- subgroups. PP scores were significantly lower in patients with AD dementia, consistent with the spread of pathology into cortical regions in more advanced disease.

Quantitative MRI analyses revealed that hippocampal volume, and the cortical thickness of the precuneus and posterior cingulate gyrus were reduced in patients with MCI, and that this atrophy was more severe and extensive in patients with AD dementia. These observations were consistent across both UK and Italy study populations and are in keeping with a number of previous MRI studies conducted in MCI and AD (Du et al., 2001, Hämäläinen et al, 2007, Fennema-Notestine et al, 2009). However, and in contrast to the place memory test performances, no differences in these ROI structural measurements were observed between MCI+ and MCI- subgroups.

PM scores correlated with total hippocampal volume in the UK patient group and with the cortical thickness of the precuneus in both UK and Italy patient populations but not with cortical thickness of the posterior cingulate gyrus. The correlation with hippocampal volume in the UK patient population is in keeping with the central study hypothesis with regard to impairment of hippocampal function in early AD, and as such it is proposed that this relationship is causal, and not merely associative, in nature. Two factors may explain the absence of any observed correlation between PM scores and hippocampal volume in the Italy patient population; first, the smaller sample size at this study site and second, the reduced dynamic range as a consequence of the high degree of overlap between MCI and AD patient scores (see Figure 2).

The correlation between PM score and cortical thickness of the precuneus can be interpreted in two main ways. The first of these is that this correlation is associative rather than causal, as a reflection of the known early involvement of the precuneus in AD (Braak and Braak, 1991; Mirra *et al*., 1991; Thal *et al*., 2002). However, this is inconsistent with the failure to observe any correlation with the cortical thickness of the posterior cingulate gyrus, given that this brain region is similarly early in AD.

The second interpretation is that the precuneus, along with the hippocampus, is directly implicated in allocentric spatial memory. Evidence from non-human primates (Selemon and Goldman-Rakic, 1988) and from human studies (reviewed by Cavanna and Trimble 2006) suggests that the precuneus is involved in spatially-related behaviours. These study findings may therefore be consistent with the viewpoint that in humans allocentric spatial memory is subserved by a functional

network that encompasses the hippocampus and the precuneus. Task-free fMRI studies show that the hippocampus and precuneus represent highly interconnected hubs within a “default mode network” underpinning spatial and episodic memory (Greicius *et al*., 2004; Vincent *et al.*, 2006) and the vulnerability of this network to early AD (Rombouts *et al*., 2005) is of note in the context of this current study.

Several aspects of this study warrant further discussion in light of the potential future application of the 4MT as a clinical diagnostic tool. First of all, this study did not aim to assess the effect of gender on 4MT place memory performance. While previous work has suggested the presence of sexual dimorphism in spatial cognition (Maguire *et al.*, 1999) no significant difference in 4MT performance between men and women was observed in the recent study by Hartley et al.(2012). However unpublished results from a large study undertaken in healthy individuals suggest a small but significant gender effect with lower scores obtained from women on a 30-item test, with mean scores of 20 and 18.6 for men and women respectively (T. Hartley, C. Bird, H. Spiers, personal communication). The issue of gender effect on 4MT performance needs further clarification and this issue is being explored within current studies involving much larger numbers of young and older cognitively normal subjects.

The second issue relates to the lower place memory scores obtained by the Italy-based control subjects and the potential negative implications for diagnostic discrimination. However, while these lower control scores are reflected in lower diagnostic specificity in this study population (50% as opposed to 90% in the UK study population) a PM score of 8/15 or below was still associated with 100% sensitivity. In terms of the explanation for the lower control scores, it is perhaps relevant that the Italy-based control subjects were significantly older than their UK counterparts ($p < 0.001$). As is the norm for dementia research studies of this kind, in this study the primary criterion for recruitment of control subjects was an absence of reported cognitive symptoms. Testing for AD biomarkers, in the form of amyloid-PET scanning or CSF examination, was not undertaken in any control subjects in either Italy or the UK. Given that the incidence of AD rises with age, one possible explanation for the lower Italy control scores that cannot be discounted, and which would be a major confounding factor, is the presence of presymptomatic AD in the older control group.

