

Development of an MRI rating scale for multiple brain regions: comparison with volumetrics and with voxel-based morphometry

R. Rhys Davies · Victoria L. Scahill · Andrew Graham ·
Guy B. Williams · Kim S. Graham · John R. Hodges

Received: 25 November 2008 / Accepted: 4 March 2009 / Published online: 24 March 2009
© Springer-Verlag 2009

Abstract

Introduction We aimed to devise a rating method for key frontal and temporal brain regions validated against quantitative volumetric methods and applicable to a range of dementia syndromes.

Methods Four standardised coronal MR images from 36 subjects encompassing controls and cases with Alzheimer's disease (AD) and frontotemporal dementia (FTD) were used. After initial pilot studies, 15 regions produced good intra- and inter-rater reliability. We then validated the ratings against manual volumetry and voxel-based morphometry (VBM) and compared ratings across the subject groups.

Results Validation against both manual volumetry (for both frontal and temporal lobes), and against whole brain VBM, showed good correlation with visual ratings for the majority of the brain regions. Comparison of rating scores across disease groups showed involvement of the anterior fusiform gyrus, anterior hippocampus and temporal pole in semantic dementia, while anterior cingulate and orbitofrontal regions were involved in behavioural variant FTD.

Conclusion This simple visual rating can be used as an alternative to highly technical methods of quantification, and may be superior when dealing with single cases or small groups.

Keywords Visual rating scale · MRI · VBM

Introduction

Visual rating of MR scans can provide anatomical data for clinical diagnosis or for research [1–3]. Advantages of the rating approach include speed, a flexibility and clinical applicability. Other approaches, however, provide more precise quantification. For instance, MRI volumetry, the 'gold standard' of structural imaging, typically involves tracing boundaries by hand [3–5]. Various automated techniques devised in recent years, notably voxel-based morphometry (VBM), lend objectivity to brain measurement and permit assessment of the whole brain [6, 7]. Disadvantages of these methods, however, are that they require sophisticated and time-consuming post-imaging analysis. They also stipulate specific scanning protocols. By contrast, clinical scans are often available from cases where instructive behavioural data have been collected but where further scanning is not possible. Valuable neuropsychological data may also be held by workers without access to automated imaging resources who may wish to perform studies involving structure–function correlation.

A number of MRI rating scales have been devised but have focused on certain lesion types (e.g. infarcts [2, 8]) or on individual regions, e.g. hippocampus [1, 9], which limits their applicability. There is scope for extension of this approach as illustrated by Galton et al. who successfully established a method for assessing atrophy of temporal lobe

R. R. Davies · A. Graham · G. B. Williams · J. R. Hodges
Department of Clinical Neurosciences, University of Cambridge,
R3 Box 83, Addenbrooke's Hospital,
Cambridge CB2 2QQ, UK

V. L. Scahill · A. Graham · K. S. Graham · J. R. Hodges
MRC Cognition and Brain Sciences Unit, Cambridge and Wales
Institute of Cognitive Neuroscience, School of Psychology,
Cardiff University,
Cardiff CB2 7EF, UK

J. R. Hodges (✉)
Cognitive Neurology, Prince of Wales Medical Research Institute,
Barker Street Randwick,
Sydney, NSW 2031, Australia
e-mail: j.hodges@powmri.edu.au

sub-regions [3]. In the light of recent insights concerning localisation of brain function and distribution of atrophy in neurodegenerative disease [4, 10], we wanted to develop a more extensive rating scale which could be readily applied to good quality clinical MRI images. We also intended the scheme to encompass brain abnormalities other than neurodegeneration, including post-operative or post-encephalitic appearances, where a given brain structure may be completely destroyed.

The primary aim of this study was to devise a reliable rating scheme focusing on temporal and frontal lobe regions important in diagnosis, in localisation of function, or both. Secondly, because pilot studies were conducted over many regions, we were able to gauge the limits of rating as an approach. Thirdly, as the method was developed using scans from patients with Alzheimer's disease (AD), behavioural variant FTD (bvFTD) and semantic dementia (SD), we were able to draw inferences about distribution of atrophy in those syndromes. Finally, we examined the relationship between rating and the automated structural imaging method of VBM. This included using VBM for the validation of rating scores and, also, a comparison of rating data with VBM analyses across the neurodegenerative disease groups.

Materials and methods

Development of rating method

The following candidate regions were selected: basal ganglia, orbitofrontal, dorsolateral frontal, ventrolateral frontal, cingulate (anterior and posterior), temporal pole, amygdala, hippocampus (anterior, mid and posterior), parahippocampal gyrus (anterior, mid and posterior), banks of the collateral sulcus, fusiform gyrus (anterior, mid and posterior), lateral temporal, insula, superior temporal, posterior temporal and inferior parietal lobule. Landmarks for the regions were taken from standard reference works [11–13]. To maximise the efficiency of the scheme, the number of images used in the protocol was limited to four easily identifiable T1 coronal scan slices. All regions of interest, apart from the amygdala, could be viewed in one of the four slices; the amygdala was consequently removed from the protocol. In order to maximise applicability, scan slice thickness and brain orientation were not specified beyond the fact that the slices must recognisably be in the coronal plane (perpendicular to the anterior commissure–posterior commissure axis, $\pm 10^\circ$).

The four slices defined by the protocol were, in antero-posterior sequence: (I) the most anterior slice in which the internal capsule may be seen in the corpus striatum; (II) the

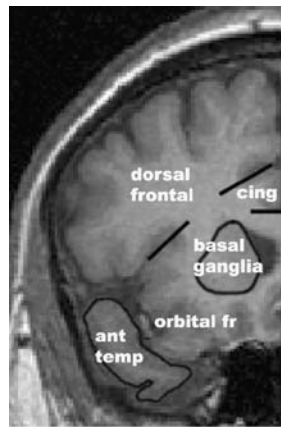
most anterior slice in which the medial opening of the intralimbic gyrus is visible; (III) the slice midway between II and IV, at the same antero-posterior level as the lateral geniculate nucleus; (IV) the most anterior slice in which the crus of the fornix may be seen in continuity alongside the pulvinar (Fig. 1). The slices were determined individually for each hemisphere to minimise the confounding effects of variation in brain orientation.

Initial pilot studies showed that rating was not viable for eight of the initial candidate regions because of a high degree of variability among controls or local degradation of image quality. These were: slice I—dorsolateral and ventrolateral frontal cortex (later combined to give the large 'lateral frontal cortex' region); slice III—mid-parahippocampal gyrus, mid-fusiform gyrus; slice IV—posterior parahippocampal, posterior fusiform, inferior parietal lobule, posterior cingulate cortex. The final selection of 15 regions is listed in Table 1 (see also Fig. 1).

