

Visual Rating versus Volumetry to Detect Frontotemporal Dementia

T.W. Chow^{a, b, e, f} F. Gao^h K.A. Links^f J.E. Ween^{a, e, g} D.F. Tang-Wai^{a, i}
J. Ramirez^h C.J.M. Scott^h M. Freedman^{a, d-f, i, j} D.T. Stuss^{a, c-f} S.E. Black^{a, d-f, h}

Departments of ^aMedicine (Division of Neurology), ^bPsychiatry (Division of Geriatric Psychiatry) and ^cPsychology and ^dInstitute of Medical Sciences, University of Toronto, ^eDivision of Neurology, ^fRotman Research Institute and ^gKunin-Lunenfeld Applied Research Unit, Baycrest, ^hL.C. Campbell Cognitive Neurology Research Unit, Sunnybrook Health Sciences Centre, ⁱToronto Western Hospital Division of Neurology, University Health Network Memory Clinic, and ^jDivision of Neurology, Mt. Sinai Hospital, Toronto, Ont., Canada

Key Words

Alzheimer's disease · Frontotemporal dementia · Magnetic resonance imaging · Visual rating

Abstract

Background/Aims: Automated, volumetrically defined atrophy in the left anterior cingulate (LAC) and anterior temporal regions (LAT) on MRI can be used to distinguish most patients with frontotemporal dementia (FTD) from controls. FTD and Alzheimer's disease (AD) can differ in the degree of anterior temporal atrophy. We explored whether clinicians can visually detect this atrophy pattern and whether they can use it to classify the 2 groups of dementia patients with the same accuracy. **Methods:** Four neurologists rated atrophy in the LAC and LAT regions on MRI slices from 21 FTD, 21 controls, and 14 AD participants. Inter-rater reliability and diagnostic accuracy were assessed. **Results:** All 4 raters agreed on the presence of clinically significant atrophy, and their atrophy scoring correlated with the volumes, but without translation into high inter-rater diagnostic agreement. **Conclusions:** Volumetric analyses are difficult to translate into routine clinical practice.

Copyright © 2011 S. Karger AG, Basel

Introduction

Although frontotemporal abnormalities on structural [1–6] and functional imaging [7–11] have been reported in frontotemporal dementia (FTD), consensus groups deemed such findings supportive and not necessary for clinical diagnosis [12, 13]. These consensus groups were convened prior to volumetric studies replicating robust differences between FTD, healthy aging, and Alzheimer's disease (AD). In a prior study [14], logistic regression modeling showed that the combination of atrophy in the left anterior cingulate (LAC) and the left anterior temporal pole (LAT) can distinguish FTD from controls with age-associated frontal atrophy with 90% accuracy. The application of these findings to routine clinical practice has not been explored. Although clinical diagnosis can be enhanced using visual ratings of coronal views of the LAT with a high correlation of the visual ratings to automated volumetric measurements [15, 16], not all clinical MRI protocols provide coronal sections.

In this study, we explored the ability of clinical neurologists experienced in dementia diagnosis and neuroimaging research to detect selective regional atrophy on 'real life' MR images. MRI studies obtained for volumet-

ric studies often contain more data and use 3D T₁-weighted images, which are not routinely acquired in clinical MRI scan protocols. Our study aim was to determine how applicable the findings from our prior study are to frontline clinical neurologists, especially those who may have Internet access (i.e. picture archive and communication systems) to patients' MRI scans.

The first study question was whether our raters could use atrophy scores to distinguish FTD from control MRIs with acceptable inter-rater reliability. We also wanted to test whether coronal views (not included in most clinical MRI brain protocols) are necessary for accurate identification of atrophy in the LAT.

It is frequently possible to distinguish healthy middle-aged patients from FTD patients with a good informant alone. We assigned an additional exercise for our raters: to use the LAC and LAT atrophy pattern to solve the more common clinical dilemma of differentiating FTD from AD [17].

Materials and Methods

Participants

We used MRI data from 21 participants diagnosed with FTD according to criteria published from a consensus meeting [13], 14 participants with AD per NINCDS-ADRDA criteria [18] matched on the basis of duration of illness to the FTD group, and 21 age- and gender-matched controls. All participants were recruited from memory clinics at Sunnybrook Health Sciences Centre (S.E.B.) and Baycrest (T.W.C. and M.F.). Ethics committees at both institutions approved the protocols. The inclusion and exclusion criteria for the study and description of sample have already been detailed in the prior study using high-resolution volumetric data [14].

In brief, all participants underwent a standardized MRI of the brain between 1995 and 2004 on the same 1.5-T MRI scanner (General Electric Signa version 8.4M4), using the dementia protocol at the Sunnybrook Health Sciences Centre. Scans were typically performed within 3 months of clinical diagnosis of FTD or AD, with the exception of 4 cases of FTD who had their first MRI for research purposes 1–2 years after diagnosis. Findings of atrophy did not bias inclusion or exclusion for this study, and MRI results contributed to the clinical diagnoses made for study inclusion. Autopsy data were used to exclude potential participants from the study.

