Can MRI Visual Assessment Differentiate the Variants of Primary-Progressive Aphasia?


ABSTRACT

BACKGROUND AND PURPOSE: Primary-progressive aphasia is a clinically and pathologically heterogeneous condition. Nonfluent, semantic, and logopenic are the currently recognized clinical variants. The recommendations for the classification of primary-progressive aphasia have advocated variant-specific patterns of atrophy. The aims of the present study were to evaluate the sensitivity and specificity of the proposed imaging criteria and to assess the intra- and interrater reporting agreements.

MATERIALS AND METHODS: The cohort comprised 51 patients with a root diagnosis of primary-progressive aphasia, 25 patients with typical Alzheimer disease, and 26 matched control participants. Group-level analysis (voxel-based morphometry) confirmed the proposed atrophy patterns for the 3 syndromes. The individual T1-weighted anatomic images were reported by 3 senior neuroradiologists.

RESULTS: We observed a dichotomized pattern of high sensitivity (92%) and specificity (93%) for the proposed atrophy pattern of semantic-variant primary-progressive aphasia and low sensitivity (21% for nonfluent-variant primary-progressive aphasia and 43% for logopenic-variant primary-progressive aphasia) but high specificity (91% for nonfluent-variant primary-progressive aphasia and 95% for logopenic-variant primary-progressive aphasia) in other primary-progressive aphasia variants and Alzheimer disease (specificity 43%, specificity 92%). MR imaging was least sensitive for the diagnosis of nonfluent-variant primary-progressive aphasia. Intrarater agreement analysis showed mean κ values above the widely accepted threshold of 0.6 (mean, 0.63 ± 0.16). Pair-wise interobserver agreement outcomes, however, were well below this threshold in 5 of the 6 possible interrater contrasts (mean, 0.41 ± 0.09).

CONCLUSIONS: While the group-level results were in precise agreement with the recommendations, semantic-variant primary-progressive aphasia was the only subtype for which the proposed recommendations were both sensitive and specific at an individual level.

ABBREVIATIONS: AD = Alzheimer disease; lvPPA = logopenic-variant PPA; nfvPPA = nonfluent-variant PPA; PPA = primary-progressive aphasia; svPPA = semantic-variant PPA

Primary-progressive aphasia (PPA) is a clinically and pathologically heterogeneous condition characterized by insidious onset and gradual worsening of language due to degeneration of brain language areas. Clinical heterogeneity, compounded by the evolution of signs and symptoms, makes accurate classification of patients a challenging task. Making a reliable clinical diagnosis, on the other hand, is important. Despite lack of a one-to-one relationship between the clinical diagnosis and the underlying pathology, previous clinicopathologic series have identified probabilistic associations among the 3 recognized clinical presentations of PPA and certain pathologies. There are established associations between semantic-variant PPA (svPPA) and frontotemporal lobar degeneration–TAR DNA binding protein 43 (TDP-43); nonfluent-variant PPA (nfvPPA) and frontotemporal lobar degeneration–tau; and logopenic-variant PPA (lvPPA) and Alzheimer pathology.1 5 The recommendations on clinical subtyping of PPA have proposed that clinical classification can be supported by imaging according to the pattern of regional atrophy or metabolic impairment.6 Left posterior frontoinsular atrophy in nfvPPA, anterior temporal atrophy in svPPA, and left posterior peri-Sylvian or parietal atrophy in lvPPA are the recommended atrophy patterns.
Remarkably, these imaging recommendations are derived from studies that either used group-averaged data—which though highly replicated,1-10 are not necessarily valid for single-patient diagnosis—or were based on observed atrophy in convenience samples11-14 without qualification of sensitivity, specificity, or reliability. Little is known about whether individual patients, as opposed to groups, fulfilling the clinical criteria for these variants reliably present with the prescribed patterns of atrophy and whether these patterns have sufficient reliability to be exploited to arrive at an accurate syndromic diagnosis.

The aims of the present study were the following: 1) to evaluate the utility of the proposed imaging criteria for the diagnosis of PPA variants by contrasting the patterns of atrophy in individual patients with PPA, as reported by senior neuroradiologists, with the recommendations from the criteria; and 2) to assess the neuroradiologists’ intra- and interrater agreement, which, in turn, would be an indication of the robustness of the observed abnormalities.

