**Supplementary Materials**

**Title:** Increased Working Memory-Related Brain Activity in Middle-Aged Women with Cognitive Complaints

Dumas, Kutz, McDonald, Naylor, Pfaff, Saykin, & Newhouse

## Method

### Participants

Participants were recruited through notices and advertisements in local newspapers and direct mailings that advertised a research study about memory and menopause for healthy women. The sample of women described below was a community sample and was not biased towards finding women with clinical levels of memory complaints. Only after the screening visit were women sorted into the complainer and non-complainer groups.

Participants were required to be postmenopausal, without menses for a minimum of one year and without surgically induced menopause. Exclusion criteria included current smoking, a history of breast cancer, and use of hormone therapy during the past year. As the larger study that all participants in the current study were a part of involved estradiol treatment and anticholinergic drug challenges the following exclusion criteria were used to ensure subject safety. Medical exclusion criteria included contraindications for hormone therapy, estrogen-dependent neoplasia, untreated hypertension greater than 160/100, history of deep vein thrombosis or other thromboembolic disease, hepatoma, severe migraines or stroke on oral contraceptives, current use of barbiturates, rifampin, insulin, carbamezepine, oral hypoglycemics, or antidepressants, known intolerance to conjugated estrogens, diabetes, untreated thyroid disease, clinical osteoporosis, and a history or presence of severe menopausal autonomic symptoms. Exclusion criteria for MRI scanning included claustrophobia, cardiac pacemakers, other implanted metal devices, injuries to the eye involving metal, tattoos on the head or neck, and other moveable metal implants in the body. In addition, we also excluded women with a history of the following: heavy alcohol (more than an average of one drink per day) or coffee use (more than three cups per day), significant cardiovascular disease, asthma, active peptic ulcer, hyperthyroidism, pyloric stenosis, narrow angle glaucoma, epilepsy, or current Axis I psychiatric disorders. The alcohol criterion was used to ensure participants were not alcohol dependent, and the caffeine criterion was used to ensure participants would not experience caffeine withdrawal on testing days.

Upon meeting these criteria, participants were approved for further screening at the University of Vermont (UVM) General Clinical Research Center (GCRC). After signing informed consent documents, participants provided a medical history and underwent a physical exam and laboratory tests assessing hematopoietic, renal, hepatic and hormonal function. Participants were cognitively evaluated using the Mini Mental State Exam (MMSE; (Folstein et al., 1975), Brief Cognitive Rating Scale (Reisberg and Ferris, 1988), and the Mattis Dementia Rating Scale (DRS, (Jurica et al., 2001) to establish a Global Deterioration Scale score (GDS) which rated the degree of cognitive impairment (Reisberg et al., 1993). Participants were required to have an MMSE score greater than or equal to 27, a DRS score greater than or equal to 123, and a GDS score of 1 or 2.

Behavioral screening consisted of a partial Structured Clinical Interview for DSM-IV-TR (SCID; (First et al., 2001) to establish the presence/absence of current or past major depressive disorder (MDD), current mania, or current dysthymia. In addition, participants completed the Beck Depression Inventory (BDI-II). A cut off score of 10 was used for the BDI; participants scoring over this threshold were discontinued from further participation. Subjects also completed a Menopause Symptom Checklist that was modified after the Sherwin Menopause Index (Sherwin 1991).

All subjects completed a set of questionnaires after Saykin et al. (2006) to establish a cognitive complaint index (CCI). This battery included the Memory Self-Rating Questionnaire (Squire et al., 1979), Neurobehavioral Function and Activities of Daily Living Rating Scale (Saykin, 1992), Informant Questionnaire on Cognitive Decline in the Elderly (Jorm et al., 1994), four cognitive items from the Geriatric Depression Scale (Yesavage et al., 1982), 10 cognitive items from a telephone-based screening for MCI (Rabin et al., 2007), and 23 items from the Memory Assessment Questionnaire adapted in part from the Functional Activities Questionnaire (Pfeffer et al., 1982). Subjects were categorized as cognitive complainers (CC) if they endorsed more than 20% of the items in the CCI, based on our prior work demonstrating that those clinically characterized as CC tended to endorse >20% of these items (e,g., Saykin et al. 2006).