The second aspect relates to the choice of 4MT for testing spatial memory. In the study conducted by Pengas *et al*. (2010) AD patients were found to be impaired on a variety of spatial memory tests and of these the Virtual Route Learning Test (VRLT)

was found to be slightly superior to the 4MT in terms of diagnostic accuracy (differentiation of AD from semantic dementia AUC 0.93 for VRLT vs AUC 0.85 for 4MT). However operational issues favoured the choice of the 4MT over the VRLT in this study. In the Pengas *et al*. (2010) study 2/32 (6%) of MCI patients were unable to complete the VRLT due to nausea from perceived motion, whereas all MCI (and all AD) patients tolerated the 4MT. Furthermore the need to employ a computer-based platform for the VRLT, allied with the requirement for extensive (and thus time-consuming) pre-testing task familiarisation, limits the potential future usage of the VRLT outside academic centres. By comparison the 4MT may be applied in paper as well as electronic forms and this, along with a short test duration, would favour future usage of the 4MT over the VRLT as a diagnostic tool in routine clinical practice.
These study findings therefore have impact for both academic and clinical practice. The diagnostic benefits of a theory-driven test of hippocampal function are demonstrated; no less importantly, the use of a test that fulfils both diagnostic requirements with regard to sensitivity and specificity and operational requirements with regard to scalability and usability is of particular relevance given the high prevalence of memory impairment in the ageing population and the fact that the majority of patients with MCI are evaluated in community memory clinics rather than academic centres.

Conclusion

Performance on the “4 Mountains Test” of spatial memory differentiates mild cognitive impairment due to Alzheimer’s disease with high diagnostic sensitivity and specificity. High diagnostic accuracy was observed in two separate study populations recruited from different countries. The correlation of test performance with structural measures of the hippocampus and precuneus is consistent with the role of these brain regions in spatial cognition and with their early involvement in the AD pathological process.
These findings indicate the value of spatial memory testing in the diagnosis of pre-dementia AD and of the 4MT as a diagnostic tool suitable for widespread use in routine clinical practice.

Acknowledgements

The UK substudy was supported by an Alzheimer’s Research UK grant to DC. The Italian substudy was supported by grants from the Italian Ministry of Health to FT.

The investigators would like to acknowledge Dr Tom Hartley, who led the design of the 4MT paradigm, and to express their gratitude to the participating patients and control subjects recruited from their respective memory clinics in Sussex and Milan, and to Daniele Orlandi for helpful assistance during data acquisition for the Italian substudy.

References

Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer‘s Association workgroups on diagnostic guidelines for Alzheimer’s disease. Alzheimers Dement 2011; 7: 270–79.

Appollonio I, Leone M, Isella V, Piamarta F, Consoli T, Villa ML, Forapani E, Russo A, Nichelli P. The Frontal Assessment Battery (FAB): normative values in an Italian population sample. Neurol Sci 2005; 26: 108-16.

Arriagada PV, Marzloff K, Hyman BT. Distribution of Alzheimer-type pathologic changes in nondemented elderly individuals matches the pattern in Alzheimer's disease. Neurology 1992; 42: 1681–88.

Benton A, Hamsher K, Sivan A. Multilingual Aphasia Examination, 3rd ed. San Antonio, TX: Psychological corporation, 1994.

Bird CM, Chan D, Hartley T, Pijnenburg YA, Rossor MN, Burgess N. Topographical short-term memory differentiates Alzheimer's disease from frontotemporal lobar degeneration. Hippocampus 2010; 20: 1154–69.

Blackburn HL, Benton AL. Revised administration and scoring of the digit span test. J Consult Psychol 1957; 21: 139–43.

Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 1991; 82: 239–59.

Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: anatomy, function, and relevance to disease. Ann New York Acad Sci 2008; 1124: 1–38.

Burgess N, Maguire EA, Spiers HJ, O'Keefe J. Temporoparietal and Prefrontal Network for Retrieving the Spatial Context of Lifelike Events. Neuroimage 2001; 14: 439-53.

Caffarra P, Vezzadini G, Dieci F, Zonato F, Venneri A. Rey-Osterreith complex figure: normative values in an Italian population sample. Neurol Sci 2002; 22: 443–47.

Cavanna A, Trimble M. The precuneus: a review of its functional anatomy and behavioural correlates. Brain 2006; 129: 564-583

Chapman RM, Mapstone M, McCrary JW, Gardner MN, Porsteinsson A, Sandoval TC, et al. Predicting conversion from mild cognitive impairment to Alzheimer's disease using neuropsychological tests and multivariate methods. J Clin Exp Neuropsychol 2011; 33: 187–99.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006; 31: 968-80.

