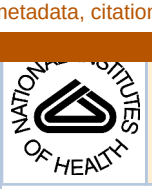

Published in final edited form as:
Clin Neurosci Res. 2007 November ; 6(6): 381–390.

Neural network approaches and their reproducibility in the study of verbal working memory and Alzheimer's disease

Christian Habeck[1] and **Yaakov Stern**[1,2,3,4]

1*Cognitive Neuroscience Division of the Taub Institute for Research in Alzheimer's disease and the Aging Brain, 622 West 168th Street, PH-18, New York, New York.*

2*Department of Neurology, College of Physicians and Surgeons of Columbia University, 630 West 168th Street, New York, New York.*

3*Department of Psychiatry, College of Physicians and Surgeons of Columbia University, 630 West 168th Street, New York, New York.*

4*Department of Psychology, College of Physicians and Surgeons of Columbia University, 630 West 168th Street, New York, New York.*

Abstract

As clinical and cognitive neurosciences mature, the need for sophisticated neuroimaging analysis becomes more apparent. Multivariate analysis techniques have recently received increasing attention because they have attractive features that cannot be easily realized by the more commonly used univariate, voxel-wise, techniques. Multivariate approaches evaluate correlation/covariance of activation across brain regions, rather than proceeding on a voxel-by-voxel basis. Thus, their results can be more easily interpreted as a signature of neural networks. Univariate approaches, in contrast, cannot directly address functional connectivity in the brain. Apart from this conceptual difference, the covariance approach can also result in greater statistical power when compared with univariate techniques, which are forced to employ very stringent, and often overly conservative, corrections for voxel-wise multiple comparisons. Multivariate techniques also lend themselves much better to prospective application of results from the analysis of one dataset to entirely new datasets. We provide two examples that illustrate different uses of multivariate techniques in cognitive and clinical neuroscience. We hope this contribution helps facilitate wider dissemination of these techniques in the research community.

Keywords

covariance analysis; functional imaging; neural networks; visual recognition memory; Alzheimer's disease

INTRODUCTION

Brain imaging in the clinical and basic neuroscience has recently paid more attention to multivariate analytic techniques that aim to uncover the operation of neural networks (for an

Correspondence should be addressed to Christian Habeck, 630 West 168th Street, P&S Box 16, New York, NY, 10032. Email: ch629@columbia.edu; Telephone: 212-305-0945; Fax: 212-342-1838.

incomplete review of landmark articles see (1–4)). Despite widespread interest in functional connectivity and the interaction between different brain areas, it is somewhat surprising that multivariate approaches have not yet quite enjoyed the success and the widespread usage of massively univariate techniques. The reasons for this deficiency mainly pertain to the increased demands in mathematical and conceptual literacy on the user and the absence of a canonical software package similar to Statistical Parametric Mapping (SPM). More than univariate techniques, multivariate techniques take into account the total variance structure of the data and are likely to reflect activity due to the experimental manipulation or disease process of interest as well as processes not under control of the experimenter. The major sources of variance might not necessarily break down in manner that lends itself to easy interpretation in terms of experimental or clinical design variables. The user has to learn to appreciate this when it happens, not as a reason for giving up and dismissing covariance analysis as useless, but as an opportunity to understand the data set better and possibly inform the design of an updated experiment in the future.

Despite these difficulties, it is worth stressing that multivariate techniques have many attractive features that cannot be easily realized by voxel-wise techniques. In particular, they evaluate correlation of activation across brain regions, rather than proceeding on a voxel-by-voxel basis. Thus, their results can be more easily interpreted as a signature of neural networks. Univariate approaches, on the other hand, cannot directly address functional connectivity in the brain. The multivariate approach can also result in greater statistical power when compared with univariate techniques, which are forced to employ very stringent, and often overly conservative, corrections for voxel-wise multiple comparisons. Multivariate techniques also lend themselves much better to prospective application of results from the analysis of one dataset to entirely new datasets, thereby aiding the test of reproducibility of results - a topic that, likewise, has not received sufficient attention in neuroimaging research so far.

This paper provides a hands-on account of some of these features. Results that were previously obtained from an experimental study of verbal working memory with fMRI (7) and from a cross-sectional study of Alzheimer's disease with $H_2{}^{15}O$-PET (5) are applied prospectively to a replication data set. Furthermore, a comparison with a massively univariate analysis is undertaken. These themes should prove instructive for both novices and veterans of the multivariate approach.

METHODS

PCA primer

We give a brief mathematical sketch for people unfamiliar with covariance analysis and the idea of dimensionality reduction. Without employing matrix algebra, we denote the data array to be analyzed as $Y(s, \mathbf{x})$, with the index s symbolizing tasks and subjects, while $\mathbf{x}$ stands for the voxel location in the brain.

The goal of any multivariate decomposition is to separate task/subject from voxel indices in the following series:

$$Y(s, \mathbf{x}) = SSF_1(s)v_1(\mathbf{x}) + SSF_2(s)v_2(\mathbf{x}) + SSF_3(s)v_3(\mathbf{x}) + \ldots$$

The data array has been decomposed into a series of terms of diminishing variance contribution. The terms are made up of task- and subject-dependent subject scaling factors, SSF_i, and voxel-dependent principal component topographies, i.e. brain images, v_i. Usually only the first few (typically < a third of the number of subjects) terms are retained to represent the data. For instance, if we decide (by some criterion that is not talked about here) that we want to retain the first two terms, we have two principal component images, $v_1(\mathbf{x})$ and $v_2(\mathbf{x})$, which are fixed across subjects and tasks, and two sets of subject scaling factors, $SSF_1(s)$ and $SSF_2(s)$, that are

varying across subjects and tasks. These latter two sets of numbers can be used for further analysis and correlation with clinical and experimental variables. The goal of dimensionality reduction thus becomes obvious: rather than keeping track of subject- and task-dependent numbers at every voxel location, we only have two subject- and task-dependent variables to deal with.

Application to an fMRI study of verbal working memory (7)

The first demonstration of spatial covariance analysis concerns a study of verbal working memory in basic neuroscience (7). The goal of the study was the identification of a neural network whose subject expression levels correlated with memory load.

Subject information

Derivation sample—40 healthy young subjects, between the ages of 20 and 35 years (age = 25.02 ± 3.68 years), participated in an event-related functional magnetic resonance imaging (efMRI) paradigm of a delayed-match-to-sample (DMS) task. All subjects were right-handed and carefully screened to insure that they had no history of medical, psychiatric, neurological, or sleep disorders. All subjects passed substance abuse screening tests. Subjects were supervised at all times, and polysomnographic monitoring confirmed that they remained awake during the sleep deprivation period.

Informed consent, as approved by the Internal Review Board of the College of Physicians and Surgeons of Columbia University, was obtained prior to study participation and after the nature and risks of the study were explained. Subjects were paid for their participation in the study.

Replication sample—Eighteen healthy elderly subjects (age = 74.13 ± 7.27 years) underwent the same behavioral and scanning protocol as the young subjects.