**Cognitive Testing: Working Memory and Episodic Memory**

After passing the behavioral and medical screening, participants were scheduled for a baseline visit for the larger study that included an MRI scan with structural and functional imaging and an extensive cognitive battery that included an episodic memory measure. The main fMRI task of interest was a visually presented verbal N-back sequential letter task used to probe working memory circuitry. Participants saw a string of consonants (except L, W, and Y), presented in upper case letters, one every 3 seconds. Four conditions were presented: 0-back, 1-back, 2-back, and 3-back. The 0-back control condition had a minimal working memory load; participants were asked to decide if the current letter matched a single target letter that was specified before the epoch began. In the 1-back condition, participants were asked to decide if the current letter matched the previous letter. During the 2-back condition, participants decided whether the current letter matched the letter that had been presented two back in the sequence. The most difficult 3-back condition followed the same pattern. The 0-, 1-, 2-, and 3-back conditions were repeated three times in a counterbalanced order such that the same condition was not repeated two times in a row. In this block design task, participants responded to nine items in each 27-second block. A rest break followed with a plus sign (+) fixation for 12 seconds. The total time of the task was 8 minutes 12 seconds. Participants practiced the task before the scanning session in a mock MRI scanner so that they were familiar with the instructions for each condition as well as the MRI environment for completing the N-back task.

Participants responded to all items by button press through an MRI compatible fiber optic button response system (Eloquence System, Invivo Corp., Gainesville, FL) to indicate whether the item matched the target condition. Stimuli were delivered through an MR-safe computer monitor. Experimental tasks were programmed using the E-prime software package. The tasks were presented by PC interface which recorded participant responses and reaction times.

**fMRI Scan Procedure and Preprocessing**

The MRI procedures were as follows. Participants were scanned on a Philips 3T Achieva MRI scanner. All participants received the following MR sequences as part of the imaging protocol: (1) A sagittal T1-weighted spoiled gradient volumetric sequence oriented perpendicular to the anterior commissure (AC)-posterior commissure (PC) plane using a repetition time (TR) of 9.9 ms, echo time (TE) of 4.6 ms, flip angle of 8 degrees, number signal averages (NSA) 1, field of view (FOV) of 256 mm, 256 X 256 matrix, and 1 mm slice thickness with no gap for 140 contiguous slices. (2) An axial T2-weighted gradient spin echo (GRASE) sequence using the AC-PC line for slice positioning. Twenty-eight contiguous slices 5 mm thick and no gap were acquired using TR 2466 ms, TE 80 ms, NSA 3 and FOV of 230 mm. All images were reviewed by a board-certified neuroradiologist to exclude intracranial pathology. fMRI was performed using EpiBOLD (echoplanar blood oxygenation level dependent) imaging using a single-shot sequence (TR 2500 ms, TE 35 ms, flip angle 90 degrees, 1 NSA for 197 volumes). Resolution was 2.5 mm x 2.8 mm x 5 mm. Thirty contiguous slices 5 mm thick with no gap were obtained in the axial oblique plane parallel to the AC-PC plane using a FOV of 240 mm and a matrix size of 128 x 96. Field map correction for magnetic inhomogeneities was accomplished by acquiring images with offset TE at the end of the functional series.

Preprocessing and random effects analyses of the functional data were performed with Brain Voyager QX software (Brain Innovation, Maastricht, The Netherlands). Before the analyses were completed the following preprocessing steps were performed. Three-dimensional motion correction to correct for small head movements was completed by aligning of all volumes to the first volume. Estimated translation and rotation movements did not exceed 2 mm for any subject. Further data preprocessing comprised of linear trend removal and filters for spatial (4 mm full width at half maximum isotropic Gaussian kernel) as well as temporal (high pass filter: 1 cycle/run) smoothing to remove aliased signal correlated with background respiration and heart rate. Anatomical and functional images were then co-registered and normalized to Talairach space. Statistical analysis was performed by multiple linear regression of the signal time course in each voxel. The expected BOLD signal change for each condition within a run was modeled by a canonical hemodynamic response function.