Autopsies confirmed FTD diagnoses in 10 participants from this group; the other 11 are still living or did not give consent for autopsy. Comparing the autopsy-confirmed group against the remainder of FTD participants revealed a higher level of education (mean 15.6 vs. 11.5 years, $t = 2.36$, $p = 0.03$), higher proportion of left-handers (8/10, $\chi^2 p = 0.004$), and smaller mean LAT volume (6.8 vs. 9.64 cm³, $t = -3.02$, $p = 0.007$) but not LAC volume for the autopsied group, and no differences in sex distribution ($\chi^2 p = 0.63$), mean age at time of MRI (63.9 vs. 64 years, $t = -0.29$, $p = 0.98$), mean duration of illness (4.3 vs. 4.0 years, $t = 0.3$, $p = 0.77$) or mean MMSE score (17.7 vs. 23.6, $t = -1.45$, $p = 0.16$).

The control group consisted of 21 healthy community volunteers without complaints of mood or behavioral disturbance, not taking psychotropic medications, and with no history suggestive of neurological or psychiatric conditions. They were matched to each of the 21 FTD participants by gender and age within 5 years.

The FTD sample for this study included 20 of the 30 FTD participants described in our prior study [14] plus 1 additional participant, and the same 18 controls, plus 3 additional control participants.

Imaging Analysis

T₁-weighted (1.2-mm slice thickness) and proton density/T₂-weighted (3-mm slice thickness) imaging parameters were selected to provide optimal intensity separation for the in-house previously published segmentation and parcellation program, SABRE [19, 20].

The LAC and LAT volumes of grey matter were normalized by dividing each volumetric component by the same individual's total supratentorial intracranial volume, in order to accommodate for normal variation in head size. The measure was then multiplied by the mean total supratentorial intracranial volume for the control sample (1,186.19, $n = 21$) to yield a sample-normalized volume in cm³ for each volume of interest (VOI) [14].

Visual Rating

A research neuroradiologist expert in dementia imaging (F.G.) used AnalyzeTM-formatted MRI data aligned to the anterior commissure and posterior commissure plane to select 5 slices that simulated the clinical format for each of the 56 participants and highlighted the LAC and LAT VOIs, based on our prior volumetric method that identified those VOIs as particularly discriminating between FTD and controls [21]. Slice simulations included: (1) the T₁-weighted axial slice just superior to the last view of insular cortex on either side of the brain (LAC), (2) the sagittal T₁-weighted slices with the most complete view of the left optic nerve on either side of the brain (LAC), (3) the most inferior T₁-weighted axial slice showing the interpeduncular cistern (LAT), (4) the sagittal T₁-weighted slice at the lateral extreme of the left hippocampus, and (5) coronal T₁-weighted slice at the level of the limen insulae, where insular cortex becomes continuous with frontal cortex (LAC and LAT; fig. 1). The volume averaging on the slices viewed from the 3D T₁ MRI generated for this exercise differed from routine clinical scans (which usually have thicker slices of approx. 5–10 mm). However, slice thickness should not affect accuracy of the atrophy rating, as the atrophy of the LAC and LAT was determined on 3 sections from different orientations. The standardization procedure for this study's slices should have enhanced reproducibility.

Although not duplicating a naturalistic setting, to review all participants in only 2 sessions, all 4 visual raters viewed the timed slideshow simultaneously. Each rater completed 1 rating sheet per participant independently. The raters were all neurologists involved in clinical research, not certified in neuroradiology. Two of the 5 images were displayed at a time, sequentially, in the same order (LAC then LAT) from participant to participant in a Powerpoint slide presentation programmed to show each slide (image pair) for 10 s. This amount of time was more than the raters needed to score atrophy and was set at this length to decrease the confound of any inter-rater differences that could be attributed to processing speed. Participants were presented with diagnosis or-