The rating scheme encompasses four stages of atrophy [1–4] with 0 indicating no atrophy and 4 being most abnormal. Rating 2 indicates that a structure is undoubtedly abnormal while rating 1 implies borderline appearances that are not categorically abnormal. In the cases of the basal ganglia, hippocampus and temporal pole, the absence of discernible tissue is rated 4. The presence of only skeletal remnants of white matter without identifiable grey qualifies for a rating of 4 for the remaining cortical structures. A rating of 3 indicates severe atrophy falling short of criteria for 4. For consistency, appearances seeming to fall between two levels were always given the higher rating.

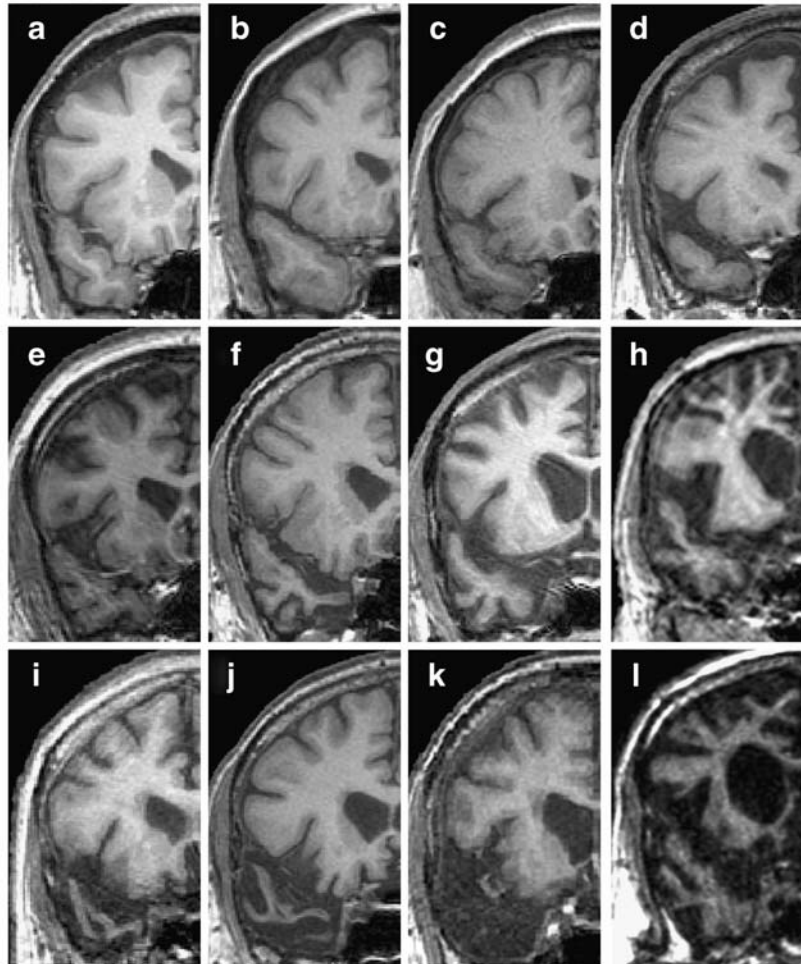
For certain regional ratings, specific criteria were adopted. The basal ganglia rating criteria were: (0) convex contour of basal ganglia with normal-sized ventricle; (1) prominence of the lateral ventricle but contour remaining convex; (2) contour of basal ganglia flattened; (3) concave basal ganglia; and (4) basal ganglia no longer discernible. The anterior hippocampal criteria were: (0) no significant cerebrospinal fluid (CSF) spaces visible; (1) mild prominence of CSF but hippocampal structure essentially normal; (2) convolutions of anterior hippocampus clearly separated by CSF; (3) remnant of hippocampus at the medial tip of the temporal horn; and (4) hippocampus no longer discernible. The mid-hippocampal criteria were: (0) no more than a small crescent of CSF visible; (1) mild prominence of CSF but hippocampal structure essentially normal; (2) temporal horn of ventricle taking up larger area on slice than hippocampus; (3) remnant of hippocampus at the medial tip of the temporal horn; and (4) hippocampus no longer discernible. The posterior hippocampal criteria were: (0) no significant CSF space between the fornix and the retrosplenial cortex; (1) prominence of CSF but hippocampal structure essentially normal; (2) CSF taking up larger area under fornix than the hippocampus; (3)

Fig. 1 Reference pamphlet slices 1 to 4. Guide for identifying regions to be rated (*upper left hand corner*) and images for comparison while rating showing range of severities (0–4) for each rated region. Slice 1: *temp (ant)* anterior temporal pole; *b ganglia* basal ganglia; *orbital fr* orbitofrontal cortex; *dorsal fr* dorsolateral frontal cortex; *cingulate* anterior cingulate cortex. Slice 2: *HPC (ant)* anterior hippocampus; *ERC* entorhinal cortex (anterior parahippocampal gyrus); *PRC* perirhinal cortex; *fus (ant)* anterior fusiform gyrus; *temp (lat)* lateral temporal region; insula. Slice 3: *HPC (mid)* mid-hippocampus; *temp (sup)* superior temporal gyrus. Slice 4: *post HPC* posterior hippocampus; 37 area 37



Slice 1

	a	b	c	d	e	f	g	h	i	j	k	l
temp (ant)	0	0	1	1	2	2	2	2	3	3	4	2
b ganglia	0	0	0	0	1	2	3	3	1	2	3	4
orbital fr	0	0	0	1	2	1	3	4	2	3	3	4
dorsal fr	0	1	0	2	2	1	2	3	2	1	2	4
cingulate	0	0	0	1	0	2	2	3	2	1	2	4