After completion of the MRI session, participants completed a cognitive battery including tests of arousal, attention, and episodic memory. The data from the episodic memory measure are presented below and the other measures will be presented elsewhere then the larger study is completed. The measure of episodic memory was the Buschke Selective Reminding Task (SRT; Buschke 1973). In the SRT, participants were read a list of 16 unrelated words balanced for word frequency and imagery, followed by an immediate recall trial. On subsequent trials (up to eight), subjects were immediately reminded of words they had failed to recall on the prior trial. Twenty minutes after the completion of the SRT after a filled retention interval when participants completed other non-verbal tasks, they were then asked to recall all 16 words again. Four measures were obtained from the SRT task: the total number of words recalled across all lists for the immediate recall conditions, recall consistency, recall failure, and delayed recall.

**fMRI Analyses**

Statistical analyses were performed using a 2 (Group: CC versus NC) X 4 (Working Memory Load: 0-, 1-, 2-, 3-back) random effects ANOVA using standard ANOVA procedures in Brain Voyager. First, we constructed contrast maps in which the two groups differed across the parametric variation in working memory load. We hypothesized that group differences in activation would increase as the working memory load increased; thus the design matrix included all N-back conditions for both participant groups. The contrast vector for the overall interaction was as follows: -6, +1, +2, +3 for the 0-, 1-, 2- and 3-back conditions in the CC group and +6, -1, -2, -3 for the 0-, 1-, 2-, 3-back conditions for the NC group. The hemodynamic response function was accounted for in these models. To probe the basis for the interaction between group and working memory load, we examined group differences in the following comparisons: 3-back minus 0-back, 2-back minus 0-back, and 1-back minus 0-back conditions in order to examine group differences at each working memory load while controlling for the visual presentation of the letters and the motor responses.

**Cognitive, Behavioral, and Mood Analyses**

Statistical analyses were performed using independent samples t-tests to compare performance, behavioral response, and mood measures for the CC and NC groups.

Working memory performance on the N-back task was analyzed using the signal detection measures of sensitivity (*d'*) and bias (*C*; (Snodgrass and Corwin, 1988). Sensitivity is a measure of how different two classes of items are as measured by *d'* and is represented in standard deviation units. Larger *d'*s represent greater sensitivity and greater accuracy. In the N-back task, the two classes of items are matches and mismatches for each of the working memory load conditions. Bias (*C*) is the tendency for a participant to endorse a letter as a match or mismatch and is also represented in standard deviation units. Liberal response bias indicates that a participant calls a large number of responses matches whereas conservative bias indicates that the participant makes many mismatch responses. Bias scores of less than 0 are liberal while bias scores greater than 0 are conservative.

## Results

### Activation Data

First, we examined brain activation patterns related to working memory maintenance and updating during the N-back task performance to establish the general N-back brain activation pattern across all subjects. We examined the parametric increase in activation from 0-back through the 1-, 2- and 3-back conditions. Overall, there was bilateral frontal, parietal, insular, and cerebellar activation as well as anterior cingulate activation that increased as the working memory load increased (See Supplementary Figure 1) as has been described in prior studies using the N-back task (Cohen et al. 1997). Decreased activity was seen in medial frontal and medial parietal regions. In addition, decreased activity was seen in the bilateral tail of the caudate as working memory load increased.

Our primary analysis of interest was the interaction between group and working memory load and is described in the main paper. This analysis showed increased activation for the CC compared to the NC group as a function of increasing working memory load in the left middle frontal gyrus (BA 10), right middle frontal gyrus (BA 9), right cingulate gyrus (BA 24), left cingulate gyrus (BA 32), right and left insula (BA 13), and right precuneus (BA 7; See Supplementary Table 12). Decreased activation was seen in the right tail of the caudate across working memory load for the CC relative to the NC group.

To probe this interaction and further understand how increasing working memory load interacted with subjective cognitive complaint status, we examined group differences at each of the working memory load conditions minus the 0-back match condition.

First, for the 3-back relative to the 0-back condition, greater activation was seen in regions similar to the activation observed for the overall interaction described above for the CC compared to the NC group (See Supplementary Table 1). Specifically, increased activation for CC compared to NC was seen in the left middle frontal gyrus (BA 10), right middle frontal gyrus (BA 9), right anterior cingulate (BA 32), right and left insula (BA 13), right precuneus (BA 7), and right superior temporal gyrus (BA 38). Second, similar regions also showed increased activation for the CC group compared to the NC group for the 2-back minus 0-back condition in the right middle frontal gyrus (BA 9), right medial frontal gyrus (BA 10), right anterior cingulate gyrus (BA 24), left anterior cingulate gyrus (BA 32), right insula (BA 13), right precuneus (BA 7), and right middle temporal gyrus (BA 21). Greater activation was seen for the NC group relative to the CC group in the right tail of the caudate (See Supplementary Table 1.). Finally, for 1-back compared to the 0-back condition no activation differences were seen between the two groups.