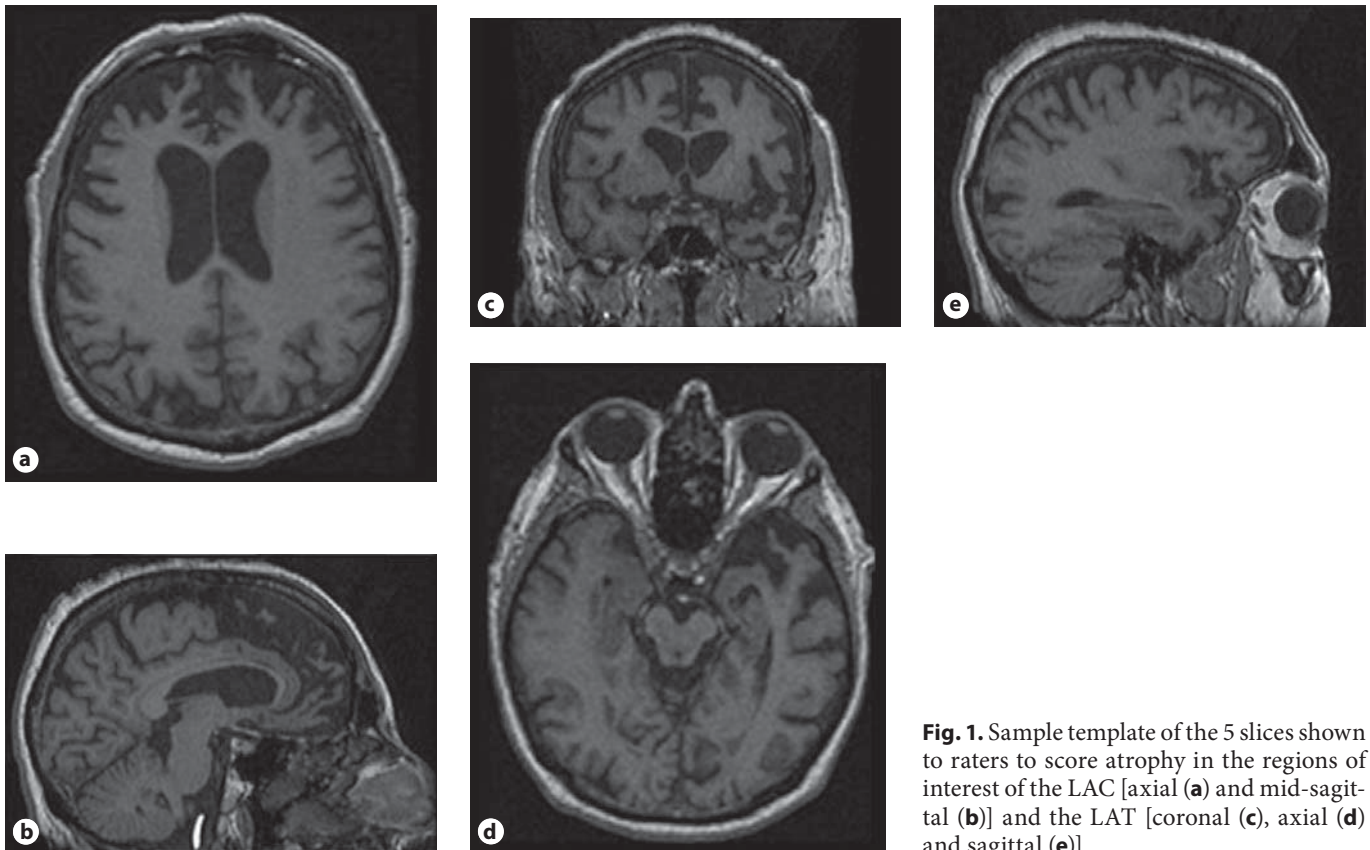


Fig. 1. Sample template of the 5 slices shown to raters to score atrophy in the regions of interest of the LAC [axial (a) and mid-sagittal (b)] and the LAT [coronal (c), axial (d) and sagittal (e)].

der (i.e. FTD, AD, control) randomized. This way, the 4 visual raters could make their ratings of the timed slide show presentation without bias from diagnosis. During the rating sessions, raters understood that approximately one third of the scans were from control participants, but were blinded to all clinical information about the other participants, including age.

Each rating sheet showed standard images for each atrophy rating, and scores ranged from 0, indicating no atrophy, to 4, the most severe atrophy, similar to the methods of Davies et al. [16]. The standard images were derived from a single control and a single participant with FTD, neither of whom were part of the study sample, so that the raters would not be able to simply match some images from the slideshow to the rating sheet. For intrarater reliability of atrophy scoring, the raters re-reviewed a subsample of 18 participants randomly chosen from the initial group of 56 during a second scoring session.

A diagnosis was selected immediately after viewing each 5-image set during the first rating session, without being able to return to any set once completed.

Debriefing after the first session revealed a need to clarify the diagnostic criteria as raters had differed on use of non-LAC and non-LAT regions to make these decisions.

Raters were instructed that the main criterion for an FTD diagnosis was presence of atrophy most marked in the LAC and/or LAT regions. Hippocampal and parietal atrophy equivalent to that seen in the LAC and/or LAT would lead a rater towards an AD diagnosis.

No demographic information on participants was provided on the slides presented, to emphasize the weight to be placed on regional atrophy, as done with SABRE in the previous study [14]. In the second session, raters stated they would have used participant ages to inform their diagnostic impressions if those data had been available.

Statistical Analysis

We subjected the raters' 0–4 atrophy scores to Spearman's rho correlations with the corresponding LAC and LAT volumes yielded by SABRE, to give a measure of raters' accuracy apart from choosing a diagnosis.

Inter-Rater Agreement: Analysis of Visual Ratings

We ran Kendall's W for weighted κ statistics to test inter-rater agreement for the 0–4 atrophy scores from each of the 4 visual raters.

We also captured a simple presence versus absence of atrophy response for LAC and LAT. We subjected these responses to Cochran's Q statistical analysis, in the event this measure had a different reliability from the atrophy scoring. The Q statistic for agreement yields larger p values (i.e. $p > 0.05$) with increasing inter-rater agreement [22].

To determine the added impact of coronal views for designation of clinically significant atrophy in the LAT, we conducted a binary logistic regression for each rater's score set. The three

Table 1. Demographics and VOIs (means \pm SD)

	FTD (n = 21 ¹)	AD (n = 14)	Controls (n = 21)
Age, years	62.8 \pm 8.1	61.4 \pm 8.1	63.3 \pm 7.5
Men	7 (33.3%)	6 (42.9%)	6 (28.6%)
Mean education level, years	13.7 \pm 4.4	12.7 \pm 3.1	15.5 \pm 2.4
Mean duration of illness, years	4.1 (range = 1–11)	3.8 (range = 1–7)	n.a.
Mean MMSE score	22.0 (n = 20)	24.1	28.9 (n = 16)
Left anterior cingulate, cm ³	23.4 \pm 2.7	26.5 \pm 2.47	26 \pm 3.6
Left anterior temporal, cm ³	8.3 \pm 2.5	10.8 \pm 1.5	10.3 \pm 3.2

¹ Including 3 with non-fluent progressive aphasia, 5 with semantic dementia, and 13 with behavioral variant FTD.

