Data supplement to Huntley et al. Adaptive working memory strategy training in early Alzheimer's disease: randomised controlled trial. Br J Psychiatry doi: 10.1192/bjp.bp.116.182048

Online supplement DS1

Participants

Thirty participants with early AD were recruited from memory and community health services of the South London and Maudsley NHS Foundation Trust. Diagnoses were made by Old Age Psychiatrists and multidisciplinary teams unconnected to the study. Twenty-seven participants had a diagnosis of 'probable' AD (according to NINCDS-ADRDA criteria [1]) and 'Dementia in Alzheimer's Disease (F00)' according to ICD-10 criteria [2]. Three participants (two training and one control participant) had a diagnosis of 'possible' AD, having been assessed as converting from mild cognitive impairment (MCI) to AD within the preceding 8 weeks. We acknowledge that patients at this very early stage of AD may overlap with the criteria for Mild Cognitive Impairment on objective cognitive assessment, however diagnoses were made by clinical teams based on the history of functional and cognitive deterioration and progression. For the purpose of this study 'early AD' was defined as a diagnosis of AD with mild cognitive and functional impairment, rather than indicating a recent diagnosis or early-onset of dementia. The inclusion criteria therefore stated that baseline MMSE[3] was required to be >22/30 in order to recruit participants at the earliest stage of AD, where cognitive impairment remained mild. All baseline scores on cognitive assessments are shown in Table 1. The mean length of time between diagnosis of AD being made and recruitment into the trial was 419.7 (591.76) days for the control group and 545.29 (513.42) days for the intervention group (see online Table DS1). Although the inclusion criteria allowed patients with age > 60 to be considered for the study, participants ages ranged from 65 to 88 years, and no participant had a diagnosis of Alzheimer's disease with early-onset (see online Table DS1).

Secondary outcome measures

Transfer of training effects to clinical measures of general cognitive function were examined using the mini mental state examination (MMSE)[3] and Alzheimer's disease assessment scale- cognitive section (ADAS-Cog)[4]. The MMSE is a clinically widely used 30-point pen-and-paper test incorporating assessments of orientation (10 points), immediate and delayed recall (6 points), reading, repetition, writing and copying of a shape (4 points), object recognition (2 points), following a three-stage instruction (3 points) and attention (5 points). Points are scored for each correct response, with a maximum score of 30. The ADAS-Cog is a widely used 70 point pen-and-paper assessment involving eleven subsections that evaluate word recall, word finding and naming, following commands,

orientation, copying shapes, performing a 5 stage task, recall of test instructions, word recognition, spoken language ability and language comprehension. It is reverse scored, therefore higher scores represent greater cognitive impairment.

Transfer of training effects to non trained cognitive domains were assessed using the Logical Memory I+II tasks and Paired Associates Learning task (PAL) to assess episodic memory. The logical memory I+II is a verbal episodic memory task and is taken from the Wechsler Memory Scale [5]. Participants were read a short story and asked to remember it. They were then asked to immediately recall as much of the story as possible (part I). After 25 minutes they were asked to recall the story again (part II). Each part is scored for 25 specific and 7 thematic components, with a total score of 32 points.

The PAL task examines visuo-spatial episodic memory and is sensitive to episodic memory deficits in early AD [6]. A number of boxes were presented at different locations on a computer screen. Each box covered a picture. The boxes were initially shown, followed by the pictures under each box. Each picture was then presented in the middle of the screen and the participant had to recall which picture appeared under which box, therefore testing both object and location recall. If a participant correctly recalled all the pictures, the next set of boxes had one more box/picture combination. If an error was made a new set of boxes was presented, with one fewer box/picture. If 3 errors were made, the task ended.

Transfer of training effects to executive function was assessed using the following tasks:

1) Verbal Fluency task [7]: Participants were asked to generate as many words as they could, beginning with the letter P in one minute, not including place or person names. They were then asked to generate as many types of animal they could in one minute, whose name began with any letter of the alphabet. The total number of words generated for each category was converted to a score out of 7 (> 17 words= 7, 14-17 words = 6, 11-13 words = 5, 8-10 words = 4, 6-7 words = 3, 4-5 words = 2, 2-3 words = 1, < 2 words = 0), with a maximum total score of 14 for the two tasks.