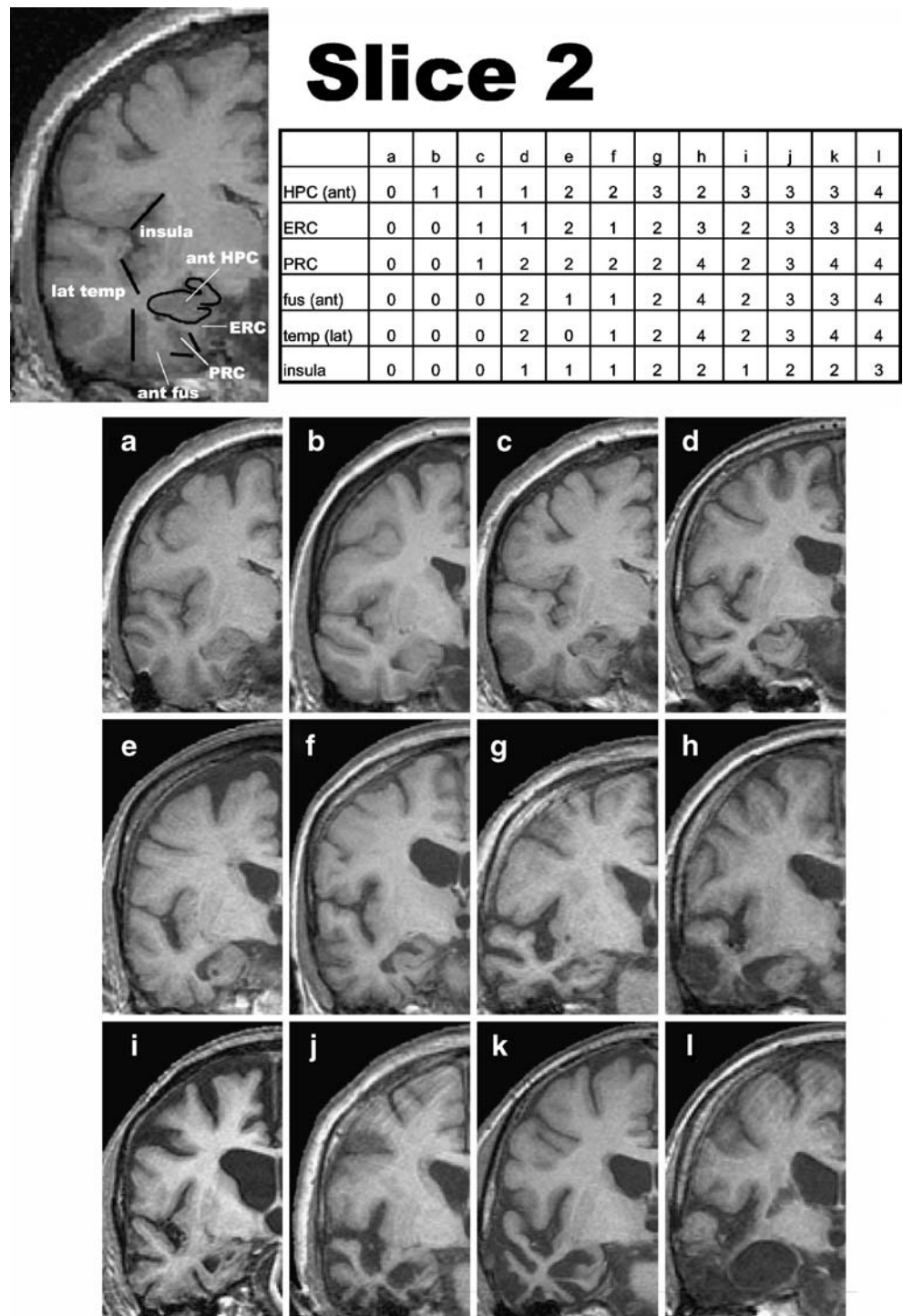
remnant of hippocampus at the base of the fornix; and (4) hippocampus no longer discernible.

A key source of reference during the rating process was an array of pre-rated images with a guide to the target regions for rating (Fig. 1) and illustrations of levels of atrophy taken from patients and controls selected to encompass the whole range of pathologies likely to be encountered by clinicians using the scale. The pamphlet images were consulted during each rating session.

Evaluation of rating method

Methodological analyses were based on the rating of 36 MR scans. These included controls ($n=8$, age=63.8±6.0), Alzheimer’s disease (AD) cases ($n=8$, age=64.9±4.6) and frontotemporal dementia (FTD) cases [14, 15] ($n=20$; semantic dementia $n=9$, age=61.1±7.8; behavioural variant FTD $n=11$, age=57.8±5.6). The scans selected had all been used in previous volumetric studies, allowing compar-

Fig. 1 (continued)



isons across data-sets [3, 4, 16]. All scans were also available in electronic form (necessary for VBM analysis).

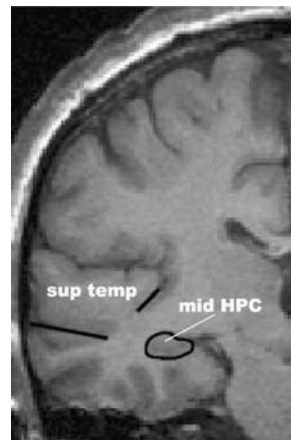
Agreement analyses

Agreement analyses were based on assessment of all scans by three independent raters blinded to clinical details and, after further blinding, repeat rating by one of the three (rater

1, RD), a behavioural neurologist, with extensive experience of reviewing scans. Rater 2 (VS), a neuropsychologist, was involved with selecting the regions for rating and the preparation of reference material. Rater 3 (AG), a trained neurologist, had not been involved with the initial phase of devising the scheme.

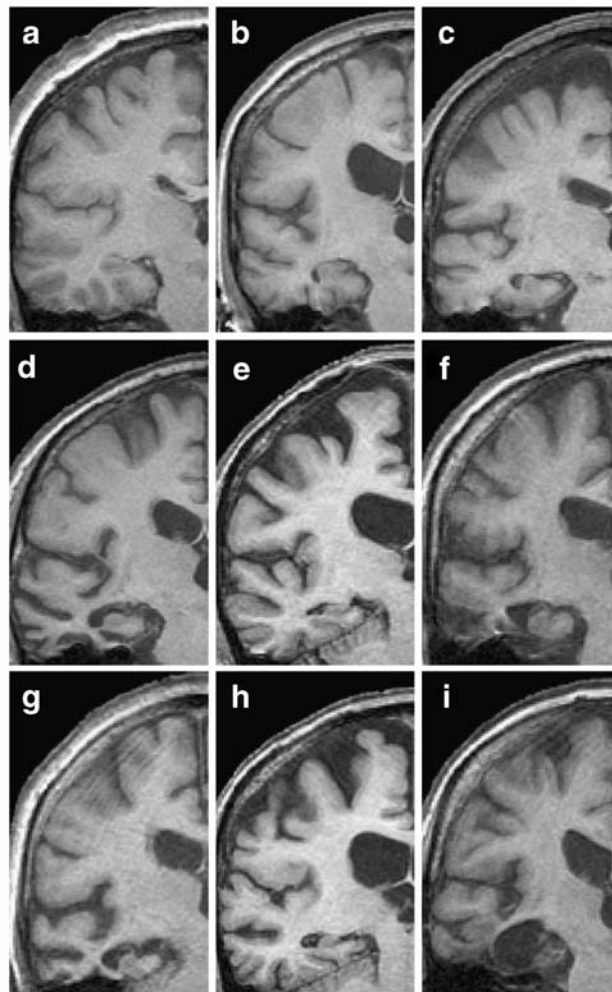
The rating of 36 scans yielded 72 measurements in total for each region (ratings for left and right for each of the 36

Fig. 1 (continued)



Slice 3

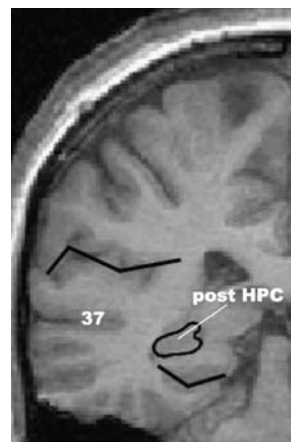
	a	b	c	d	e	f	g	h	i
HPC (mid)	0	1	1	2	2	2	3	3	4
temp (sup)	0	1	1	1	2	1	2	2	2



cases). Agreement between raters was measured overall, by region, and by subject group, using Cohen’s kappa statistic and the intra-class correlation coefficient. For each pair of raters, three measures of agreement were calculated: the raw proportion of ratings in absolute agreement and kappa, unweighted and weighted. The weighted kappa statistic is more appropriate than the unweighted as discrepancies carry greater weight if raters