**Working Memory Performance**

We analyzed the performance on the N-back task during the MRI session with a 2 (group: CC versus NC) X 4 (working memory load: 0-back, 1-back, 2-back, 3-back) mixed model ANOVA for the *d',* percent correct, and *C* measures. Group was a between-subjects factor and working memory load was a within-subjects factor. As detailed below, there were no differences between the CC and NC groups for the *d*', percent correct, or *C* measures.

Analysis of *d'* showed a main effect of working memory load ($F(3,63) = 29.74$, $p < .001$; see Supplementary Table 2) where sensitivity declined as the working memory load increased. Sensitivity on the 3- and 2-back conditions was less than sensitivity on the 1- and 0-back conditions (smallest $t(22) = 5.20$; largest $p < .001$). Sensitivity for the 2- and 3-back conditions was not different ($t(22) = -.28$, $p = .78$). There was a trend for a difference in sensitivity for the 0- and 1-back conditions ($t(22) = 2.04$, $p = .054$). Overall the data patterns for the percent correct measure were the same as *d'*.

There was also a main effect of working memory load for the bias measure *C* ($F(3,63) = 3.43$, $p = .02$) that showed that subjects were more conservative on the 3-back condition relative to the 2-back condition ($t(22) = -2.57$, $p = .02$). There were no other significant differences between working memory load conditions for *C* (largest $t(22) = -1.94$; smallest $p = .07$; see Supplementary Table 2).

**Correlations between Activation, Performance and BDI score**

To further examine the underlying basis for the group differences in activation across working memory load we examined correlations between the beta values from the significant clusters generated from the overall ANOVA of group and each of the working memory load conditions minus the 0-back with performance and BDI scores. No significant correlations were found between any of the significant clusters and performance on the N-back task or scores on the BDI.

**References**

First MB, R.L. S, Gibbon M, Williams JBW (2001) Structured Clinical Interview for DSM-IV-TR Axis I Disorders-Patient Edition. Washington, D.C.: American Psychiatric Press Inc.

Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research 12:189-198.

Jorm AF, Christensen H, Henderson AS, Korten AE, Mackinnon AJ, Scott R (1994) Complaints of cognitive decline in the elderly: a comparison of reports by subjects and informants in a community survey. Psychological Medicine 24:365-374.

Jurica PJ, Leitten CL, Mattis S (2001) In: Dementia Rating Scale-2 Lutz, FL: Psychological Assessment Resources Inc.

Pfeffer RI, Kurosaki TT, Harrah CH, Chance JM, Filos S (1982) Measurement of functional activities in older adults in the community. Journal of Gerontology 37:323-329.

Rabin LA, Saykin AJ, Wishart HA, Nutter-Upham KE, Flashman LA, Pare N, Santulli RB (2007) The Memory and Aging Telephone Screen: development and preliminary validation. Alzheimers Dementia 3:109-121.

Reisberg B, Ferris SH (1988) Brief cognitive rating scale (BCRS). Psychopharmacol Bull 24:629-635.

Reisberg B, Ferris SH, de Leon MJ, Crook T (1993) The Global Deterioration Scale for assessment of primary degenerative dementia. American Journal of Psychiatry 139:1136-1139.

Saykin AJ (1992) Neurobehavioral function and activities of daily living rating scale (NBFADL-63 item version). Hanover, NH: Dartmouth Medical School.

Snodgrass J, Corwin J (1988) Pragmatics of measuring recognition memory: applications to dementia and amnesia. Journal of Experimental Psychology 117:34-50.

Squire LR, Wetzel CD, Slater PC (1979) Memory complaint after electroconvulsive therapy: assessment with a new self-rating instrument. Biological Psychiatry 14:791-801.

Yesavage JA, Brink TL, Lose TL (1982) Development and validation of Geriatric Depression Rating Scale: a preliminary report. Journal of Psychiatric Research 17:37-49.