2) Grammatical Reasoning Task[8]: In this task a picture of a square and circle were presented on a computer screen. A sentence describing the relationship between the circle and square was presented above the picture and the participant had to choose whether the sentence describing the picture was true or false. The participant had 90 seconds to answer as many true/false questions as they could.

3) Odd One Out task [9]: In this task a 3 x 3 grid of objects were presented on a computer screen. Each object was made of up of one or multiple shapes or colours. One object differed from all of the others, owing to it being a different shape, combination of parts or colour. The participant had to select which object they thought was the 'odd one out'. The participant had 3 minutes to answer as many trials as possible in the time.

4) Trail Making tasks A and B [10]. In Task A, participants were asked to connect a series of numbered circles on a piece of paper as quickly as possible. In Task B, participants were again asked to connect a series of circles containing ascending numbers or letters of the alphabet. On this occasion they were asked to alternate between numbers and letters (e.g. 1-A-2-B-3-C etc) and connect up all of the circles as quickly as possible. Prior to doing the task, participants were given short practice examples to complete. If an error was made, the examiner was allowed to point this out to the participant for them to correct. Each part was timed, and a time to completion for each part of the task was recorded. If the combined time was > 300s the task was discontinued[11]. Results of part B are not reported as 7/15 control participants and 11/15 training participants were unable to complete the task at baseline.

5) Self Ordered Search task [9]: In this task a series of boxes were presented on a screen. The aim was to search through the boxes in order to find a gold coin hidden in one of the boxes. Gold coins appeared sequentially in the boxes, with a new coin appearing in one of the remaining boxes after each coin had been found. There were two rules to the task. Firstly, a coin was never hidden in the same box twice; therefore if a coin had already been found in a box, and the participant looked in that box again, they lost a "life". Secondly, if a participant looked in the same empty box twice whilst looking for a coin, they lost a "life". The task proceeded with the participant deciding which boxes to look in, and continued until a gold coin has been found in each box. If an error was made, the participant lost a "life" and a new trial started with one less box. If the participant successfully found all the gold coins, a new trial began with one additional box. The task therefore tested the participant's ability to plan and execute a strategy and also recall the spatial location of boxes searched and coins previously found.

Statistical analyses

For the primary outcome measures, mean span accuracy scores were analysed using a mixed repeated measures analysis of variance (ANOVA) in statistical Package for the Social Sciences (SPSS v 22.0)[12]. In these analyses time (pre vs post intervention) and trial type (structured trials vs random trials) were within subjects factors, and group (training vs control) was the between subjects factor. If there was evidence of significant time x group, or time x trial type x group interactions, these were then further explored by further repeated measures ANOVAs for each trial type separately and paired T tests for each group separately.

For the secondary outcome measures, maximum scores were analysed using mixed repeated measures ANOVAs with time (pre vs post intervention) as the within subjects factor and group as the between subjects factor. The effect of training on primary and secondary outcomes was also examined by calculating change scores (post –pre) and effect sizes (r). Assumptions of parametric data were assessed for all data. If the assumptions of parametric data were violated, Mann–Whitney and Wilcoxon signed-rank non parametric tests were conducted. For all analyses the α significance level was set at 0·05.

fMRI acquisition

All participants underwent pre and post intervention fMRI on a Siemens 3T scanner, at the Centre for Neuroimaging Sciences, Institute of Psychiatry, Psychology and Neuroscience, London. The mean duration between pre and post scan for all participants was 82.97 (28.11) days, with no significant difference between groups. Functional images were collected with an EPI sequence using an event-related design sequence (8-channel head coil, 30ms TE, 2s TR, 75 deg. flip angle, 64-by-64 matrix, FOV 21·1cm (such that the voxel size is isotropic 3·3mm3), 4 DDAs, 246 volumes). Whilst undergoing fMRI, participants performed a 5-digit span WM task adapted for AD subjects from a previous fMRI study of young healthy individuals [13], requiring them to encode, retain and then verbally recall the 5 digits in order. The task difficulty at both baseline and post intervention fMRI sessions was fixed at 5 span in order to control for performance differences between the fMRI sessions. Any observed changes in activation between fMRI sessions would therefore be due to effects of the intervention, rather than due to confounding effects of differential task difficulty or performance during scanning sessions. Three blocks of twenty trials were performed and structured or random span sequences were presented pseudo-randomly.