Delayed-match-to-sample task

The Delayed-Match-To-Sample task was a variant of the Sternberg task (8,9). The sequence of trial events was the following (figure 1): first, during the stimulus period of the task, an array of one, three or six capital letters were presented for three seconds ("stimulus phase"). With the offset of the visual stimulus, subjects were instructed to focus on the blank screen and hold the stimulus items in mind for a seven-second retention interval ("retention phase"). Finally, a probe appeared for three seconds ("probe phase"), which was a lowercase letter centered in the field of view. In response to the probe, subjects indicated by a button press whether or not the probe matched a letter in study array. (Left index finger indicated "Yes", right index finger indicated "No".)

Details about the event-related design of the task have been described in detail elsewhere (6, 7), and are thus not presented again in the current account.

fMRI Acquisition and Processing

Functional images were acquired using a 1.5 Tesla magnetic resonance scanner (General Electric) retrofitted for echoplanar imaging. A gradient echo EPI sequence [TE=50 ms; TR=3 sec; flip angle = 90°] and a standard quadrature head coil was used to acquire T_2* weighted images with an in-plane resolution of 3.124 mm×3.124 mm (64×64 matrix; 20 cm² field of view). Based on T_1 “scout” images, 8 mm transaxial slices (15–17) were acquired. Following the fMRI runs, a high (in-plane) resolution T_2 image at the same slice locations used in the fMRI run was acquired using a fast spin echo sequence [TE=105 ms; TR=2500 ms; 256 × 256 matrix; 20 cm² field of view].

All image processing and analysis was done using the SPM99 program (Wellcome Department of Cognitive Neurology, London, UK) and other code written in Matlab 5.3 (Mathworks, Natick, MA). FMRI time series were corrected for order of slice acquisition. All functional volumes in a given subject were realigned to the first volume from the first run of each study. The T_2 anatomical image was then co-registered to the first functional volume, using the mutual information co-registration algorithm implemented in SPM99. This co-registered structural image was then used in determining non-linear spatial normalization ($7 \times 8 \times 7$ nonlinear basis functions) parameters for a transformation into a Talairach standard space defined by the Montreal Neurological Institute template brain applied with SPM99. These normalization parameters were then applied to the functional data (using SINC-interpolation to reslice the images to 2mm × 2mm × 2mm).

First-level data analysis

The fMRI responses to the three separate temporal components of the task, in each of three set-size conditions, were fit to separate sets of predictor variables (10). This was done in the same fashion irrespective of the number of items in the stimulus array. A constant intercept which was convolved with a canonical hemodynamic response waveform (a sum of two gamma functions, as specified in the SPM99 program) was used as a predictor variable for the fMRI time series in stimulus and probe phases, whereas a first-order Fourier basis set convolved with the canonical hemodynamic response function was used during the retention phase (11). The band-pass filtered (low pass by a Gaussian with a FWHM of 4 sec and a high pass cutoff of 14.5 mHz) fMRI time series at each voxel were regressed onto these predictor variables. A first order autoregressive autocorrelation model was fit to the residuals to make statistical inference more robust to the intrinsic temporal autocorrelation structure (12).

At every voxel in the image, user-defined contrasts assessed the amplitudes (normalized regression coefficients) of the components of the event-related responses that matched the canonical hemodynamic response waveform. This method of time-series modeling and contrast estimation at each voxel reduces the number of images to one per subject per condition, and thus makes a multivariate analysis feasible. To account for gain differences between fMRI sessions, activation values were normalized by their voxel averages. The resulting images were smoothed using an isotropic Gaussian kernel (FWHM = 8 mm) and used as the data in the subsequent analysis.

Multivariate analysis

Ordinal Trend Canonical Variates Analysis (OrT CVA) (4) was performed on the data, using the first-level contrast maps, i.e. one per subject per memory-load level. This analysis contains the core step of applying principal components analysis (PCA) to the data matrix that has been transformed using a matrix representing the experimental design (2,3). OrT CVA was designed to identify a covariance pattern in MR signal across all voxels, the expression of which increases as many subjects as possible across memory load. The property of consistent change of pattern expression across task conditions on a subject-by-subject, rather than just on a subject mean, basis is called an "ordinal trend". For our inferential statistics, a permutation test was performed in which the task assignment was broken within-subjects and the complete analysis was performed on the permuted data set 500 times. The test statistic for the permutation test was the repeated-measures F-statistic ('rm-F') of the activation pattern's subject expression across memory-load levels(1/3/6). P-levels can thus be attached to the point-estimate value of rm-F observed in the unpermuted subject sample.

Activation patterns resulting from multivariate analysis assign different weights to all voxels included in the analysis, depending on the salience of their covariance contribution. Whether a voxel weight is reliably different from zero is assessed by a bootstrap estimation procedure

(13). Denoting the point estimate of a voxel weight as **w** and the standard deviation resulting from the bootstrap resampling procedure as $s_{\mathbf{w}}$, we can compute the inverse coefficient of variation, which should obey a standard normal distribution and can therefore interpreted as a Z-score, i.e.

$$Z = \mathbf{w}/s_{\mathbf{w}} \sim N(0, 1)$$

Sufficiently small variability of a voxel weight around its point estimate value in the resampling processes results in a Z value of large magnitude and indicates a reliable contribution to the covariance pattern. One-tailed p-levels of 0.01, 0.001, for instance, imply Z thresholds of 2.33, 3.09, respectively. It is important to realize that bootstrap resampling is intended to give a visual depiction of the covariance pattern, similar to voxel-wise t-test in SPM. However, for the computation of subject-scaling factors the use of the actual point estimate of the covariance pattern is important, as *all* regions contribute, even if they were not deemed reliable in the bootstrap test. The Z-map obtained from the bootstrap test is *never* used for any purpose other than visual depiction in a brain image.

Once an activation pattern was identified that systematically increased in expression as a function of memory load, we examined the correlation between individual subjects' change in network expression from the lowest (1 letter) to the highest (6 letters) memory load and change in their scores on the task performance measures. Further, this activation pattern can also be forward-applied to any other scan, regardless whether it came from the derivation data set or not, according to the inner product rule. Let $\mathbf{a}_i$ be the activation pattern to be applied and $\mathbf{s}_i$ stands for the target scan, both with the voxel index i. The level of expression can then be obtained as the sum of the products $\mathbf{a}_i \times \mathbf{s}_i$ over all voxel positions.

Univariate analysis

We also performed massively univariate analysis, again using the first-level contrast maps as before. First we tried to mimic the OrT CVA purpose, by performing a voxel-wise mean-contrast analysis, i.e. seeing whether in the derivation sample we could find a significant mean trend in the maintenance data across memory-load levels at the uncorrected p-level of 0.001. We also performed brain-behavioral correlations in a voxel-wise manner, i.e. we tested which voxels had activation values that correlated with a behavioral outcome measures. Again, we chose an uncorrected p-level of 0.001 in order not to make it too difficult for our univariate analysis to produce results.

RESULTS

Behavioral Performance

We give a brief summary of the behavioral-performance results, although they will not be used further for the purposes of this report. Outcomes measures were (within-subject) mean reaction times as well as (within-subject) coefficients of variation (CV) in reaction time.