Du AT, Schuff N, Amend D, Laakso MP, Hsu YY, Jagust WJ, Yaffe K, Kramer JH, Reed B, Norman D, Chui HC, Weiner MW. Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. J Neurol Neurosurg Psychiatry 2001; 71:441-7.

Dubois B, Feldman HH, Jacova C, Cummings JL, DeKosky ST, Barberger-Gateau P, et al. Revising the definition of Alzheimer's disease: a new lexicon. Lancet Neurol 2010; 9: 1118–27.

Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, Jack CR, et al. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. Neurobiol Aging 2012; 30: 1203–14.

Fennema-Notestine C, Hagler DJ, McEvoy LK, Fleisher AS, Wu EH, Karow DS, Dale AM et al. Structural MRI biomarkers for preclinical and mild Alzheimer's disease. Hum Brain Mapp. 2009; 30: 3238–3253.

Fischl B. FreeSurfer. Neuroimage 2012; 62: 774–81.

Folstein MF, Folstein SE, McHugh PR. “Mini-mental state.” A practical method for grading the cognitive state of patients for the clinician.J Psychiatr Res 1975; 12: 189–98.

Fowler KS, Saling MM, Conway EL, Semple JM, Louis WJ. Paired associate performance in the early detection of DAT. J Int Neuropsychol Soc 2002; 8: 58-71.

Gainotti G, Quaranta D, Vita MG, Marra C. Neuropsychological predictors of conversion from mild cognitive impairment to Alzheimer's disease. J Alzheimers Dis 2014; 38: 481–95.

Giovagnoli AR, Del Pesce M, Mascheroni S, Simoncelli M, Laiacona M, Capitani E. Trail making test: normative values from 287 normal adult controls. Ital J Neurol Sci 1996; 17: 305-9.

Greicius MD, Krasnow B, Reiss AL, Menon V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc Natl Acad Sci USA 2003; 100: 253–8

Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. Proc Natl Acad Sci USA 2004; 101: 4637–42.

Hämäläinen A, Tervo S, Grau-Olivares M, Niskanen E, Pennanen C, Huuskonen J et al. Voxel-based morphometry to detect brain atrophy in progressive mild cognitive

impairment. Neuroimage 2007; 37: 1122-1131.

Hartley T, Bird CM, Chan D, Cipolotti L, Husain M, Vargha-Khadem F, et al. The hippocampus is required for short-term topographical memory in humans. Hippocampus 2007; 17: 34–48.

Hartley T, Harlow R. An association between human hippocampal volume and topographical memory in healthy young adults. Front Hum Neurosci 2012; 6: 338.

Iaria G, Chen J-K, Guariglia C, Ptito A, Petrides M. Retrosplenial and hippocampal brain regions in human navigation: complementary functional contributions to the formation and use of cognitive maps. Eur J Neurosci 2007; 25: 890-9.

James M, Warrington EK: The visual object and space perception battery. Thames Valley Test Company; 1991.

Maguire EA, Burgess N, Donnett JG, Frackowiak RS, Frith CD, O'Keefe J. Knowing where and getting there: a human navigation network. Science 1998; 280: 921–4.

Maguire EA, Burgess N, O'Keefe J. Human spatial navigation: cognitive maps, sexual dimorphism, and neural substrates. Curr Opin Neurobiol 1999; 9: 171–7.

McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7: 263-9.

Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. Int J Geriatr Psychiatry 2006; 21: 1078-85

Mirra SS, Heyman A, McKeel D, Sumi SM, Crain BJ, Brownlee LM, et al. The Consortium to Establish a Registry for Alzheimer"s Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 1991; 41: 479–86.

Morey RA, Petty CM, Xu Y, Pannu Hayes J, Wagner HR II, Lewis DV, et al. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. Neuroimage 2009; 45: 855–66.

Moroney JT, Baqiella E, Desmond DW, Hachinski VC, Molska PK, Brun A, et al. Meta-analysis of the Hachinski Ischaemic score in pathologically verified dementias. Neurology 1997; 49; 1096-1105

Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993; 43: 2412–4

Mulder C, Verwey NA, van der Flier WM, Bouwman FH, Kok A, van Elk EJ, et al. Amyloid-beta(1-42), Total Tau, and Phosphorylated Tau as Cerebrospinal Fluid Biomarkers for the Diagnosis of Alzheimer Disease. Clinical Chemistry 2010; 56: 248–53.