fMRI analysis

Structural MPRAGE images were registered to a template generated from the mean of all participants, using the DARTEL toolbox in SPM8 [14]. Individual participant functional data was corrected for slice timing, realigned for motion and co-registered to the participant specific structural image. Data were normalised to MNI space using the DARTEL structural template and individual participant flow fields. One participant was excluded from imaging analysis due to incomplete structural imaging data. In the first level analysis, events of interest were parameterised to ensure orthogonal contrasts. Structured trials, random trials and all incorrect responses at the encoding, maintenance and recall stages were included as regressors in the design matrix, along with 6 movement regressors. If there was excessive movement between images, (defined as > 4mm or 5 degrees of rotation), these images were included as an additional regressor of no interest. The specified time series of events were convolved with the haemodynamic response to create predictor functions. These were fitted to time BOLD series at each voxel using the General Linear model in SPM8 along with six movement parameters. The high pass filter was set to 128s to remove low-frequency drifts in signal.

Random effects analysis was conducted on group-level data. A 2 x 2 x 2 full factorial design was used with PrePost (pre vs. post) and Chunking (structured trials vs. random trials) as within subjects and group as the between subjects factor.

A region of interest (ROI) approach was applied based on the *apriori* hypotheses that the structured WM task would be associated with prefrontal and parietal activation, as had been found in previous studies in young healthy adults. Bilateral prefrontal cortex and parietal cortex ROIs were defined from the study group data set to allow for the anticipated structural and task related functional differences between AD participants used in the current study and healthy young populations examined in previous studies [13, 15, 16]. In order to avoid selection bias, the SPM of the whole brain positive effect of condition contrast (overall performance of WM task) was used to define ROIs as this contrast was orthogonal to the contrasts of interest (pre vs post intervention and structured vs random trials). Regions of interest were defined using the MarsBar toolbox in SPM8, and estimated beta values were extracted, winsorised, (replacing any values mean +/- 2·5 x SD with that value), and analysed in SPSS using a repeated measures ANOVA. All fMRI data were processed and analysed using SPM8 software 2011.

As the study included an fMRI paradigm, the sample size was calculated from previous studies using a similar paradigm which produced significant results in healthy controls with group sizes of n = 14, producing effect sizes of

0·9 and 1·7 [13, 15]. Recent cognitive training studies have yielded significant results in controls with group sizes of n = 8 producing an effect size of 1·75 [17]. Based on these studies, power calculations gave 80% power to detect a significant difference ($p<0·05$) with group sizes of > 12.

Additional references

1. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. 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*. 1984 Jul;**34**(7):939-44.

2. World Health Organization. The ICD-10 classification of mental and behavioural disorders : clinical descriptions and diagnostic guidelines. Geneva: World Health Organization; 1992.

3. 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 Nov;**12**(3):189-98.

4. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry*. 1984 Nov;**141**(11):1356-64.

5. Wechsler D. Wechsler Memory Scale- Third Edition. San Antonio, Texas: The Psychological Corporation; 1997.

6. Gould RL, Brown RG, Owen AM, Bullmore ET, Williams SC, Howard RJ. Functional neuroanatomy of successful paired associate learning in Alzheimer's disease. *Am J Psychiatry*. 2005 Nov;**162**(11):2049-60.

7. Lezak MD, Howieson DB, Bigler ED, Tranel D. Neuropsychological assessment (5th ed.). New York: Oxford University Press; 2012.

8. Baddeley AD. A three-minute reasoning test based on grammatical transformation. *Psychometric science*. 1968;**10**:341-2.

9. Hampshire A, Highfield RR, Parkin BL, Owen AM. Fractionating human intelligence. *Neuron*. 2012 Dec 20;**76**(6):1225-37.

10. Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. *Percept Mot Skills*. 1958;**8**:271-6.

11. Ashendorf L, Jefferson AL, O'Connor MK, Chaisson C, Green RC, Stern RA. Trail Making Test errors in normal aging, mild cognitive impairment, and dementia. *Arch Clin Neuropsychol*. 2008 Mar;**23**(2):129-37.

12. IBM. Statistical Package for Social Sciences (SPSS) version 22.0. 22.0 ed2013.

13. Bor D, Cumming N, Scott CE, Owen AM. Prefrontal cortical involvement in verbal encoding strategies. *The European journal of neuroscience*. 2004 Jun;**19**(12):3365-70.

14. Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage*. 2007 Oct 15;**38**(1):95-113.

15. Bor D, Duncan J, Wiseman RJ, Owen AM. Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron*. 2003 Jan 23;**37**(2):361-7.