Reaction times showed a load-dependent increase: RT(1)=847 ± 170 ms, RT(3)=1015 ± 202 ms, RT(6)=1148 ± 227 ms (H–F corrected rm-$F_{2,78}$=106.18, p<0.0001), variability in reaction time was biggest for the 3-item array: CV RT(1)=0.305 ± 0.117, CV RT(3)=0.348 ± 0.085 ms, CV RT(6)=0.299 ± 0.092 (H–F corrected rm-$F_{2,78}$=4.01, p<0.05). Recognition accuracies stayed constant across memory load: P(1)=98 ± 3 %, P(3)=97 ± 5 %, P(6)=97 ± 4 % (H–F corrected rm-F2,78=1.61, p=0.20). The fraction of non-response trials stayed constant and negligible too, NR(1)=1 ± 4 %, NR(3)=1 ± 2 %, NR(3)=1 ± 2 % (H–F corrected rm-$F_{2,78}$=0.78, p=0.39).

fMRI Data

OrT CVA was applied to all task phases and the results are discussed with interpretations in detail in the original report (7); for the current review, we concentrated on the maintenance phase only.

Our analysis produced significant results: a linear combination of the first 2 principal components produced by OrT CVA displayed ordinal trend properties (permutation test: rm-F=38.66, $p<0.001$), suggesting an unambiguous memory load-related neural correlate of rehearsal. The pattern accounted for 11% of the variance of the raw data (=untransformed by any design matrix). Brain regions that concomitantly decreased in activation (as ascertained by the bootstrap test) for the majority of subjects were found mainly in lateral PFC (BA 9,44), the parietal lobe (BA 7,40), anterior cingulate (BA 32), and cerebellum. Decreasing activation was found in the occipito-temporal lobe (BA 19,39,22), insula (BA 13) as well as the medial prefrontal cortex (BA 9,10) and limbic areas (BA 24,33).

We then tested whether individuals' pattern expression correlated with individuals' behavioral performance or neuropsychological measures recorded outsider the scanner. The correlation with the memory-load related increase in reaction times was very Neural networks in modest, albeit statistically significant ($R^2=0.12$, $p<0.05$) and positive, meaning the more subject increased their expression of the load-related pattern, the more they slowed down from load levels 1 to 6. There was also a significant negative correlation with IQ as estimated by the American version of the National Adult Reading Test ($R^2=0.15$, $p<0.05$).

Thus, the more subjects increased their expression of the load-related network, the lower their NARTIQ, and the slower they responded to the probe at test. NARTIQ and the load-related reaction time increase themselves though were not correlated ($R^2=0.02$, $p=0.3816$). Apparently, the covariance pattern obtained during maintenance captures a *necessary* aspect of neural processing that, while displaying a relationship with behavioral outcomes, is not *advantageous* for task performance. Greater utilization of the covariance pattern did not result in better, but worse performance. Also, subjects who employed the covariance pattern more had lower NARTIQ. It may be that better-performing subjects change their strategies (hence changing the constellation of activated regions) with increasing memory loads. Readers who are interested in a more detailed discussion of the results of all task phases are referred to the original article (7).

Forward application to elderly replication sample

Forward application of the covariance pattern to the elderly replication sample preserves the ordinal trend (rm-F=24.52, $p<0.0001$). Apparently, the elderly subjects also necessarily employ the areas for memory-load related rehearsal in the same fashion as the young subjects. Correlations with NARTIQ Neural and reaction times were not significant. The pattern accounted for 8% of the variance of the raw scans.

For further inspection, we subjected the raw scans of the older subjects to a principal components analysis (without any prior transformation by a design matrix), and forward-apply the load-related covariance pattern obtained from the young subjects to this new set of principal components. The resulting loadings show that the load-related patterns mostly strong loads onto principal components 1–4, with the biggest loading on component 2. This and the lower total variance-contribution of 8% indicates that the covariance pattern underlying load-related maintenance in the young subjects is still operational in the elderly, but that other subject-related variance components are superimposed, hampering the clean identification of the load-related effect. After all, there were 40 young subjects in the derivation sample, but the construction of the load-related pattern only required the inclusion of the first two principal

components, indicating very good concentration of task-related variance, but this is not the case in the elderly.

This exercise illustrates the convenient feature of forward application in spatial covariance analysis: we could still detect the load-related pattern as operating in the elderly subjects, even though the variance accounted for was relatively low. Since the inferential judgment only concerns the subject-expression values resulting from the forward application, in our case a permutation test with the repeated-measures F-score as the test statistic, the variance accounted for by the forward-applied pattern itself is not important. This allows the verification that a pattern is still present through the correlation with experimental and clinical variables, while identification of the pattern in the new data set would be impossible on account of diminished variance contributions. Specifically, in the replication sample of elderly subjects the load-related activation patterns identified in the young is still present, but subject outliers would make its identification in a fresh analysis impossible.

Comparison with univariate analysis

The last section demonstrates one convenient feature of spatial covariance analysis, but it did not tell us how spatial covariance analysis directly compares to univariate analysis. The practical and worthwhile question many researchers have in mind is whether both techniques would highlight the same brain regions. We did not seek the easy way out and resort to the mathematical disclaimer that both methods analyze different aspects of the neural signal – univariate methods strictly look at first-order moments, whereas covariance techniques look at second-order moments and are therefore not beholden to the requirement of identical brain maps.

We tried to mirror the results of the OrT CVA with the corresponding univariate analyses. We first searched for voxels that showed a significant mean trend across memory-load levels. Then we used the activation differences between memory-load levels 6 and 1 and searched for voxels that showed a correlation with (1) the corresponding reaction time difference, (2) NARTIQ. These analyses all used uncorrected p-levels of 0.001.

The results do not lend themselves to a parsimonious interpretation in the way that the multivariate analysis did. The mean-trend analysis across subjects and task conditions revealed a subset of areas of the ordinal-trend covariance patterns that also have good face-validity on the basis of the extant literature (14–16), particularly right dorsolateral prefrontal cortex (BA 9), Broca's area, left BA 44, and left inferior parietal cortex (BA 40) have been shown to increase their activation during retention with increasing memory load, whereas medial prefrontal cortex (BA 10) has been shown to de-activate with increasing task difficulty in a number of tasks, hinting at a possible shifting of resources (17). The univariate and multivariate correlates of increasing memory load are thus consistent with one another, keeping in mind that the univariate analysis fails to identify a number of areas that surfaced in the multivariate analysis.

The areas exhibiting mean-trend activation changes across memory load unfortunately did *not* correlate with any of the behavioral variables the multivariate analysis identified – as a by-product without using the behavioral information in the first place. Worse, no pair of the three univariate analyses showed any overlapping areas; every analysis produced its own unique parametric brain map.

In summary we can thus say that, although there might not be any agreement yet as to a "gold standard" for the results of complex experimental paradigms of higher cognition, we could observe some consistencies between the areas highlighted in the OrT CVA covariance patterns and the corresponding univariate mean-contrast analysis. However, the additional univariate

analyses that probed the voxel-wise correlation between activation and behavioral variables resulted in a proliferation of *different* sets of areas depending on the dependent variable used in the univariate correlational analysis. In the OrT CVA approach, on the other hand, utilizing correlation between different voxels helps in pinpointing a unitary set of co-activating areas whose aggregate behavior can account for both design-related monotonic changes across task-conditions *as well as* behavioral variables. Univariate voxel-wise analyses apparently destroy this unity.