Mulder ER, de Jong RA, Knol DL, van Schijndel RA, Cover KS, Visser PJ, et al. Hippocampal volume change measurement: Quantitative assessment of the

reproducibility of expert manual outlining and the automated methods FreeSurfer and FIRST. Neuroimage 2014; 92: 169–81.

Nelson H, Willison J. National Adult reading test manual (NART). Windsor, UK: NFER-Nelson, 1991.

Novelli G, Papagno C, Capitani E, Laiacona M, al E. Tre test clinici di memoriaverbale a lungotermine: Taraturasusoggettinormali. Archivio di Psicologia, Neurologia e Psichiatria 1986; 47: 278-96.

O'Keefe J, Dostrovsky J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res 1971; 34: 171–5

O'Keefe J, Nadel L. The hippocampus as a cognitive map. Oxford, UK: Oxford University Press; 1978.

Patenaude BB, Smith SMS, Kennedy DND, Jenkinson MM.A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage 2011; 56: 907–22.

Pengas G, Patterson K, Arnold RJ, Bird CM, Burgess N, Nestor PJ. Lost and found: bespoke memory testing for Alzheimer's disease and semantic dementia. J Alzheimers Dis 2010; 21: 1347–65.

Pengas G, Williams GB, Acosta-Cabronero J, Ash TWJ, Hong YT, Izquierdo-Garcia D, et al. The relationship of topographical memory performance to regional neurodegeneration in Alzheimer's disease. Front Aging Neurosci 2012; 4: 17.

Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004; 256: 183-94.

Price JL, McKeel DW, Buckles VD, Roe CM, Xiong C, Grundman M, et al. Neuropathology of nondemented aging: Presumptive evidence for preclinical Alzheimer disease. Neurobiol Aging 2009; 30: 1026–36.

Reitan RM. Validity of the trail making test as an indicator of organic brain damage. Perceptual and motor skills 1958; 8: 271-6.

Rey A. L’examen psychologique dans les cas d’encephalopathie traumatique, Archives de Psychologie 1941; 28: 21.

Rombouts SA, Barkhof F, Goekoop R, Stam CJ, Scheltens P. Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: an fMRI study. Hum Brain Mapp 2005; 26: 231–9.

Sachdev P, Kalaria R, O'Brien J, Skooq I, Alladi S, Black SE et al. (2014) Diagnostic criteria for vascular cognitive disorders: a VASCOG statement. Alzheimer Dis Assoc Disord. 28, 206-18

Sánchez-Benavides G, Gómez-Ansón B, Sainz A, Vives Y, Delfino M, Peña-Casanova J. Manual validation of FreeSurfer's automated hippocampal segmentation in normal aging, mild cognitive impairment, and Alzheimer Disease subjects. Psychiatry Res 2010: 181: 219–25.

Selemon LD, Goldman-Rakic PS. Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural network subserving spatially guided behavior. Journal of Neuroscience 1988; 8: 4049–68.

Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol 2009: 65: 403–13.

Spinnler H, Tognoni G. Standardizzazione e taratura di test Neuropsicologici. Ital J Neurol Sci 1987; 8 (supplement 6): 1-20

Thal DR, Rüb U, Orantes M, Braak H. Phases of A beta-deposition in the human brain and its relevance for the development of AD. Neurology 2002; 58: 1791–800.

Tombaugh TN, Kozak J, Rees L. Normative data stratified by age and education for two measures of verbal fluency: FAS and Animal Naming. Arch Clin Neuropsychol 1999; 14: 167-177

Van der Elst W, van Boxtel MPJ, van Breukelen GJP, Jolles J. Rey's verbal learning test: normative data for 1855 healthy participants aged 24-81 years and the influence of age, sex, education, and mode of presentation. J Int Neuropsychol Soc 2005; 11: 290–302.

Vincent JL, Snyder AZ, Fox MD, Shannon BJ, Andrews JR, Raichle ME, et al. Coherent spontaneous activity identifies a hippocampal-parietal memory network. J Neurophysiol 2006: 96: 3517–31.

Wardlaw JM, Smith EE, Biessels GJ, Cordonnier C, Fazekas F, Frayne R et al. (2013) Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. Lancet Neurol 12, 822-38

Wechsler D. A Standardized Memory Scale for Clinical Use. The Journal of Psychology 1945; 19: 87–95.