16. Bor D, Owen AM. A common prefrontal-parietal network for mnemonic and mathematical recoding strategies within working memory. *Cereb Cortex*. 2007 Apr;**17**(4):778-86.

17. Olesen PJ, Westerberg H, Klingberg T. Increased prefrontal and parietal activity after training of working memory. *Nat Neurosci*. 2004 Jan;**7**(1):75-9.

Table DS1 Demographic and screening variables

	CONTROL (n = 15) Mean (SD)	TRAINING (n = 15) Mean (SD)	Sig (p)
AGE	80·13 (5·19)	79·40 (6·19)	0·728
MMSE	25·93 (2·09)	26·00 (2·30)	0·934
YRS ED	12·57 (2·82)*	12·33 (2·94)	0·832†
IQ	115·63 (6·78)	117·14 (6·80)	0·548
GDS	3·73 (2·25)	4·33 (1·99)	0·433†
GENDER	6 F 9 M	6 F 9 M	1·000
MEDS	12	11	0·679
LENGTH	419.7 (591.8)	545.3 (513.4)*	0.548

Abbreviations: MMSE= Mini mental state examination, YRS ED= years of education, GDS = Geriatric Depression scale, M=male, MEDS= participant taking prescribed antidementia medication (cholinesterase inhibitors or memantine). LENGTH= length of illness, measured in days from date of diagnosis to inclusion in study. *n=14. †Mann-Whitney U and Wilcoxon W Tests, due to non parametric data.

Fig. DS1 Examples of span trial types. A) Structured trial B) Random trial for both verbal and spatial span tasks:

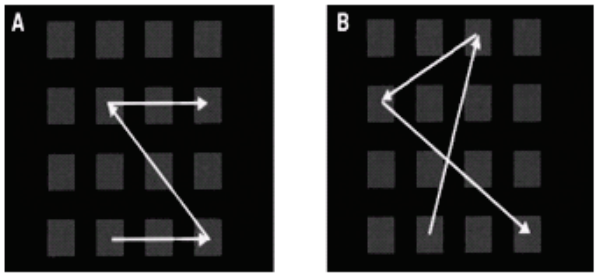
Verbal Span Sequences	Spatial Span Sequences
A) 2 4 6 9 7 5 B) 8 1 6 2 9 4	

Fig. DS2 Flow chart of recruitment.