Application to an $H_2{}^{15}O$ PET study of Alzheimer's disease (5)

Our second example of spatial covariance concerns resting PET scans capturing regional cerebral blood flow (rCBF); no design matrix had to be applied to aid the identification of any task-related effects in the data. We used principal components analysis purely in a cross-sectional sense to discriminate between patient and healthy-control groups, thereby identifying a disease-related covariance pattern.

Subject information

Seventeen subjects with mild Alzheimer's disease (AD; 11 male and 6 female), 23 with minimal to mild cognitive impairment but no dementia (CI) (15 male and 8 female) and 16 healthy elderly subjects (8 male and 8 female) met criteria for entry into the study. AD and minimal to mild cognitively impaired (MMCI) subjects were recruited from outpatients who presented to the Alzheimer's Disease Research Center at Columbia University.

AD subjects met DSM-III-R criteria for dementia and NINCDS-ADRDA criteria for probable AD (18). Only AD subjects rated as Clinical Dementia Rating (CDR)=1 (19) were used in this study. Other causes of dementia were excluded with appropriate laboratory tests.

The MMCI subjects were participating in a longitudinal study, the details of which are described elsewhere (20,21). Inclusion criteria were designed to recruit subjects with diagnoses ranging from "nondemented with minimal cognitive impairment" (CDR of 0) or "questionably demented" (CDR of .5) (19). These inclusion criteria were more liberal than in other studies of mild cognitive impairment (22–24) because we intended to enroll a relatively broad group of subjects who presented with cognitive complaints and fell between the "normal" and "dementia" categories, whose cognitive deficits were ranging from minimal to mild severity.

Both AD patients and the MMCI subjects underwent extensive neuropsychological evaluation. For more detailed information about this and the AD and MMCI criteria, readers are referred to the original publication (5). Potential healthy elderly subjects were carefully screened to exclude those with dementia or cognitive impairment, as well as other neurological, psychiatric, or severe medical disorders.

PET Acquisition and Processing

$H_2{}^{15}O$ resting PET scans were acquired. A bolus of 30 mCi $H_2{}^{15}O$ was injected intravenously. Scan acquisition was triggered by the detection of a threshold level of true counts from the camera. Employing a Siemens EXACT 47 PET camera (Knoxville, TN, USA), two 30-s scan frames were acquired in two-dimensional mode, which were subsequently averaged to yield a single image per subject. After measured attenuation correction (15-min transmission scan) and reconstruction by filtered back projection, image resolution was 4.6 mm full width at half maximum (FWHM). Arterial blood sampling was not conducted; thus, only non-quantitative count images (and not absolute rCBF measures) could be obtained, subsequently leading to the discovery of *relative*, and not absolute, flow changes as a function of disease status in the group-level analysis.

The image of each subject was spatially transformed to the PET Montreal Neurological Institute brain space template included with SPM99 program (Wellcome Department of Neurology, London, UK). The spatially transformed images were smoothed with an isotropic Gaussian kernel (FWHM = 12 mm).

Multivariate data analysis

To identify network-correlates of early dementia, we employed an analysis technique that originated from the Scaled Subprofile Model (SSM), a covariance-analysis method (1,25) that has been used previously in resting imaging studies of normal aging and a variety of neurological diseases (26–29). SSM involved a principal components analysis of the scans from the control and AD groups, and captured the major sources of between- and within-group variation, producing a series of principal components (PCs).

To identify a covariance pattern that best discriminated AD patients from controls, each subject's expression (SSF) of a selected set of PCs was entered into a linear regression model as the independent variable. Group membership (AD versus controls) was the dependent variable. This regression results in a linear combination of the PCs that best discriminated the 2 groups. This linear combination of the PCs can itself be conceived of as signifying a 'pattern' or 'network'. We used Akaike's information criterion (30) to determine how many PCs should be included in the regression in order to achieve optimal bias-variance trade-off. The set of PCs that yielded the lowest value in Akaike's information criterion was selected the predictor set in the regression model. We restricted the set of PCs to a cumulative one out of the first few PCs, explaining at least 75% of the variance.

For the estimation of the robust regional weights in the covariance pattern, we again reported to the bootstrap resampling procedure (13) explained earlier, performing our analysis several hundred times on the resampled data to produce a Z-map.

Forward application to MMCI subjects

The AD–related covariance pattern obtained in the cross-section group discrimination between AD subjects and controls was forward–applied to the MMCI group. Resulting SSFs were correlated with clinical outcome measures for additional validation.

Blinded clinical reads and univariate analysis

AD–control group discrimination was first attempted via expert clinical reading of PET scans by a dementia neurology specialist with 10 years of experience in brain SPECT and PET evaluation who was blinded to all clinical information. A dichotomous rating for the presence or absence of the classic AD-related temporoparietal-frontal hypometabolism pattern with sparing of cerebellum, basal ganglia and sensorimotor cortex was generated. This was followed by a rating of high and low levels of confidence in the presence of the normal or AD pattern.

For the univariate analyses, the PET images were intensity-normalized by their average perirolandic count values. We used the average rCBF values of a region of interest corresponding to primary sensorimotor cortex as the intensity normalization factor because this region is typically spared in AD. To effect this, a single perirolandic region of interest, constant for all subjects, was defined on the MNI template brain (using MEDX; Sensor Systems, Inc, Sterling, VA). In each subject, the average PET count value in this region of interest was determined in the spatially normalized PET image (before spatial smoothing). This value was used to divide the image values of the PET image of that subject. We reasoned that intensity normalization should reduce inter-subject scaling noise without "washing out" spatially coherent effects of AD pathology. After normalization, voxel-wise t statistics for between group (AD vs. controls, AD vs. CI and CI vs. controls) comparisons were computed

using procedures in SPM99. The false positive rate was controlled at p = 0.05 per map via Bonferroni correction for the number of resolution elements (RESELS) across which regressions were calculated. Comparisons on a cluster level were also performed using height thresholds corresponding to uncorrected P-values of 0.001.

RESULTS

Group discrimination

Our covariance analysis successfully identified a disease-related covariance pattern in the first 4 PCs, whereas both clinical reading attempts and univariate analysis failed.

Although the pattern accounted for only 6% of the variance on the data, indicating unfavorable signal-to-noise characteristics, it could discriminate between the groups with 76% sensitivity and 81% specificity. Positive loadings (indicating concomitant increased flow) were noted in bilateral insula, lingual gyri and cuneus; left fusiform and superior occipital gyri; and right parahippocampal gyrus and pulvinar. Negative loadings (indicating concomitant decreased flow) were noted in bilateral inferior parietal lobule and cingulate; and left middle frontal, inferior frontal, precentral and supramarginal gyri (cf figure 4 and check original publication (5)).