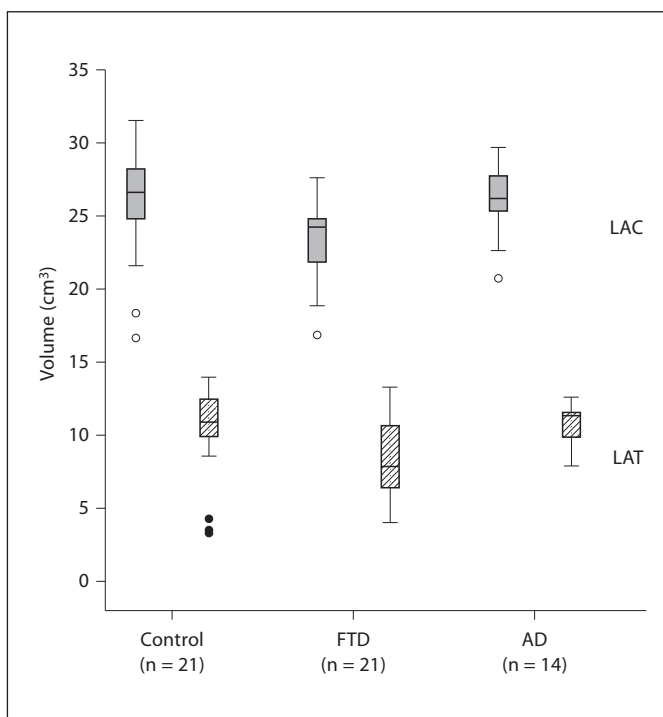


Fig. 2. Boxplots of median values of LAC and LAT volumes for each of the 3 participant groups show that participants with FTD generally had the smallest volumes, but outliers within the control group had lower volumes than participants with dementia.

planes of view for LAT (axial, sagittal and coronal) were entered in that order for a binary logistic regression, with the atrophy response (yes/no) as the dependent variable. The lack of contribution of the coronal atrophy score to the model for atrophy would indicate that the rater would have labeled the LAT as atrophied regardless of whether the coronal slices were available to view along with standard axial and sagittal slices.

Overall Diagnostic Accuracy

The clinical consensus-based diagnoses used to select the sample were considered the gold standard for comparison. We calculated the percentage accuracy by dividing the tally of the unanimously correctly identified FTD and AD participants by the total number of patients with FTD or AD (n = 35). Because our original study reported significant VOIs for FTD versus control, we performed a χ^2 test for any significant difference between the accuracies of FTD versus control and FTD versus AD diagnoses.

SPSS software version 16.0 (for Windows) was used for all statistical analyses.

Results

Although 18 participants with AD met the inclusion criteria by being duration-of-illness-matched to the FTD group and had undergone MRI of the brain, motion or dental implant artifacts rendered the images unreadable by the volumetrics software in 4 participants. The AD group therefore had 14 members in this study. The participants with AD, while selected for duration-of-illness matches to the FTD participants, as opposed to age at the time of the MRI, happened to have early-onset ages. Demographic data for the groups appear in table 1.

The FTD and AD groups were in a mild stage of illness near the time of MRI scanning, based on Mattis Dementia Rating Scale scores close to 118. Mean MMSE scores for the 2 dementia groups did not differ ($p = 0.324$).

Figure 2 depicts the volumes of the LAC and LAT for each of the 3 groups.

Accuracy of Visual Rating

Atrophy scores (0–4) for LAC in the sagittal view, LAT in the coronal view, and LAT in the axial view by all 4 raters correlated appropriately (inversely) with actual vol-

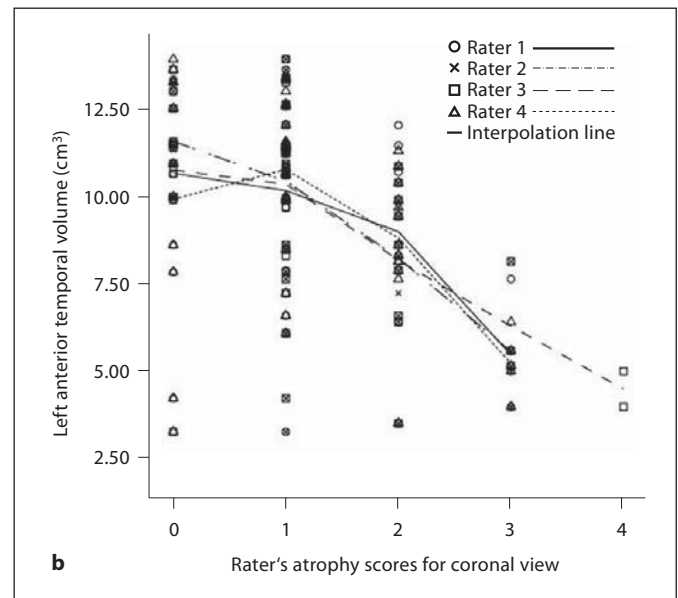
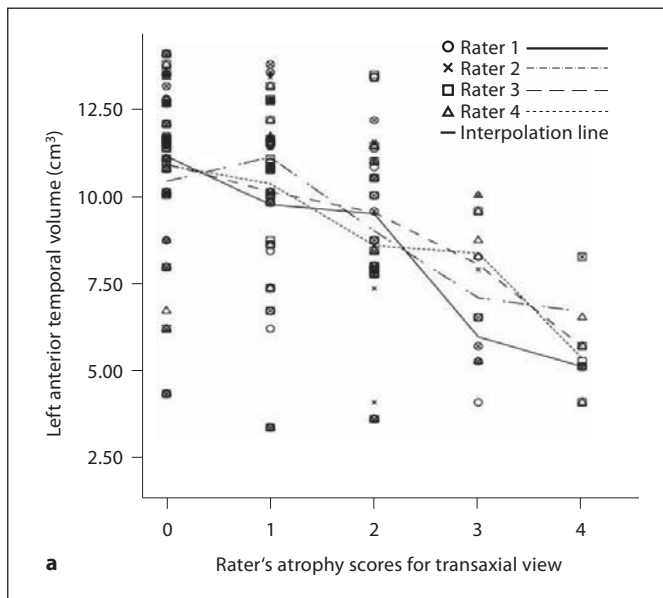