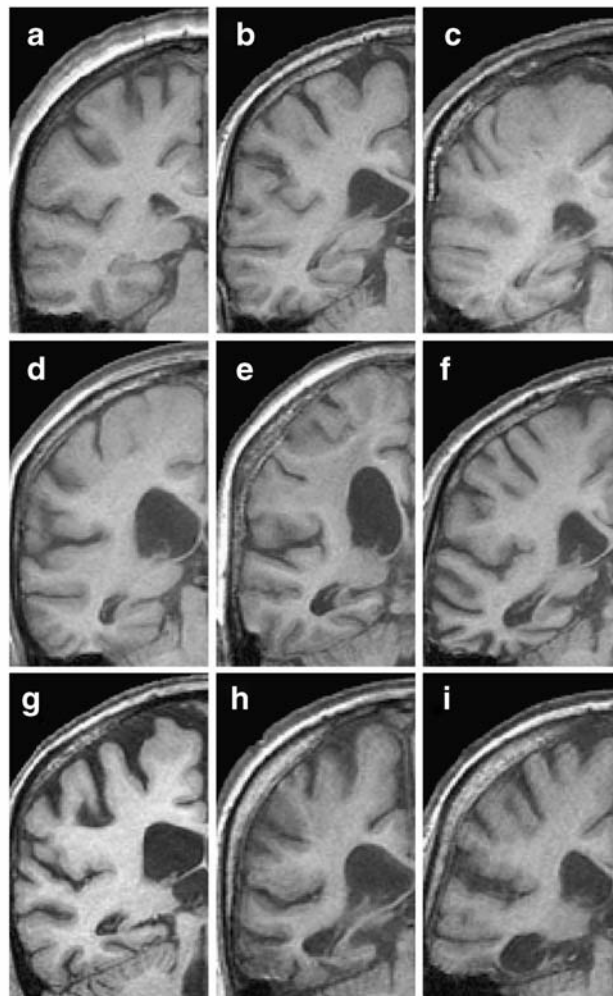
have disagreed by a greater margin. The intraclass correlation coefficient (ICC), unlike kappa, permits simultaneous analysis of agreement across all raters but has the disadvantage of being influenced more by variance in the sample assessed (more homogeneous samples will yield lower ICCs for the same measure). Scoring from the two sessions undertaken by rater 1 were compared in a similar manner to the inter-rater data.

Fig. 1 (continued)



Slice 4

	a	b	c	d	e	f	g	h	i
HPC (pos)	0	1	1	2	2	2	3	3	4
37	0	1	0	1	1	3	2	2	3



Validation analyses: manual volumetrics

To validate the scale against independent volume measures, rating data were compared with manual volumetrics and with VBM (discussed below). A subset of 25 scans (eight controls; eight AD patients; nine SD patients) had been used in a manual volumetric study of four temporal lobe

regions (temporal pole, perirhinal, entorhinal, anterior hippocampus) [4]. For the frontal lobe, manual volumetric data were available from four regions (orbitofrontal, insular, lateral frontal, anterior cingulate) in 14 scans (three controls and 11 bvFTD) [16]. Temporal lobe volumetric data were based on grey matter boundaries defined by Insausti (temporopolar, perirhinal and entorhinal cortices) or Watson

Table 1 Brain regions included in the final rating protocol.

Slice	Region	Approximate Brodmann designation	Talairach coordinates (x, y, z)	Small volume correction radius (mm)
I	Temporal pole	BA 36/38	-41 16 -30	10
	Anterior basal ganglia	N/A	-19 16 0	5
	Orbitofrontal gyri	BA 11/25	-35 16 8	5
	Lateral frontal gyri	BA 8/9/46	-36 24 32	15
	Anterior cingulate gyrus	BA 24	-6 24 27	3
II	Anterior hippocampus	N/A	-22 -12 -21	3
	Anterior parahippocampal gyrus (entorhinal)	BA 28	-20 -12 -31	2
	Collateral sulcus (perirhinal)	BA 35	-29 -12 -31	2
	Anterior fusiform gyrus	BA 36/20	-32 -12 -36	2
	Lateral temporal gyri	BA 21/22	-53 -12 -26	5
	Insula	BA 13	-39 -8 -2	3
III	Mid hippocampus	N/A	-29 -20 -15	2
	Superior temporal gyrus	BA 41/42	-43 -20 -1	3
IV	Posterior hippocampus	N/A	-27 -32 -8	3
	Posterior temporal gyri	BA 37	-55 -32 -16	5

BA Brodmann area; N/A not applicable

(hippocampus) [12, 17]. Frontal lobe regions were defined using a hub and spoke manner, from the most lateral point of the lateral ventricle, into orbitofrontal, insula, inferior, middle and superior frontal areas [16]. Spearman analyses were used to correlate rating scores and volumetrics; as with the agreement analyses, left and right scores for a given region were analysed together.

Validation analyses: VBM

VBM an automated procedure for comparing relative grey matter, white matter or CSF density between groups of subjects on a voxel-by-voxel basis has been used in a variety of settings, including dementia [18–21] and ageing [22]. Typically, the comparison is between groups of controls and subject population. For our purposes, however, groups were formed according to rating scores for each region (see below). The expectation was that groups rated as having milder atrophy would show also less atrophy on VBM.

Pre-processing was undertaken to form grey matter images in Talairach space [23] by means of an optimised VBM protocol involving the creation of customised *a priori* probability density images of white matter, grey matter and CSF, followed by an iterative procedure of segmentation and normalisation [22]. Images were smoothed with a 12 mm Gaussian kernel in order to reduce the individual variation in gyral anatomy. The normalised grey matter probability density images produced for each subject allowed comparison of relative regional grey matter concentrations between the groups. As ratings were performed independently on the left and right hemispheres,

the brains were flipped left-to-right to double the sample size for each regional analysis ($n=72$).

Since we had a strong prior hypothesis that we would find atrophy at the location of rated regions, we used a small volume correction to correct for multiple comparisons [24] (see Table 1 for coordinates and sphere radius). Corrected probabilities in statistical parametric mapping software (SPM99) are based upon random field theory. This estimates the number of independent measurements (resels) occurring in a given search volume (there will be fewer resels than voxels due to correlation between nearby voxels and smoothing effects). The resulting statistical correction is, appropriately, less conservative than Bonferroni correction based on the number of voxels. Corrected thresholds for VBM may be misrepresented, however, when variation in spatial density of resels over the brain is not accounted for. We avoided this problem by repeating the analyses calculating separately the number of resels in the small volume studied for each analysis.