Forward application to cognitively impaired subjects

We next applied the disease-related pattern to the group of minimally to mildly cognitively impaired people for additional validation. Our expectation was that subject expression in this replication sample should correlate with clinical and neuropsychological performance variables with the correct directionality: people exhibiting *worse* performance should express the AD-pattern to a *higher* degree. Expression of the pattern was higher in the 9 CI subjects with CDR = 0.5 as compared to the 14 CI subjects with CDR = 0 ($p<0.01$; also cf figure 4). Thus, CI subjects with greater functional deficits expressed the pattern more like the AD group, while those CI subjects without functional deficits expressed the pattern more like the healthy elderly control group. Because of slight age differences between the two CI groups, we controlled for age in a supplementary analysis: the results were unchanged.

Next we checked correlation of pattern expression and neuropsychological performance in the MMCI group: greater pattern expression did account for worse performance on the Selective Reminding Test ($R^2=0.25$, $p<0.05$) and the modified Mini Mental Status Examination ($R^2=0.20$, $p<0.05$) (uncorrected for 8 computations).

Because sufficient follow-up information was available from the time the PET scans were performed, we also investigated whether baseline expression of the AD-related pattern could predict disease progression in the MMCI subjects (31). We first investigated whether the baseline expression of the AD-related CBF pattern in the MMCI subjects could predict their conversion to AD status (CDR=1). For this we conducted a logistic regression with the binary outcome measure of dementia status (1/0) as the dependent variable and baseline pattern expression as the independent variable. 6/23 of the MMCI subjects progressed to dementia. We performed a median split of the baseline expression levels of the AD-related CBF pattern, dichotomizing the independent variable. The median for the subject expression of the AD-related pattern was 23.40; this value corresponds to a sensitivity of 76% and a specificity of 81% in the AD-Normal discrimination that produced the covariance pattern originally. We conducted a logistic regression with this dichotomized independent variable and obtained a regression weight of $\beta=1.727\pm1.198$ ($p=0.15$), translating to an odds ratio of 5.625, meaning a more than five-fold increased risk of becoming demented in subjects with the higher baseline

expression of the AD-related pattern. While this logistic regression did not reach statistical significance, it suggests an effect size that will clearly be detectable with a higher sample size.

We also investigated whether the subject expression of the AD-related covariance pattern at baseline had any predictive power with regard to subsequent functional and neuropsychological decline in the 23 MMCI patients. Follow-up neuropsychological evaluation of these subjects was done roughly every 6 months from 1.98 to 7.07 years. The relation between baseline SSF and cognitive decline was evaluated with a generalized equation estimation (GEE) model (32). This statistical method takes into account the multiple visits per subject and the fact that the characteristics of a single individual over time are likely to be correlated with one another. The repeated measures for each subject are treated as a cluster. A second advantage of GEE is that it takes into account the status or changing value of each covariate at each visit.

We chose the Selective Reminding Test (SRT) Delayed Recall score (33) as the dependent variable. The GEE model included two independent variables, time (test interval measured in years) and the baseline expression of the covariance pattern, as well as the interaction term of time by subject baseline expression. Decreasing SRT scores indicate increasing disease severity, prompting the expectation of a *negative* regression weight of the years of follow-up (as time goes on, the disease becomes worse), and possibly a *negative* regression weight of initial baseline expression of the AD-related pattern (the higher the baseline expression, the worse the disease). The interaction term tests for differential change over time as a function of the initial expression of the AD-related covariance pattern. We also hypothesized a *negative* regression weight of the interaction term between years of follow-up and baseline pattern expression: i.e. as time goes on, people with higher baseline expression of the pattern will show a more rapid decrease in the SRT score.

The results of the GEE analysis were as follows. The SRT Delayed Recall score improved significantly over time, $\beta=1.685\pm0.421$, $p<0.0001$, which is a bit surprising. Apparently, as a group, the MMCI group enjoyed a learning effect with repeated testing. However, there was a significant interaction between time and baseline expression of the AD-related CBF pattern with a negative regression weight $\beta=-0.071\pm0.019$, $p<0.0001$. Per unit of increased baseline expression of the AD-related CBF pattern, MMCI subjects improved 0.071 points less on the SRT Delayed Recall test.

Overall, these results of the forward application to the MMCI replication sample provide good support for the use of the AD-related CBF pattern as a prognostic and diagnostic tool (31).

DISCUSSION

We close our report with a discussion of our results and connect it to a few general comments and observations, some of which have emerged from our extensive experience with univariate and multivariate analytic techniques. Covariance techniques have been around for some time and have been applied with good success now for more than a decade (see for instance (1,2, 26,34)); but they only recently have attracted increasing attention in the neuroscience community. We hope our article can contribute to making them more palatable to the community at large and re-iterate their attractive feature form our point of view.

We presented two examples that utilized covariance analysis in slightly different ways: one was a cognitive-neuroscience fMRI paradigm employing a standard DMS task to investigate verbal working memory (7), the second was a clinical PET study of Alzheimer's in a cross-sectional comparison (5). In both cases valuable insights could be supplied that surpassed the results of mass-univariate analyses. For our first example, covariance provided a parsimonious account of the neural correlates of necessary memory-load related rehearsal, and showed how it related to behavioral performance and NARTIQ. The parallel univariate analyses were not

successful in accounting for task-and subject effects simultaneously; instead every dependent variable produced its own set of non-overlapping brain images. This is clearly unsatisfactory for good model- and theory-building, and presents the analyst with a plethora of choices after the fact.

In the example of the clinical PET study of Alzheimer's disease, no results were produced by the univariate analysis at all, hinting at better sensitivity of multivariate compared with univariate techniques. Obviously, which of the two is better applicable to the data set under consideration depends on the effect of interest itself. One can conceive of a continuum with best suitability for both types of analytic techniques at either end: for truly focal effects in the data, involving only one circumscribed brain region, univariate techniques might indeed be more powerful than multivariate ones which risk committing a type-II error on account of the effect's diminished variance contribution. For a truly distributed effect, involving subtle signal differences that are widely dispersed in the brain, the multivariate approach is likely to be superior, since stringent false-positive corrections (like Bonferroni etc.) risk "correcting away" the true positives.

But even in the gray zone between the two ends of the spectrum where clear-cut superiority of one approach over the other is *a priori* plausible, we have found consistently that a multivariate analysis provides further benefits. The data sets presented in this report are representative of the to-be-expected quality of brain-scan data in terms of signal-to-noise characteristics and experimental control. Particularly in clinical studies, these parameters are not up to the experimenter's choice anyway. In absence of any special knowledge that posits preferential treatments for a selected set of brain areas, we have found that multivariate approaches consistently provide more information than univariate approaches. Even their failure is usually instructive and performs a better job at quality control than their univariate cousins. This virtue of utilizing the brain-wide activation signal in a more complete way than univariate analysis occasionally presents challenges to the novice and delivers a break-down of variance components that is not easily interpretable in terms of the theory that informed the experimental design in the first place.

We would encourage the researcher to persist nevertheless, and venture that a complete account of the data with undesirable results is still superior to palatable results that were obtained from a highly selective analysis process that only focused on data from a few brain regions, while neglecting the remainder information gathered. In this sense, we find the terms "exploratory" and ";hypothesis-led" misleading. In practice, every neuroimaging analysis will be somewhere in between, i.e. trying to validate prior notions about brain functions, while still analyzing the acquired data above and beyond these notions with the goal of finding fresh insights that can inform the model-building process.