Fig. 3. Scatter plots of atrophy scores from each rater against the actual volumes indicates that neither axial (**a**) nor coronal (**b**) views prevented disagreement about whether the VOI is atrophied. Note markers corresponding to approximately 4 cm³ volumes, not all of which were considered to be even moderately at-

rophied by some raters. Control n = 21, FTD n = 21, AD n = 14. Scoring key: 0 = normal; 1–2 = mild deepening of sulci, could be normal for participants aged >60 years; 3 = moderate atrophy beyond that consistent with age of participant; 4 = severe atrophy, e.g. 'paper-thin' cortex.

umes. More severe atrophy indicates smaller volume, rho ranging among individual raters from -0.61 to -0.42 , p values <0.01 .

Inter-Rater Agreement

Inter-rater agreement for the 0–4 severity of atrophy was high for both LAC and LAT, yet inconsistent between axial, sagittal, or coronal views. Scoring was more uniform among raters for atrophy in the LAC on axial views (Kendall's $W = 0.062$, $p = 0.016$) than sagittal views (larger Kendall's $W = 0.070$, $p = 0.008$). This was true regardless of whether the raters focused on cortical atrophy or ventricular enlargement. LAT atrophy scoring was in high agreement between raters both when looking at a coronal view and when examining an axial view but significantly reduced when raters were asked to score ventricular enlargement (Kendall's $W = 0.204$, $p < 0.00001$) in any of these orientations.

Inter-rater agreement was strong for the presence versus absence of atrophy in LAC ($Q = 5.769$, $p = 0.123$), as observed with scoring for this VOI. Matching of atrophy response among raters was not close for the LAT VOI (much higher Q of 32.124 , $p = 4.866 \times 10^{-7}$). This would have contributed to discrepancies among raters for distinguishing FTD from controls, because the combined

presence of LAC and LAT atrophy were used to make an FTD designation.

Binary logistic regressions for LAT atrophy responses from any of the raters using the axial, then sagittal, then coronal view atrophy scores as dependent variables showed that none of the views singularly created a model for prediction of atrophy. We temper our interpretation of this finding, because raters could review all slice ratings for a participant for about 1 min before the next set of images appeared. This could potentially allow the coronal views to influence scoring of atrophy on the axial and sagittal slices. Figure 3 shows rater-by-rater scores on axial and coronal views shown against actual volumes.

Accuracy of Differential Diagnosis

Accuracy for categorizing MRI scans as control, FTD, or AD was low in this study. Unanimous rater agreement for the correct diagnoses occurred in no more than one-third of each diagnostic group: 3/21 controls, 9/21 FTD, and 5/14 AD. Six out of 10 (67%) autopsy-confirmed FTD and more non-autopsied FTD participants (7 out of 11, 77%) received an accurate diagnosis from 3 or more of the 4 raters. The raters were able to correctly distinguish between FTD and AD scans with an average of 63.1% accuracy.

Table 2. Demographic, cognitive, functional, and behavioral inventory scores for autopsy-confirmed FTD mistaken for AD by 3 raters, the rest of the FTD sample, and the AD sample

	FTD mistaken for AD (n = 2)	Correctly identified FTD (n = 19)	AD	ANOVA p value
Age at time of scan, years	66.5 ± 14.1	63.7 ± 7.4	61.4 ± 8.5	0.7
Duration of illness, years	1.0 ± 0	4.5 ± 2.4	3.8 ± 2.9	0.06
Duration of education, years	14.0 ± 1.4	13.4 ± 1.1	12.2 ± 3.9	0.86
MMSE	25.0 ± 7.1	21.7 ± 9.2 (n = 18)	24.1 ± 2.8	0.57
Dementia Rating Scale (total)	118.5 ± 10.6	110.4 ± 32 (n = 7)	122.4 ± 10.6	0.43
Cornell Depression Scale	10.5 (n = 1)	19.7 ± 11.8 (n = 9)	17.5 ± 18.6	0.85
BPADRS (total)	4 ± 2.8	4.1 ± 4.3 (n = 9)	3.5 ± 5.1	0.95

Means ± SD are presented. BPADRS = Behavioral Pathology in Alzheimer's Disease Rating Scale.