Fifteen different grey matter concentration analyses were performed, one for each region, with each image placed in a group corresponding to its rating, 0–4. If there were fewer than three images in the highest rating category, these were included at the next lowest point on the scale (e.g. for temporal pole—rating 0, $n=21$; rating 1, $n=23$; rating 2 $n=19$; rating 3 or 4, $n=9$). Analysis of variance (ANOVA) was performed to look for significant differences between the groups' mean concentration of grey matter with one-tailed *t* tests to confirm the direction of main effects shown by ANOVA. For the latter, statistical power was increased by including all scans in each test. Hence, group 0

was compared with groups 1, 2, 3 and 4; groups 0 and 1 were compared with groups 2, 3 and 4; groups 0, 1 and 2 were compared with groups 3 and 4 and groups 0, 1, 2 and 3 were compared with group 4. The *a priori* prediction was that, for each comparison, there should be a directed difference between the two sets of categories; lower-rated groups should have higher concentrations of grey matter. In the same manner, we undertook three further analyses of mean CSF concentrations in groups defined by hippocampal ratings (anterior, mid and posterior), in the expectation that the more mildly rated groups would have lower CSF concentrations.

Analyses by diagnostic category

Scans were categorised into diagnostic groups: controls ($n=8$), AD ($n=8$), SD ($n=9$) and bvFTD ($n=11$). ANOVAs were undertaken across all patient groups and all regions. Rating data across all regions were compared by Kruskal–Wallis tests, with *post hoc* Mann–Whitney tests. Discriminant analysis was used to determine which regional ratings were most important in distinguishing between the syndromes. Further, ANOVA was also undertaken on the VBM data with cases categorised by diagnosis. A note was made of the location of the most significantly different voxel and corresponding cluster size; coordinates of regions that had been found to differ between groups on rating score were also examined.

Statistics

Weighted kappas were calculated using Analyse-It (Analyse-It Software Ltd., Leeds, UK), an add-on for Microsoft Excel. VBM data processing was performed in SPM99 (<http://www.fil.ion.ucl.ac.uk/spm>) using Matlab 6.5 (Mathworks, MA, USA). SPSS 11.0 was used for all other analyses (SPSS Inc., Chicago, IL, USA).

Results

Agreement analyses: between raters

Inter-rater reliability data are given in Table 2. The mean weighted Cohen's kappa of 0.71 demonstrates significant

overall agreement between raters ($p<0.05$). This value compares favourably to those reported for other rating scales, for instance, a kappa value between two raters for a visual measure of hippocampal atrophy of 0.63 [3]. The kappa (0.71) can also be interpreted as demonstrating a “good” level of agreement between raters [25]. The ICC was calculated separately for each of the different subject groups: normal controls, 0.64; AD, 0.53; SD, 0.81; and bvFTD, 0.82 ($p<0.05$ for all). The lower level of agreement is a reflection of less variance in the rating data for the former two groups. The ICC was also used to compare agreement across regions: the region of least agreement was the anterior cingulate (ICC=0.71), while the region of best agreement was the anterior hippocampus (ICC=0.90) ($p<0.05$ for all correlations). These results indicate significant agreement between raters across all subject groups and brain regions studied.

Agreement analyses: within-rater

Intra-rater reliability analyses gave the raw proportion of ratings in absolute agreement for rater 1 across sessions of 0.75; the unweighted Cohen's kappa statistic was 0.64 and the weighted Cohen's kappa statistic 0.75 ($p<0.05$ for all). Again, these results demonstrate significant intra-rater reliability and compare favourably to those reported by previous visual ratings scales [3]. The ICC across sessions for rater 1 varied between 0.58 and 0.86 for the different subject groups ($p<0.05$ for all), indicating significant intra-rater agreement.

Validation analyses: manual volumetrics

The correlations between the visual ratings (rater 1) and the temporal and frontal lobe measures generated by volumetrics are summarised in Table 3: correlations were significant across all regions, indicating good agreement between the visual rating scale and independent methods of volume estimation for all regions studied.

Validation analyses: VBM

Figure 2 shows the results for the temporal pole. Significant differences ($p<0.05$ corrected) between the rating categories were found for 12 out of the 15 regions (all except

Table 2 Raw absolute agreement and Cohen's kappa values (unweighted and weighted) for all pairs of raters ($*p<0.05$).

	Rater 1 versus rater 2	Rater 1 versus rater 3	Rater 2 versus rater 3	Mean
Raw absolute agreement	0.69*	0.76*	0.70*	0.71*
Cohen's kappa (unweighted)	0.55*	0.66*	0.57*	0.59*
Cohen's kappa (weighted)	0.69*	0.76*	0.70*	0.71*

Table 3 Correlations between ratings for rater 1 and regional brain volumes obtained by manual volumetrics ($*p<0.05$).

	Region	Spearman's rho
Temporal lobe	Temporal pole	0.55*
	Perirhinal cortex	0.50*
	Entorhinal cortex	0.61*
	Anterior hippocampus	0.60*
Frontal lobe	Orbitofrontal cortex	0.80*
	Insular cortex	0.71*
	Dorsolateral prefrontal cortex	0.91*
	Anterior cingulate cortex	0.62*

posterior hippocampus, superior and posterior temporal). In ten of these 12 regions, the predicted direction of differences was confirmed in all analyses (0 versus 1, 2, 3, 4; 0, 1 versus 2, 3, 4 etc). The exceptions were anterior basal ganglia and anterior cingulate cortex. For both of these, however, comparison between “normal” (rated 0) and “abnormal” (rated 1, 2, 3 or 4) did show a significant difference in the predicted direction. For the regions in which we did not find significant differences at the pre-determined coordinates (posterior hippocampus, superior

and posterior temporal), significant clusters were, however, found in the vicinity of those coordinates. Furthermore, complementary analyses using the hippocampal small-volume correction spheres with ANOVA based on CSF, rather than grey matter, concentration yielded significant results (for all analyses except 0, 1, 2 versus 3, 4 for posterior hippocampus), showing that ventricular size increases with severity of hippocampal rating. The antero-posterior gradient of atrophy is clearly evident in the visual representation of analysis of cases categorised by anterior, mid and posterior hippocampal ratings (Fig. 3).