After making this general remark, we also want to defend multivariate approaches against the oft-leveled charge that they are purely exploratory, or that they "will always find something no matter what" (actual quote of a reviewer). To our knowledge, all recent multivariate techniques, whether based on principal or independent components analysis, have clearly formulated inferential frameworks. Stringent methods exist to ascertain whether their findings came about as a result of type-I errors or not. Some of the inferential frameworks employ non-parametric resampling algorithms, which are occasionally regarded by some researchers as a matter of bad taste, but we find such disdain illogical. Modern computers are powerful enough to offer very efficient and fast computation of p-levels for *any* test statistic of choice. In this light, the insistence on closed-formula solutions, which are dependent on untested assumptions about distribution characteristics of the data, does not make sense.

A further defense against the charges of unprincipled "data dredging" is the convenient feature of forward application: both examples in this report used this feature extensively for further validation. Clearly, with multivariate techniques one is not confined to just rely on p-levels that come out of the inferential framework, but can go further and put these to the test in replication experiments; this is possible in a manner that is much less dependent on the details of experimental design and drawing of regions-of-interests than for univariate analyses.

Multivariate analytic techniques are thus a worthy pursuit as this report (and others in this special issue) has argued. Much work remains to be done to facilitate dissemination in the community at large: a first step would be easier-to-use software packages, didactic reports, tutorials and conference workshops. It is our contention that the efforts to bring this about will benefit neuroscience in the long term.

Acknowledgments

This research was supported by Federal Grants AG 14671, AG 17761, RR 00645, the Alzheimer's Association and DARPA.

References

1. Moeller JR, Strother SC, Sidtis JJ, Rottenberg DA. Scaled subprofile model: a statistical approach to the analysis of functional patterns in positron emission tomographic data. J Cereb Blood Flow Metab 1987;7(5):649–658. [PubMed: 3498733]
2. McIntosh AR, Bookstein FL, Haxby JV, Grady CL. Spatial pattern analysis of functional brain images using partial least squares. Neuroimage 1996;3(3 Pt 1):143–157. [PubMed: 9345485]
3. Worsley KJ, Poline JB, Friston KJ, Evans AC. Characterizing the response of PET and fMRI data using multivariate linear models. Neuroimage 1997;6(4):305–319. [PubMed: 9417973]
4. Habeck C, Krakauer JW, Ghez C, Sackeim HA, Eidelberg D, Stern Y, et al. A new approach to spatial covariance modeling of functional brain imaging data: ordinal trend analysis. Neural Comput 2005;17 (7):1602–1645. [PubMed: 15901409]
5. Scarmeas N, Habeck CG, Zarahn E, Anderson KE, Park A, Hilton J, et al. Covariance PET patterns in early Alzheimer's disease and subjects with cognitive impairment but no dementia: utility in group discrimination and correlations with functional performance. Neuroimage 2004;23(1):35–45. [PubMed: 15325350]
6. Habeck C, Rakitin BC, Moeller J, Scarmeas N, Zarahn E, Brown T, et al. An event-related fMRI study of the neurobehavioral impact of sleep deprivation on performance of a delayed-match-to-sample task. Brain Res Cogn Brain Res 2004;18(3):306–321. [PubMed: 14741317]
7. Habeck C, Rakitin BC, Moeller J, Scarmeas N, Zarahn E, Brown T, et al. An event-related fMRI study of the neural networks underlying the encoding, maintenance, and retrieval phase in a delayed-match-to-sample task. Brain Res Cogn Brain Res 2005;23(2–3):207–220. [PubMed: 15820629]
8. Sternberg S. High-speed scanning in human memory. Science 1966;153(736):652–654. [PubMed: 5939936]
9. Sternberg S. Memory-scanning: mental processes revealed by reaction-time experiments. Am Sci 1969;57(4):421–457. [PubMed: 5360276]
10. Zarahn E. Testing for neural responses during temporal components of trials with BOLD fMRI. Neuroimage 2000;11(6 Pt 1):783–796. [PubMed: 10860802]
11. Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R. Event-related fMRI: characterizing differential responses. Neuroimage 1998;7(1):30–40. [PubMed: 9500830]
12. Friston KJ, Josephs O, Zarahn E, Holmes AP, Rouquette S, Poline J. To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. Neuroimage 2000;12(2):196–208. [PubMed: 10913325]
13. Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. New York: CRC Press, LLC; 1994.
14. Rypma B, D'Esposito M. The roles of prefrontal brain regions in components of working memory: effects of memory load and individual differences. Proc Natl Acad Sci U S A 1999;96(11):6558–6563. [PubMed: 10339627]

15. Postle BR, Zarahn E, D'Esposito M. Using event-related fMRI to assess delay-period activity during performance of spatial and nonspatial working memory tasks. Brain Res Brain Res Protoc 2000;5 (1):57–66. [PubMed: 10719266]

16. Rypma B, Berger JS, D'Esposito M. The influence of working-memory demand and subject performance on prefrontal cortical activity. J Cogn Neurosci 2002;14(5):721–731. [PubMed: 12167257]

17. Engle RW, Conway ARA, Tuholsky SW, Shisler RJ. A Resource Account of Inhibition. Psychological Science 1995;6:122–125.

18. 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;34(7):939–944. [PubMed: 6610841]

19. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry 1982;140:566–572. [PubMed: 7104545]

20. Devanand DP, Michaels-Marston KS, Liu X, Pelton GH, Padilla M, Marder K, et al. Olfactory Deficits in Patients With Mild Cognitive Impairment Predict Alzheimer's Disease at Follow-Up. Am J Psychiatry 2000;157(9):1399–1405. [PubMed: 10964854]

21. Tabert MH, Albert SM, Borukhova-Milov L, Camacho Y, Pelton G, Liu X, et al. Functional deficits in patients with mild cognitive impairment: Prediction of AD. Neurology 2002;58(5):758–764. [PubMed: 11889240]

22. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999;56(3):303–308. [PubMed: 10190820]

23. Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, et al. Current concepts in mild cognitive impairment. Arch Neurol 2001;58(12):1985–1992. [PubMed: 11735772]

24. Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001;56(9): 1133–1142. [PubMed: 11342677]

25. Alexander GE, Mentis MJ, Van Horn JD, Grady CL, Berman KF, Furey ML, et al. Individual differences in PET activation of object perception and attention systems predict face matching accuracy. Neuroreport 1999;10(9):1965–1971. [PubMed: 10501542]

26. Moeller JR, Ishikawa T, Dhawan V, Spetsieris P, Mandel F, Alexander GE, et al. The metabolic topography of normal aging. J Cereb Blood Flow Metab 1996;16(3):385–398. [PubMed: 8621743]

27. Moeller JR, Nakamura T, Mentis MJ, Dhawan V, Spetsieres P, Antonini A, et al. Reproducibility of regional metabolic covariance patterns: comparison of four populations. J Nucl Med 1999;40(8): 1264–1269. [PubMed: 10450676]

28. Hutchinson M, Nakamura T, Moeller JR, Antonini A, Belakhlef A, Dhawan V, et al. The metabolic topography of essential blepharospasm: a focal dystonia with general implications. Neurology 2000;55(5):673–677. [PubMed: 10980732]

29. Nakamura T, Ghilardi MF, Mentis M, Dhawan V, Fukuda M, Hacking A, et al. Functional networks in motor sequence learning: abnormal topographies in Parkinson's disease. Hum Brain Mapp 2001;12 (1):42–60. [PubMed: 11198104]

30. Burnham, KP.; Anderson, DR. Model Selection and Multimodel Inference. New York: Springer Verlag; 2002.

31. Devanand DP, Habeck C, Tabert MH, Scarmeas N, Pelton GH, Moeller JR, et al. PET network abnormalities and cognitive decline in patients with mild cognitive impairment. Neuropsychopharmacology. in press

32. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrica 1986;73:13–22.