Unanimous agreement for incorrect diagnoses by all 4 raters occurred in 2 cases (1 control mistaken for AD and 1 FTD mistaken for AD). FTD scans were mistaken as 'AD' by 3 of the 4 raters in 2 of the 21 cases. These 2 had autopsy-confirmed diagnoses of FTD, 1 of whom had motor neuron disease and died only 1 year into illness. Their mean LAC and LAT volumes landed right between the remainder of the FTD sample (n = 19) and the AD group despite this short duration of illness (mean LAC volume 25.4 ± 3.2, 23.2 ± 2.6, and 26.5 ± 2.5 cm³, and LAT 9 ± 2.4, 8.2 ± 2.6, and 10.8 ± 1.5 cm³, respectively). Post hoc contrasts showed LAC and LAT to be smaller in accurately diagnosed FTD (n = 19) than in AD: LAC (F = 5.49, p = 0.009; post hoc FTD vs. AD contrast p = 0.002) and LAT (F = 6.58, p = 0.004; post hoc FTD vs. AD contrast p = 0.001). Table 2 shows that there were no significant differences between the 2 mistaken FTD cases, the rest of the FTD group, and the AD group in demographics or functional/behavioral assessments, but this conclusion is limited by the unavailability of Dementia Rating Scale, Cornell Depression Scale, and BEHAVE-AD scores for half of the larger FTD group.

Only 1 AD scan was mistaken for FTD by more than 2 raters.

The average individual rater accuracy for controls versus FTD was 59.5%. Rater 1 was better at differentiating FTD versus control ($\chi^2 = 12.1$, p < 0.001), while raters 2–4 had higher accuracy making FTD versus AD discriminations than FTD versus controls (rater 2, $\chi^2 = 9.5$, p < 0.01; rater 3, $\chi^2 = 10.7$, p < 0.01; rater 4, $\chi^2 = 6.1$, p < 0.05). Despite that, rater 4's accuracy for FTD versus AD was lower than the rest of the group (23.8% for FTD, 64.3% for AD; p < 0.02).

Discussion

The 4 raters showed strong inter-rater agreement for recognizing and scoring atrophy. In the present study, a determination of moderate or worse atrophy of the LAT could have come from the axial and sagittal views alone, but inter-rater agreement was recorded best on coronal views of the LAT.

Agreement on the presence of atrophy did not translate to agreement or accuracy for classifying the 2 types of dementia. Our experienced clinicians could not differentiate controls from FTD at the 90% accuracy achieved by semi-automated volumetry [21]. This has been a challenge for several investigators [24–27]. This difficulty echoes the conclusion of our prior paper showing that many frontotemporal regions are mildly atrophied with normal aging and may overlap with FTD-related mild atrophy [14]. Others have identified the insula as a volumetric discriminator for FTD versus controls [16, 28], and the exclusion of insular VOIs from the SABRE protocol may have accounted for some of our inaccuracy in making the FTD distinction.

This study bears many other possible confounds: first, the control participants shown on figure 2 who had low LAC or LAT volumes would have looked like FTD. The likely larger confound was that 14 AD MRI scans were interspersed with the 42 FTD and controls. Though 3D MRI protocols can be done in a few minutes and have been recommended, for example, in the consensus guidelines for Vascular Cognitive Impairment [29], 76.8% of our errors were FTD and controls scans misattributed to AD. The boxplots in figure 2 show the overlap in LAC and LAT atrophy between AD and FTD, even if the mean AD volumes were still larger than for FTD.

Our methods did not entirely replicate a clinical situation: the slices prepared for review were in 16-bit Analyze format, whereas routine MRIs are presented to clinicians in 32-bit. Visual raters saw only the 5 selected MRI slices in sequence and never other brain images for a participant. Other investigators have commented on the importance of more naturalistic approaches [30]. Clinicians can also go back to see views as needed to update their interpretations. In this study, at least 1 rater changed criteria for scoring the atrophy once he had gone through the slides for the first 30 participants.

Bias in favor of volumetric analysis over visual rating may exist in our study due to combining the behavior and language presentation FTD subgroups, but participants for this study had shown great enough similarity between LAC and LAT critical cutoff values for accurate classification in the prior study [14] that we felt justified in taking this step. Kipps et al. [31] found that 75% of behavioral variant FTD patients had marked frontal and/or temporal atrophy which could be detected by raters on visual inspection, but 100% of their semantic dementia cases had focal LAT atrophy. Otherwise, 47% of patients with specifically behavioral variant FTD could not be discriminated from controls when raters examined frontal and temporal areas on MRI scans. Other limitations of the present study may include a relatively small sample size and the use of the regional, as opposed to lobar foci [32]. Studies using other brain regions for visual rating have been more successful [16, 33].

With the pertinent history usually available in a clinical setting, structural imaging may be more helpful for further characterizing the severity of the illness or anticipating further behavioral and cognitive changes for caregivers than it is for predicting the etiology of the dementia. In the absence of historical information, however, structural imaging may help to rule out other neurological disorders, such as neoplasm, stroke, infection, cerebral autosomal dominant arteriopathy with subcortical infarcts, and leukoencephalopathy.

That visual rating procedures are not reliable even among experts is a frequent concern [33–37]. Neuroradiologists primed to the diagnosis of behavioral variant FTD versus semantic dementia versus AD show higher accuracy of diagnostic categorization and inter-rater agreement than the neurologists in this and a previous study [27], although sensitivity remains low at approx. 60% [33]. Difficulties with regional neuroimaging assessments are not unique to the MRI modality. Comparisons of perfusion on SPECT imaging for differentiating FTD from AD yield similar results [38].