Analyses by disease category

Comparison of rating data across disease groups showed significant group effects for all regions ($p<0.001$). *Post hoc* pairwise comparisons between controls and each of the disease groups confirmed significant differences in all instances (i.e. significantly lower ratings for controls). Corresponding analyses between disease groups gave significant differences ($p<0.05$) as follows (Table 4): (1) AD versus SD—temporal pole, anterior parahippocampal gyrus, anterior fusiform gyrus and lateral temporal gyri (all worse in SD); (2) AD versus bvFTD—orbitofrontal gyri,

Fig. 2 Result of VBM ANOVA with cases categorised by temporal pole ratings—glass brains showing voxels differing significantly in grey matter concentration across rating groups

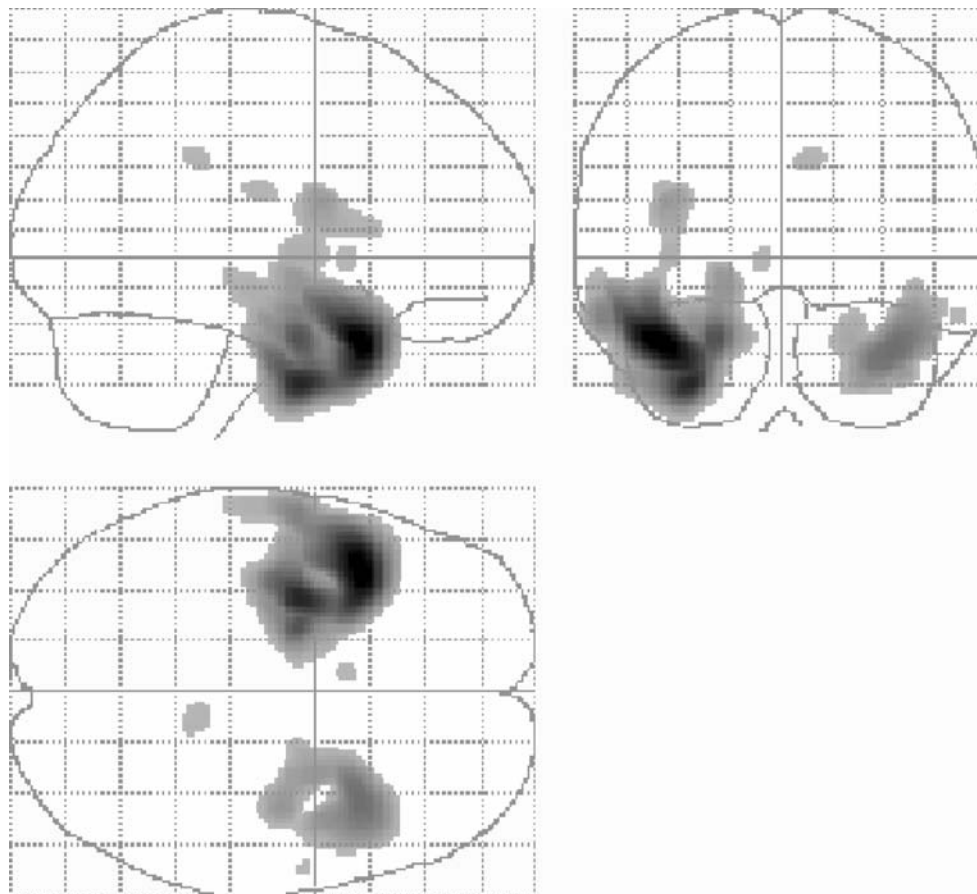
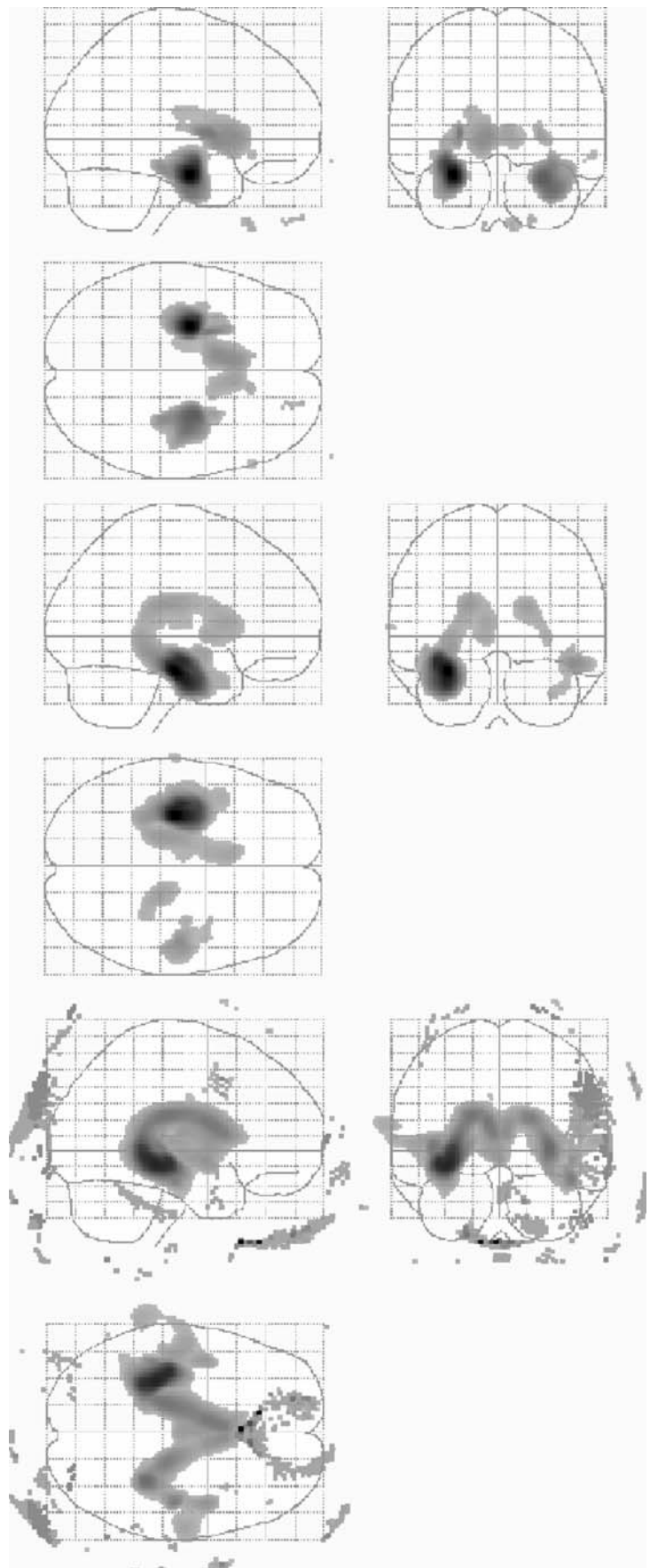


Fig. 3 Result of VBM ANOVA with cases categorised by ratings for anterior hippocampus, mid-hippocampus and posterior hippocampus—glass brains showing voxels differing significantly in CSF concentration across rating groups. Note antero-posterior axis of increased CSF; also note extracerebral artefact in posterior hippocampal analysis



anterior parahippocampal gyrus and lateral temporal gyri (all worse in bvFTD); and (3) SD versus bvFTD—temporal pole and anterior fusiform (worse in SD) and the anterior cingulate (worse in bvFTD).

To explore the key regions that differentiated subject groups, we used discriminant analyses and showed differences as follows: (1) controls versus AD—insula, anterior hippocampus, orbitofrontal gyri and temporal pole; (2) controls versus SD—anterior fusiform; (3) controls versus bvFTD—insula; (4) AD versus SD—temporal pole, anterior hippocampus and lateral temporal (all worse in SD); (5) AD versus bvFTD—orbitofrontal gyri, anterior hippocampus and lateral frontal cortex (all worse in bvFTD); (6) SD versus bvFTD—anterior cingulate (worse in bvFTD) and temporal pole (worse in SD).