33. Buschke H, Fuld PA. Evaluating storage, retention, and retrieval in disordered memory and learning. Neurology 1974;24(11):1019–1025. [PubMed: 4473151]

34. McIntosh AR, Rajah MN, Lobaugh NJ. Interactions of prefrontal cortex in relation to awareness in sensory learning. Science 1999;284(5419):1531–1533. [PubMed: 10348741]

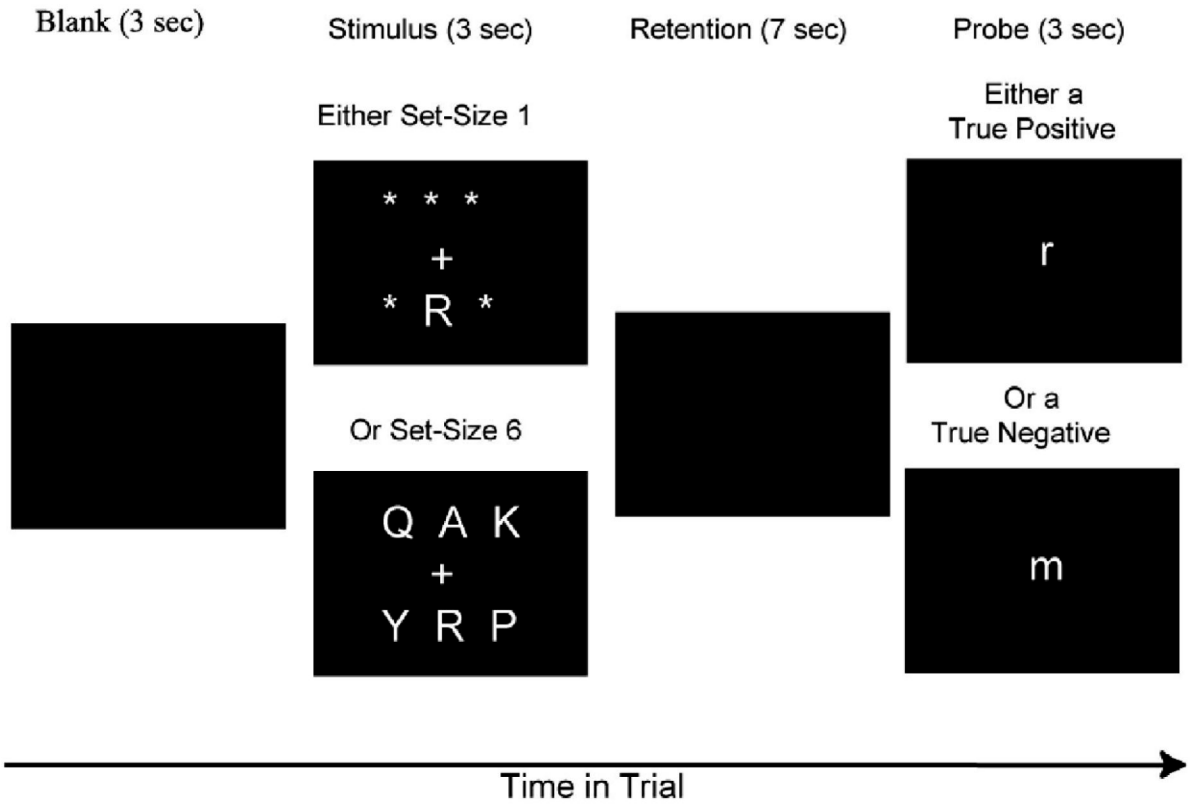

Figure 1.
Schematic sketch of delayed-match-to-sample paradigm of the Sternberg variant.

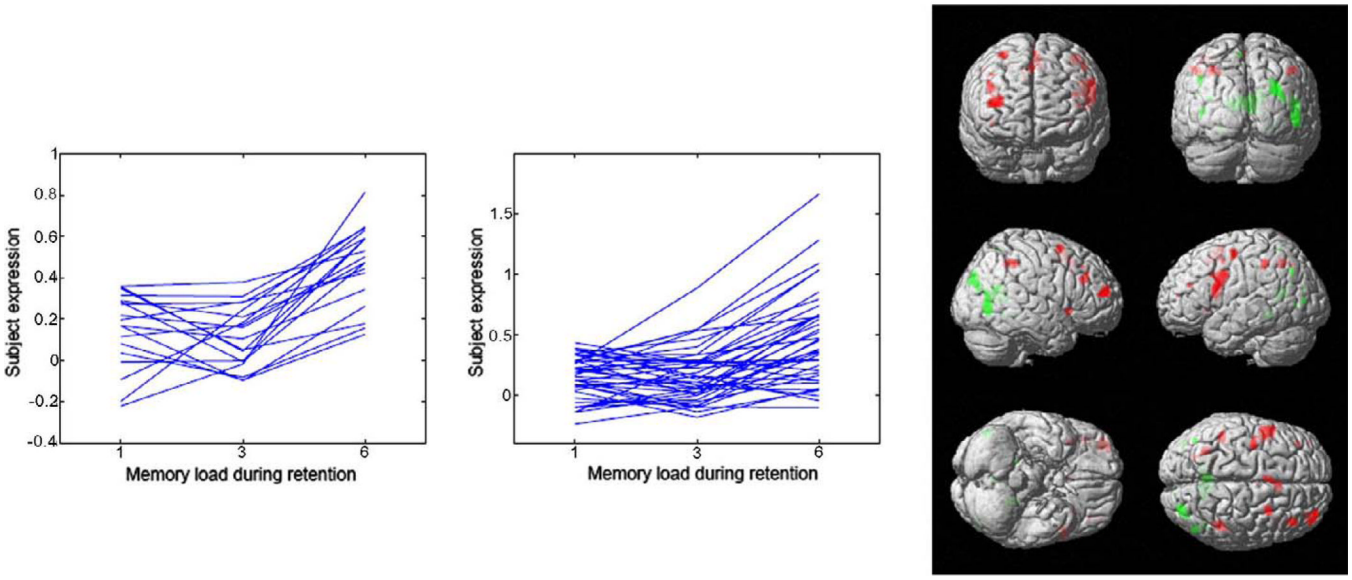

Figure 2.