Conclusion

It is difficult to test whether volumetric analyses can be translated into routine clinical practice, particularly when the goal is to differentiate between FTD and early-onset AD. Structural neuroimaging, in conjunction with a complete history and neuropsychological testing, can help distinguish between FTD and normal aging, but routine clinical MRIs from participants with FTD, early-onset AD, and controls with age-related atrophy may appear very similar to clinicians.

Acknowledgments

NIA grant F32 AG022802 (T.W.C.); the Saul A. Silverman Family Foundation, as part of a Canada International Scientific Exchange Program project (M.F.); University of Toronto Dean's Fund for New Faculty (457494 to T.W.C.); an endowment to the Sam and Ida Ross Memory Clinic (T.W.C., M.F.); Canadian Institutes of Health Research MT – 12853, MRC-GR-14974, James S. McDonnell Foundation 21002032 (D.T.S.); Canadian Institutes of Health Research 13129 (S.E.B.).

D.T.S. is the Reva James Leeds Chair in Neuroscience and Research Leadership and S.E.B. holds the Brill Chair in Neurology, Department of Medicine, University of Toronto and Sunnybrook Health Sciences Centre.

References

- 1 Rosen HJ, Gorno-Tempini ML, Goldman WP, Perry RJ, Schuff N, Weiner M, Feiwell R, Kramer JH, Miller BL: Patterns of brain atrophy in frontotemporal dementia and semantic dementia. *Neurology* 2002;58:198–208.
- 2 Rosen HJ, Allison SC, Schauer GF, Gorno-Tempini ML, Weiner MW, Miller BL: Neuroanatomical correlates of behavioural disorders in dementia. *Brain* 2005;128(part 11):2612–2615.
- 3 Williams GB, Nestor PJ, Hodges JR: Neural correlates of semantic and behavioural deficits in frontotemporal dementia. *Neuroimage* 2005;24:1042–1051.
- 4 Liu W, Miller BL, Kramer JH, Rankin K, Wyss-Coray C, Gearhart R, Phengrasamy L, Weiner M, Rosen HJ: Behavioral disorders in the frontal and temporal variants of frontotemporal dementia. *Neurology* 2004;62:742–748.