VBM analyses by disease group showed the most significant reductions in grey matter concentration at the following locations (Table 4): (1) AD minus controls—posterior cingulate cortex (BA31) ($p=0.000$, cluster=475 voxels); (2) SD minus controls—anterior fusiform (BA36) ($p=0.000$, cluster=4,608 voxels); (3) bvFTD minus controls—lateral orbitofrontal gyrus (BA10) ($p=0.000$, cluster=2,478 voxels); (4) SD minus AD—anterior fusiform gyrus (BA36) ($p=0.000$, cluster=437 voxels); (5) bvFTD minus AD—medial prefrontal gyri (BA9/10) ($p=0.003$, cluster=25 voxels) and (6) bvFTD versus SD, bvFTD minus SD—white matter underlying lateral frontal gyri, SD minus bvFTD—white matter underlying posterior insula ($p=0.000$, 0.000; clusters=329, 89 respectively). No further voxels in regions deemed to differ between the subject groups by rating survived a corrected threshold of 0.05.

Discussion

We describe a visual rating method that provides anatomical data for 15 brain regions. By means of a combination of written criteria and reference images, we obtained favourable reliability data for all 15 regions, across raters of varied baseline expertise thus fulfilling the primary aim of the study.

Validation of the rating data against independent structural data also gave favourable results. Available volumetric data in eight regions spanning temporal and frontal lobes correlated significantly with rating scores. To the best of our knowledge, validation of a rating scale against VBM is novel. Significant differences in grey matter concentration at pre-defined locations were found across groups, categorised by rating severity, for 12 of the 15 regions. For the remaining three regions (e.g. posterior hippocampus), validation by VBM may have failed because of distance to grey matter boundaries, which are readily detected by VBM.

Analysis of the rating data by subject group was largely in keeping with the literature on the various syndromes represented in our sample of 36 cases and the finding that all regions were rated significantly differently from controls in all three disease groups reinforces the favourable validity analyses. Discriminant analyses of control data and patient groups showed the anterior fusiform rating to be the region in distinguishing SD from controls [3, 10] and the insula to be key in discriminating bvFTD [21, 26]. Multiple regions were relevant in discriminating the AD from controls (insula, anterior hippocampus, orbitofrontal gyri and temporal pole) in keeping with the more diffuse atrophy seen in AD. The key regions differentiating AD and SD were the temporal pole, anterior hippocampus and lateral temporal gyri [3, 4, 10, 19]. The SD versus bvFTD discrimination was based on the temporal pole being more atrophic in the former and the anterior cingulate being more atrophic in the latter, in keeping with hypothesised roles for these regions respectively in semantics and social cognition [4, 18, 27, 28]. Direct comparison of regional ratings between the syndromes paralleled the discriminant analyses.

The key voxels in the VBM analyses of the controls versus each of the three disease groups were located in regions previously implicated in the respective syndromes—posterior cingulate cortex in AD [29], anterior fusiform in SD [3, 10] and lateral orbitofrontal gyrus in bvFTD [20]. The anterior fusiform was also highlighted in the AD-versus-SD analysis. The remaining whole brain analyses comparing

Table 4 Regions of significant relative atrophy identified by rating and by VBM.

Groups compared	Atrophied regions detected by RATING	Atrophied regions detected by VBM
AD versus controls	All rated regions	Posterior cingulate
SD versus controls	All rated regions	Anterior fusiform
bvFTD versus controls	All rated regions	Lateral orbitofrontal gyri
SD versus AD	Temporal pole anterior parahippocampal gyrus anterior fusiform gyrus lateral temporal gyri	Anterior fusiform
bvFTD versus AD	Orbitofrontal gyri anterior parahippocampal gyrus lateral temporal gyri	Medial prefrontal gyri
SD versus bvFTD	Temporal pole anterior fusiform gyrus	Lateral frontal gyri
bvFTD versus SD	Anterior cingulate	Posterior insula

disease groups proved difficult to interpret. Notably, neither the temporal pole nor hippocampus was found to differ between controls and any subject group by VBM although differences were manifest on simple visual inspection. By contrast, the posterior cingulate proved difficult to rate (this region was removed from the rating protocol in pilot phase) but was shown by VBM to differ between AD and controls in accordance with recent evidence on the status of the posterior cingulate in AD [29, 30].

Logistical considerations significantly limit the application of volumetric and automated methods, including VBM. Such data are also open to more fundamental criticism. Firstly, individual case data is, of necessity, distorted during the processing required for automated analysis; indeed, the application of VBM to single cases may be fundamentally flawed. Secondly, certain automated analyses may be inherently insensitive to changes in particular structures, such as the hippocampus [7]. Thirdly, gross distortion of anatomy in some diseases may interfere with the application of automated imaging software and, to a lesser extent, volumetric protocols. Visual inspection, by its nature, provides a holistic assessment. A previous VBM study in SD, for instance, failed to show hippocampal atrophy [19] as did the VBM undertaken here, although atrophy was clearly evident on volumetric and visual inspection methods [3, 10]. Volumetric protocols for cerebral cortex typically aim to delineate entire regions but are crucially dependent on gross anatomical landmarks which vary in position from case to case, making some error inevitable. In taking *post mortem* tissue specimens, by contrast, the approach is usually to sample a circumscribed area well within recognised boundaries. The rating approach, it may be argued, is analogous. Rating therefore lends itself to comparison of *in vivo* and *post mortem* appearances for longitudinal assessment [31].

The disadvantages of the rating approach are transparent. Rating will not give precise volumes. Rating is also prone to sampling error in that only one view of a structure is usually assessed. Rating can be further criticised for lacking objectivity although this can be improved by rigorous blinding and by the use of specific rating criteria. Logistical advantages of rating are equally clear. It is quick and widely applicable, requiring little training and no special equipment. For certain purposes, we argue that validated visual rating methods may be preferable both to descriptive reporting and to attempts at more precise quantification.

We hope that this rating system will find widespread application in studies of brain structure–function relationships. It is particularly applicable to single-case and small-group studies when the variabilities in volumetric analyses limit their use and in correlative studies when there is gross brain destruction.

Acknowledgements We are grateful for the practical support of Addenbrooke's Hospital MRI Department and that of the Wolfson Brain Imaging Centre. RD and AG were funded by Wellcome Trust Clinical Training Fellowships; RD is also in receipt of a Sackler Fund Award. KG holds an Alzheimer's Research Trust Programme Grant; JRH was funded by the Medical Research Council and is now supported by an Australian Research Council Federation Fellowship (FF0776229).

Conflict of interest statement We declare that we have no conflict of interest.