Summary of the characteristics relating to subject expression across memory-load levels (1/3/6/ letters) and topographic composition of the load-related covariance pattern, isolated from the maintenance phase of the DMS task in young people. **Left:** subject expression of the pattern in the elderly replication sample. Monotonically increasing expression is apparent on a subject-by-subject basis (rm-F=24.52, $p<0.0001$). **Middle**: subject expression of the covariance pattern in the original derivation sample of young subjects. The ordinal trend is also clearly visible here (rm-F=38.66, $p<0.001$). **Right**: Surface rendering of brain regions in the activation pattern, as ascertained by the bootstrap resampling procedure ($p<0.01$). Red color denotes brain regions whose activation associated with the pattern increases as a function of memory load; green color denotes regions whose activation decreases.

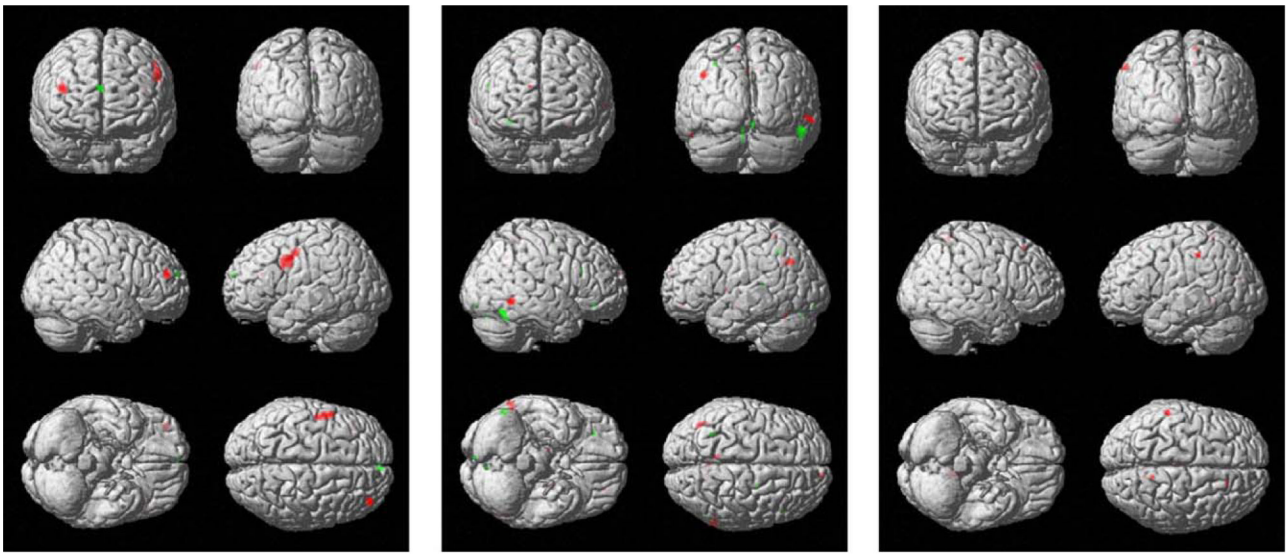

Figure 3.
Voxel-wise univariate results from the data of the maintenance phase of the derivation sample of young subjects, conducted in parallel to the multivariate findings. **Left**: mean trend of load-related activation across memory load. **Middle**: correlation of the 6–1 differences in activation with NARTIQ. In both maps, red color denotes positive values and green color denotes negative values. **Right**: correlation of 6-1 difference in activation with corresponding difference in RT, only a negative correlation was found. All maps were thresholded at uncorrected $p<0.001$. There is no overlap between any two maps.

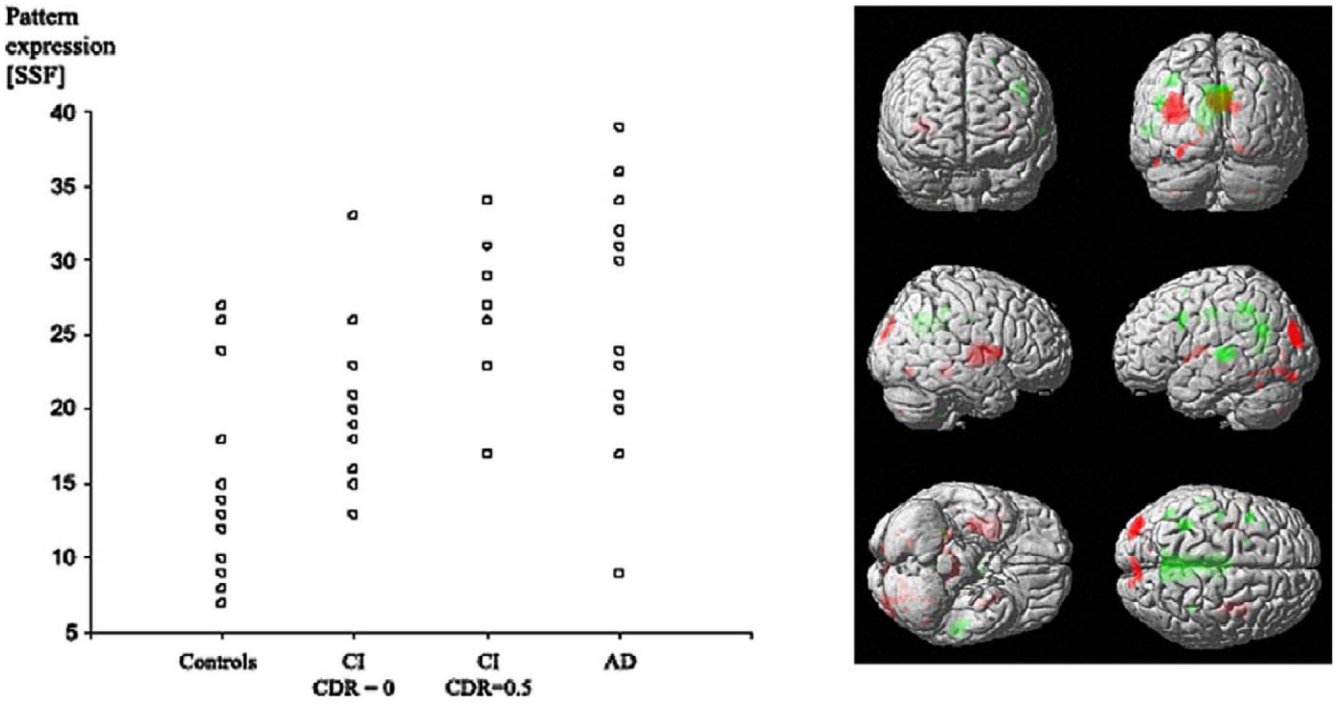

Figure 4.

Summary of the characteristics relating to subject expression and topographic composition of the AD-related covariance pattern, obtained from the cross-sectional comparison of AD subjects and healthy controls. **Left**: subject expression of the pattern in all subjects, both from the replication as well as derivation samples. **Middle**: subject expression of the covariance pattern in the original derivation sample of young subjects. One can discern a mean increase in subject expression across disease severity, from healthy controls (CONTROLS), over minimally cognitively impaired (CI: CDR=0), mildly cognitively impaired (CI: CDR=0.5), to AD (CDR=1). **Right**: Surface rendering of brain regions in the covariance pattern, as ascertained by the bootstrap resampling procedure ($p<0.01$). Red color denotes brain regions whose associated relative rCBF is increased as a function of disease status; green color denotes regions whose relative rCBF is decreased.