- 5 Mummery CJ, Patterson K, Price CJ, Ashburner J, Frackowiak RSJ, Hodges JR: A voxel-based morphometry study of semantic dementia: relationship between temporal lobe atrophy and semantic memory. *Ann Neurol* 2000;47:36–45.
- 6 Rosen HJ, Kramer JH, Gorno-Tempini ML, Schuff N, Weiner M, Miller BL: Patterns of cerebral atrophy in primary progressive aphasia. *Am J Geriatr Psychiatry* 2002;10: 89–97.
- 7 Elfgrén CI, Ryding E, Passant U: Performance on neuropsychological tests related to single photon emission computerized tomography findings in frontotemporal dementia. *Br J Psychiatry* 1996;169:416–422.
- 8 Garrard P, Hodges JR: Semantic dementia: clinical, radiological and pathological perspectives. *J Neurol* 2000;247:409–422.
- 9 Heutink P, Stevens M, Rizzu P, Bakker E, Kros JM, Tibben A, Niermeijer MF, van Duijn CM, Oostra BA, van Swieten JC: Hereditary frontotemporal dementia is linked to chromosome 17q21–q22: a genetic and clinicopathological study of three Dutch families. *Ann Neurol* 1997;41:150–159.
- 10 Miller BL, Cummings JL, Villanueva-Meyer J, Boone K, Mehinger CM, Lesser IM, Mena I: Frontal lobe degeneration: clinical, neuropsychological and SPECT characteristics. *Neurology* 1991;41:1374–1382.
- 11 Rosso SM, Donker KL, Baks T, Joosse M, de Koning I, Pijnenburg Y, de Jong D, Dooijes D, Kamphorst W, Ravid R, Niermeijer MF, Verheij F, Kremer HP, Scheltens P, van Duijn CM, Heutink P, van Swieten JC: Frontotemporal dementia in The Netherlands: patient characteristics and prevalence estimates from a population-based study. *Brain* 2003; 126:2016–2022.
- 12 McKhann GM, Albert MS, Grossman M, Miller B, Dickson D, Trojanowski JQ: Clinical and pathological diagnosis of frontotemporal dementia: report of the Work Group on Frontotemporal Dementia and Pick's Disease. *Arch Neurol* 2001;58:1803–1809.
- 13 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: Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* 1998;51: 1546–1554.
- 14 Chow TW, Binns MA, Freedman M, Stuss D, Ramirez J, Scott CJM, Black SE: Overlap in frontotemporal atrophy between normal aging and patients with frontotemporal dementias. *Alzheimer Dis Assoc Dis* 2008;22: 327–335.
- 15 Rohrer JD, Omar R, Anderson VM, Price SL, Warren JD, Fox NC: Serial MRI measurement of temporal lobe and global atrophy progression in semantic dementia. *Alzheimer's Dement* 2006;2:S702.
- 16 Davies RR, Scallan VL, Graham A, Williams GB, Graham KS, Hodges JR: Development of an MRI rating scale for multiple brain regions: comparison with volumetrics and with voxel-based morphometry. *Neuroradiology* 2009;51:491–503.
- 17 Feldman H, Levy AR, Hsiung GY, Peters KR, Donald A, Black SE, Bouchard RW, Gauthier SG, Guzman DA, Hogan DB, Kertesz A, Rockwood K, ACCORD Study Group: A Canadian cohort study of cognitive impairment and related dementias (accord): study methods and baseline results. *Neuroepidemiology* 2003;22:265–274.
- 18 McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM: Mental and clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;34: 939–944.
- 19 Dade LA, Gao FQ, Kovacevic N, Roy P, Rockel C, O'Toole CM, Lobaugh NJ, Feinstein A, Levine B, Black SE: Semiautomatic brain region extraction: a method for parcellating brain regions from structural magnetic resonance images. *Neuroimage* 2004;22:1492–1502.
- 20 Kovacevic N, Lobaugh NJ, Bronskill MJ, Levine B, Feinstein A, Black SE: A robust method for extraction and automatic segmentation of brain images. *Neuroimage* 2002;17:1087–1100.
- 21 Chow TW, Binns MA, Freedman M, Stuss D, Ramirez J, Black SE: The diagnostic utility of regional atrophy in frontotemporal degeneration. *Alzheimer's Dement* 2006;2:P2–P287.
- 22 Cochran WG: The comparison of percentages in matched samples. *Biometrika* 1950; 37:256–266.
- 23 Omar R, Sampson EL, Loy CT, Mummery CJ, Fox NC, Rossor MN, Warren JD: Delusions in frontotemporal lobar degeneration. *J Neurol* 2009;256:600–607.
- 24 Likeman M, Anderson VM, Stevens JM, Waldman AD, Goldbolt AK, Frost C: Visual assessment of atrophy on magnetic resonance imaging in the diagnosis of pathologically confirmed young-onset dementias. *Arch Neurol* 2005;62:1410–1415.
- 25 Klöppel S, Stonnington CM, Barnes J, Chen F, Chu C, Good CD: Accuracy of dementia diagnosis – a direct comparison between radiologists and a computerized method. *Brain* 2008;131:2969–2974.
- 26 Whitwell JL, Josephs KA, Nossor MN, Stevens JM, Revez T, Holton JL: Magnetic resonance imaging signatures of tissue pathology in frontotemporal dementia. *Arch Neurol* 2005;62:1402–1408.
- 27 Victoroff J, Mack WJ, Grafton ST, Schreiber SS, Chui HC: A method to improve interrater reliability of visual inspection of brain MRI scans in dementia. *Neurology* 1994;44:2267–2276.
- 28 Zhou J, Greicius M, Gennatas E, Growdon ME, Jang JY, Rabinovici GD, Kramer JH, Weiner M, Miller BL, Seeley WW: Divergent network connectivity changes in behavioral variant frontotemporal dementia and Alzheimer's disease. *Brain* 2010;133:1352–1367.
- 29 Hachinski V, Iadecola C, Petersen RC, Breteler MM, Nyenhuis DL, Black SE, Powers WJ, DeCarli C, Merino JG, Kalaria RN, Vinters HV, Holtzman DM, Rosenberg GA, Wallin A, Dichgans M, Marler JR, Leblanc GG: National Institute of Neurological Disorders and Stroke-Canadian Stroke Network vascular cognitive impairment harmonization standards. *Stroke* 2007;38:1118.
- 30 Davies RR, Kipps CM, Mitchell J, Kril JJ, Halliday GM, Hodges JR: Progression in frontotemporal dementia: identifying a benign behavioral variant by magnetic resonance imaging. *Arch Neurol* 2006;63:1627–1631.
- 31 Kipps CM, Davies RR, Mitchell J, Kril JJ, Halliday GM, Hodges JR: Clinical significance of lobar atrophy in frontotemporal dementia: application of an MRI visual rating scale. *Dement Geriatr Cogn Disord* 2007;23: 334–342.
- 32 Bock C, Rockel C, Roy P, Gao F, Black SE: Topographical patterns of lobar atrophy in frontotemporal dementia and Alzheimer's disease. *Dement Geriatr Cogn Disord* 2006; 21:364–372.
- 33 Suárez J, Tartaglia MC, Vitali P, Erbetta A, Neuhaus J, Laluz V, Miller BL: Characterizing radiology reports in patients with frontotemporal dementia. *Neurology* 2009;73: 1073–1074.
- 34 Scheltens P, Launer LJ, Barkhof F, Weinstein HC, van Gool WA: Visual assessment of medical temporal lobe atrophy on magnetic resonance imaging: interobserver reliability. *J Neurol* 1995;242:557–560.
- 35 Scheltens P, Fox N, Barkhof F, De Carli C: Structural magnetic resonance imaging in the practical assessment of dementia: beyond exclusion. *Lancet Neurol* 2002;1:13–21.
- 36 Erkinjuntti T, Lee DH, Gao F, Steenhuis R, Eliasziw M, Fry R, Merskey H, Hachinski V: Temporal lobe atrophy on magnetic resonance imaging in the diagnosis of early Alzheimer's disease. *Arch Neurol* 1993;50: 305–310.
- 37 Hanyu H, Tetsuichi A, Sakurai H, Takasaki M, Shindo H, Abe K: Magnetization transfer measurements of the hippocampus in the early diagnosis of Alzheimer's disease. *J Neurol Sci* 2001;188:79–84.
- 38 Horn JF, Habert MO, Kas A, Malek Z, Maksud P, Lacomblez L, Giron A, Fertel B: Differential automatic diagnosis between Alzheimer's disease and frontotemporal dementia based on perfusion SPECT images. *Artif Intell Med* 2009;47:147–158.