References

1. Korf ES, Wahlund LO, Visser PJ, Scheltens P (2004) Medial temporal lobe atrophy on MRI predicts dementia in patients with mild cognitive impairment. *Neurology* 63:94–100
2. Fan YH, Lam WW, Mok VC, Huang RX, Wong KS (2003) Variability and validity of a simple visual rating scale in grading white matter changes on magnetic resonance imaging. *J Neuroimaging* 13:255–258. doi:10.1177/1051228403013003009
3. Galton CJ, Gomez-Anson B, Antoun N, Scheltens P, Patterson K, Graves M, Sahakian BJ, Hodges JR (2001) Temporal lobe rating scale: application to Alzheimer's disease and frontotemporal dementia. *J Neurol Neurosurg Psychiatry* 70:165–173. doi:10.1136/jnnp.70.2.165
4. Davies RR, Graham KS, Xuereb JH, Williams GB, Hodges JR (2004) The human perirhinal cortex and semantic memory. *Eur J Neurosci* 20:2441–2446. doi:10.1111/j.1460-9568.2004.03710.x
5. Laakso MP, Frisoni GB, Kononen M, Mikkonen M, Beltramello A, Geroldi C, Bianchetti A, Trabucchi M, Soininen H, Aronen HJ (2000) Hippocampus and entorhinal cortex in frontotemporal dementia and Alzheimer's disease: a morphometric MRI study. *Biol Psychiatry* 47:1056–1063. doi:10.1016/S0006-3223(99)00306-6
6. Ashburner J, Friston KJ (2000) Voxel-based morphometry—the methods. *Neuroimage* 11:805–821. doi:10.1006/nimg.2000.0582
7. Good CD, Schill RI, Fox NC, Ashburner J, Friston KJ, Chan D, Crum WR, Rossor MN, Frackowiak RS (2002) Automatic differentiation of anatomical patterns in the human brain: validation with studies of degenerative dementias. *Neuroimage* 17:29–46. doi:10.1006/nimg.2002.1202
8. Arnaiz E, Jelic V, Almkvist O, Wahlund LO, Winblad B, Valind S, Nordberg A (2001) Impaired cerebral glucose metabolism and cognitive functioning predict deterioration in mild cognitive impairment. *Neuroreport* 12:851–855. doi:10.1097/00001756-200103260-00045
9. Scheltens P, Leys D, Barkhof F, Huglo D, Weinstein HC, Vermersch P, Kuiper M, Steining M, Wolters EC, Valk J (1992) Atrophy of medial temporal lobes on MRI in “probable” Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *J Neurol Neurosurg Psychiatry* 55:967–972. doi:10.1136/jnnp.55.10.967
10. Chan D, Fox NC, Schill RI, Crum WR, Whitwell JL, Leschziner G, Rossor AM, Stevens JM, Cipolotti L, Rossor MN (2001) Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Ann Neurol* 49:433–442. doi:10.1002/ana.92
11. Duvernoy H (1991) The human brain: surface, three-dimensional sectional anatomy and MRI. Springer, New York
12. Insausti R, Juottonen K, Soininen H, Insausti AM, Partanen K, Vainio P, Laakso MP, Pitkanen A (1998) MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *AJNR Am J Neuroradiol* 19:659–671

13. Paxinos G (1990) The human nervous system, vol. 1, 2nd edn. Academic, San Diego
14. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34:939–944
15. Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, Freedman M, Kertesz A, Robert PH, Albert M, Boone K, Miller BL, Cummings J, Benson DF (1998) Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* 51:1546–1554
16. Perry RJ, Graham A, Williams G, Rosen H, Erzinclioğlu S, Weiner M, Miller B, Hodges J (2006) Patterns of frontal lobe atrophy in frontotemporal dementia: a volumetric MRI study. *Dement Geriatr Cogn Disord* 22:278–287. doi:10.1159/000095128
17. Watson C, Jack CR Jr, Cendes F (1997) Volumetric magnetic resonance imaging. Clinical applications and contributions to the understanding of temporal lobe epilepsy. *Arch Neurol* 54:1521–1531
18. Williams GB, Nestor PJ, Hodges JR (2005) Neural correlates of semantic and behavioural deficits in frontotemporal dementia. *Neuroimage* 24:1042–1051. doi:10.1016/j.neuroimage.2004.10.023
19. Mummery CJ, Patterson K, Price CJ, Ashburner J, Frackowiak RS, Hodges JR (2000) A voxel-based morphometry study of semantic dementia: relationship between temporal lobe atrophy and semantic memory. *Ann Neurol* 47:36–45. doi:10.1002/1531-8249(200001)47:1<36::AID-ANA8>3.0.CO;2-L
20. Gorno-Tempini ML, Dronkers NF, Rankin KP, Ogar JM, Phengrasamy L, Rosen HJ, Johnson JK, Weiner MW, Miller BL (2004) Cognition and anatomy in three variants of primary progressive aphasia. *Ann Neurol* 55:335–346. doi:10.1002/ana.10825
21. Rosen HJ, Gorno-Tempini ML, Goldman WP, Perry RJ, Schuff N, Weiner M, Feiwell R, Kramer JH, Miller BL (2002) Patterns of brain atrophy in frontotemporal dementia and semantic dementia. *Neurology* 58:198–208
22. Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS (2001) A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* 14:21–36. doi:10.1006/nimg.2001.0786
23. Talairach J, Tournoux P (1998) Co-planar Stereotaxic Atlas of the human brain: 3-dimensional proportional system : an approach to cerebral imaging. Illustrated ed. Thieme, New York
24. Worsley K, Marrett S, Neelin P et al (1996) A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp* 4:58–73. doi:10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O
25. Robson C (1993) Real world research: a resource for social scientists and practitioner-researchers. Blackwell, Oxford
26. Matsuo K, Mizuno T, Yamada K, Akazawa K, Kasai T, Kondo M, Mori S, Nishimura T, Nakagawa M (2008) Cerebral white matter damage in frontotemporal dementia assessed by diffusion tensor tractography. *Neuroradiology* 50:605–611. doi:10.1007/s00234-008-0379-5
27. Anderson SW, Damasio H, Damasio AR (2005) A neural basis for collecting behaviour in humans. *Brain* 128:201–212. doi:10.1093/brain/awh329
28. Murray EA, Richmond BJ (2001) Role of perirhinal cortex in object perception, memory, and associations. *Curr Opin Neurobiol* 11:188–193. doi:10.1016/S0959-4388(00)00195-1
29. Nestor PJ, Fryer TD, Ikeda M, Hodges JR (2003) Retrosplenial cortex (BA 29/30) hypometabolism in mild cognitive impairment (prodromal Alzheimer's disease). *Eur J Neurosci* 18:2663–2667. doi:10.1046/j.1460-9568.2003.02999.x
30. Karas G, Scheltens P, Rombouts S, van Schijndel R, Klein M, Jones B, van der Flier W, Vrenken H, Barkhof F (2007) Precuneus atrophy in early-onset Alzheimer's disease: a morphometric structural MRI study. *Neuroradiology* 49:967–976. doi:10.1007/s00234-007-0269-2
31. Broe M, Hodges JR, Schofield E, Shepherd CE, Kril JJ, Halliday GM (2003) Staging disease severity in pathologically confirmed cases of frontotemporal dementia. *Neurology* 60:1005–1011