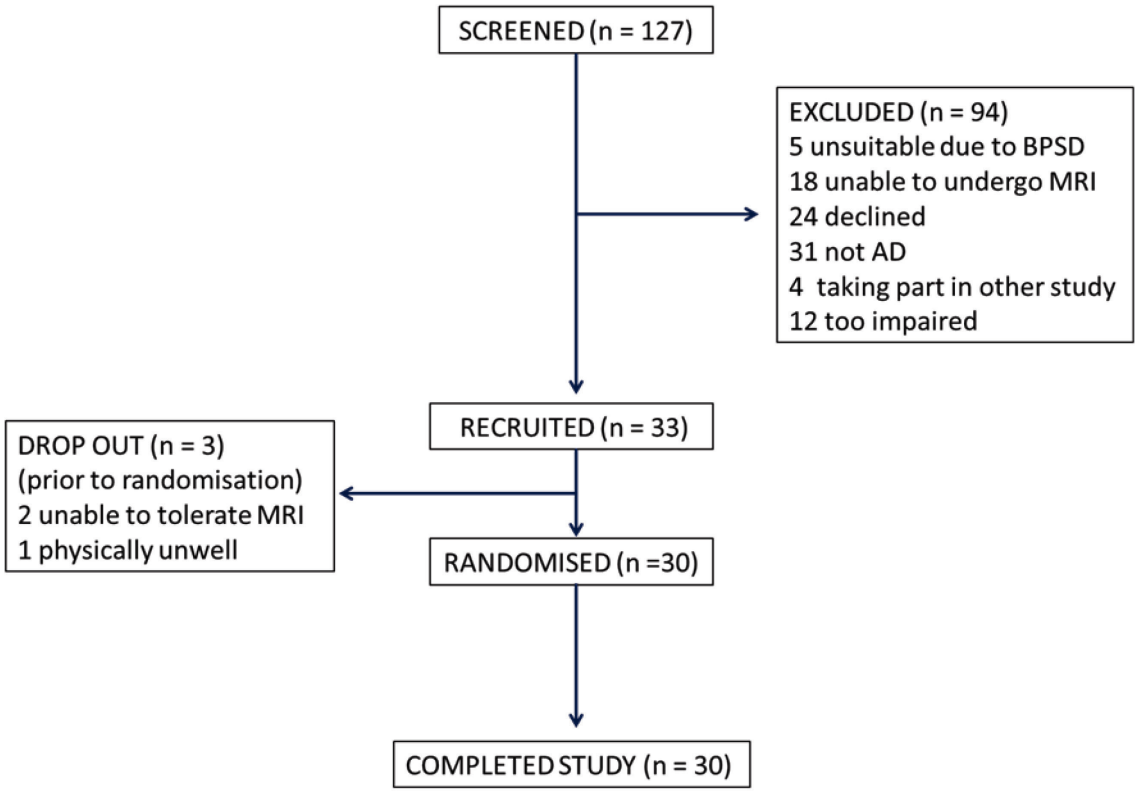
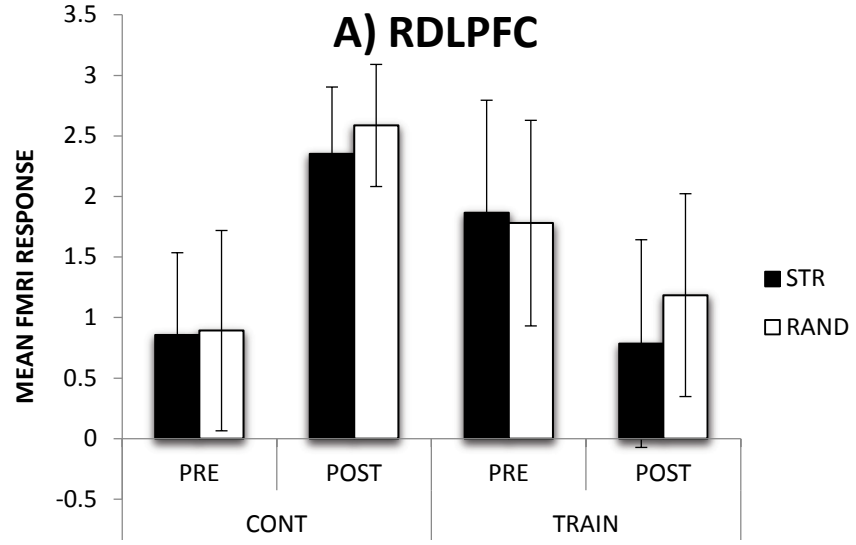


Fig. DS3 Mean fMRI response (parameter estimates) for A) Right dorsolateral prefrontal cortex B) Left dorsolateral prefromtal cortex regions of interest. CONT= control group, TRAIN = training group, PRE= pre intervention, POST = post intervention. Black bars = structured trials, White bars= random trails, Error bars are SEM



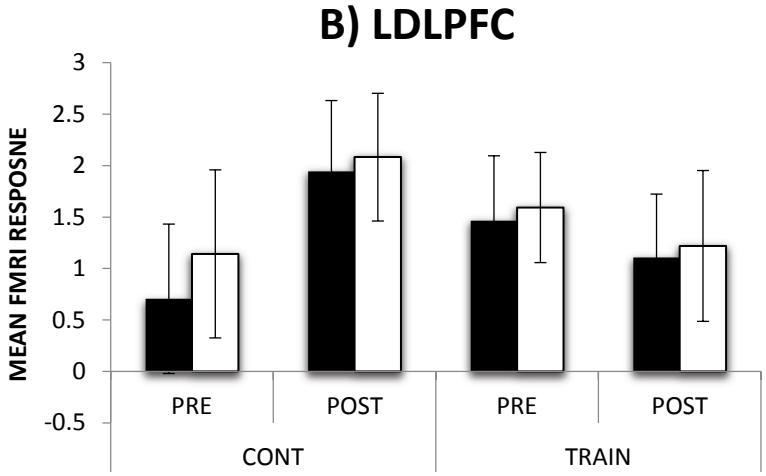


Fig. DS4 Mean fMRI response (beta values) for A) Left parietal cortex B) Right parietal cortex regions of interest. CONT= control group, TRAIN = training group, PRE= pre intervention, POST = post intervention. Black bars = structured trials, White bars= random trails, Error bars are SEM