Supplementary Table 1. Differences in activation between the cognitive complainer (CC) and non-complainer (NC) groups including Talairach coordinates, cluster size, region descriptions (Brodmann's areas, BA), *t* values and uncorrected voxel-level *p* values.

| Contrast | Coordinates<br>x y z | Cluster Extent | Region Description | *t* value | *p* value |
|---|---|---|---|---|---|
| CC – NC | | | | | |
| Increasing WM load | 33 24 27 | 1539 | Right middle frontal gyrus (BA 9) | 3.95 | 0.0009 |
| | 7 -31 44 | 624 | Left middle frontal gyrus (BA 10) | 3.28 | 0.0020 |
| | 2 -58 34 | 923 | Right precuneus (BA 7) | 3.25 | 0.0008 |
| | 20 6 37 | 1983 | Right cingulate gyrus (BA 24) | 3.75 | 0.0008 |
| | 5 1 16 | 5559 | Left cingulate gyrus (BA 32) | -3.24 | 0.0022 |
| | 35 2 16 | 692 | Right insula (BA 13) | 3.47 | 0.0013 |
| | -41 15 17 | 559 | Left insula (BA 13) | 3.46 | 0.0015 |
| NC – CC | | | | | |
| Increasing WM load | 22 39 13 | 397 | Right caudate tail | 3.49 | 0.0015 |
| CC – NC | | | | | |
| 3 back – 0 back | 32 25 27 | 1588 | Right middle frontal gyrus (BA 9) | 4.07 | 0.0007 |
| | -33 41 8 | 1354 | Left middle frontal gyrus (BA 10) | 3.20 | 0.0024 |
| | 1 -57 32 | 563 | Right precuneus (BA 7) | 3.67 | 0.0010 |
| | 1 17 30 | 5434 | Right cingulate gyrus (BA 32) | 3.33 | 0.0018 |
| | 35 2 16 | 510 | Right insula (BA 13) | 3.34 | 0.0017 |
| | -33 6 25 | 443 | Left insula (BA 13) | 3.59 | 0.0012 |
| | -43 2 -12 | 992 | Left superior temporal gyrus (BA 38) | 3.40 | 0.0016 |

| | | | | | |
|---|---|---|---|---|---|
| CC – NC | | | | | |
| 2 back – 0 back | 33 24 28 | 636 | Right middle frontal gyrus (BA 9) | 3.51 | 0.0014 |
| | 20 42 -6 | 382 | Right medial frontal gyrus (BA 10) | 3.38 | 0.0017 |
| | 4 -58 34 | 793 | Right precuneus (BA 7) | 3.34 | 0.0018 |
| | -1 15 35 | 3707 | Left cingulate gyrus (BA 32) | 3.49 | 0.0012 |
| | 22 2 38 | 1147 | Right cingulate gyrus (BA 24) | 3.36 | 0.0018 |
| | 33 2 17 | 426 | Right insula (BA 13) | 3.26 | 0.0021 |
| | 49 -31 -6 | 418 | Right middle temporal gyrus (BA 21) | 3.47 | 0.0016 |

Supplementary Table 2. Sensitivity (*d'*) and bias (C) (standard errors) on the 0-, 1-, 2- and 3-back conditions for the participant groups (CC and NC). Working memory load was significant for all three measures (*ps* < .05).

| | | CC<br>N=12 | NC<br>N=11 |
|---|---|---|---|
| 0-back | | | |
| | *d'* | 2.83 (.44) | 3.02 (0.47) |
| | *C* | 0.19 (0.20) | 0.24 (0.14) |
| 1-back | | | |
| | *d'* | 2.59 (0.70) | 2.73 (0.40) |
| | *C* | 0.19 (0.21) | 0.11 (0.17) |
| 2-back | | | |
| | *d'* | 1.63 (0.66) | 1.74 (0.72) |
| | *C* | 0.15 (0.27) | 0.06 (0.24) |
| 3-back | | | |
| | *d'* | 1.85 (0.61) | 1.62 (0.56) |
| | *C* | 0.28 (0.32) | 0.35 (0.54) |

Figure Legend

Supplementary Figure 1. Activation map for all subjects across the parametrically increasing working memory load on the N-back task from 0-back to 3-back ($p < .005$). Orange colors represent activation in brain areas that increased as working memory load increased. Blue colors represent activation in brain areas that decreased as working memory load increased.

**Supplementary Figure 1.**