MATERIALS AND METHODS

Participants

The cohort comprised 51 patients with a root diagnosis of PPA, 25 patients with mild typical Alzheimer disease (AD) as a neurodegenerative control group, and 26 healthy age- and education-matched control participants. The breakdown of the subjects with PPA based on clinical variants was 21 with svPPA, 14 with nfvPPA, 14 with mixed PPA, and 2 with lvPPA. Clinical diagnoses were made in accordance with the published criteria for the diagnosis of PPA15 and probable Alzheimer disease.16 The diagnosis of PPA variants was based on a quantitative application of the consensus recommendations6 as detailed elsewhere.17 No patients had pedigrees to suggest an autosomal dominant genetic cause. The mixed-PPA group, however, was designated “mixed” on the basis of strict application of the proposed clinical criteria. The patients almost certainly corresponded, however, to what others have designated lvPPA in that they had neither svPPA nor nfvPPA, and they had the same group-level atrophy pattern as in previous lvPPA cohorts.18 Furthermore, some researchers have proposed to diagnose lvPPA through a hierarchic decision tree in which the key feature of this group is that they are neither svPPA nor nfvPPA.19 Applying such an algorithm to the present mixed cases would also have them classified as lvPPA. The subjects with mixed PPA in this study should, therefore, be considered analogous to those with lvPPA and are referred to henceforth as such.

The study was approved by the institutional review board of Cambridge University hospitals, UK.

Neuropsychological Battery and Connected Speech Analysis

All patients underwent comprehensive neuropsychological and connected speech assessment before imaging, full details of which have been published previously.20 The On-line Table provides a summary of some of these data.

Imaging

Image Acquisition. Study participants were scanned within an average of 1.6 ± 0.8 months from cognitive assessment. All MR imaging was performed on the same Magnetum Trio 3T system (Siemens, Erlangen, Germany). T1-weighted anatomic images were acquired by using 3D MPRAGE with the following imaging parameters: TR/TE/T1/flip angle = 2300/2.86/900 ms/90°, 144 sections, 192 × 192 matrix dimensions, and 1.25 × 1.25 × 1.25 mm3 voxel size. Receiver bandwidth and echo spacing were 240 Hz/pixel and 6.7 ms, respectively.

Data Processing and Group-Level Data Analysis

All obtained T1 volumes were preprocessed as reported previously.21 Preprocessing and warping procedures need reasonable initial estimates; hence, the origin of each structural volume was set manually to the anterior commissure before preprocessing. All volumes were then spatially normalized and segmented by using the unified segmentation model in statistical parametric mapping 5 (SPM5) (http://www.fil.ion.ucl.ac.uk/spm/software/spm5/).22 The segments were also modulated to compensate for volumetric differences introduced into the warped images. Finally, gray matter segments were smoothed by using an 8-mm full width at half maximum isotropic Gaussian kernel. Total intracranial volumes were calculated by using the automated SPM technique as described elsewhere,23 and the obtained values, along with age, were fed into the statistical models as nuisance covariates. Following these steps, a 2-sample t test implemented in SPM522 contrasted the gray matter volumes of the patient groups against those in controls. The statistical maps were thresholded at P < .01, corrected for multiple comparisons (false discovery rate = .01).

Visual Reporting of Individual Scans

Three senior neuroradiologists who were blinded to the clinical diagnoses of the study participants separately reported all unprocessed T1 sequences displayed by using the FMRIB Software Library (FSL, Version 4.1.2; http://www.fmrib.ox.ac.uk/fsl).24 Thirty-five scans (n = 7 for each diagnostic group including controls) were duplicated, bringing the total number of scans to 137 to assess intrarater agreement.

The neuroradiologists were asked to report the scans for the presence and patterns of disproportionate regional or global atrophy in 2 stages. In the first stage, the outcome of which was used for calculations of intra- and interrater agreement and sensitivity, the radiologists were asked to report the scans in their own preferred styles. Subsequent calculations for this stage were based on the reported lobar distribution of the abnormalities. “Global atrophy” and “no atrophy” were also accepted as valid entries. In agreement with the published recommendations for AD25 and PPA,6 the following lobar distributions were deemed consistent with the syndromic PPA variants and Alzheimer disease: temporal lobe atrophy for svPPA; left frontal or left frontotemporal atrophy for nfvPPA; left temporal, left parietal, or left temporoparietal atrophy for lvPPA; and temporal, parietal, or temporoparietal atrophy for typical AD. Rating a scan as showing “global atrophy” was not deemed acceptable for any of the syndromic variants of PPA or for AD because the reporting radiologists had not observed “disproportionate” atrophy of a target region. In the second stage, the outcome of which was the basis for specificity and separate sensitivity and agreement calculations, however, the raters were specifically asked to comment on whether there was “disproportionate” left posterior frontoinsular atrophy (indicative of
nfvPPA), anterior temporal lobe atrophy (indicative of svPPA), left posterior peri-Sylvian or parietal lobe atrophy (indicative of lvPPA), and medial temporal or parietal atrophy (indicative of typical AD). Instructions for this second stage were only given after stage 1 was completed to ensure that the initial ratings were not biased by expected atrophy patterns.

**Statistical Considerations**

Predictive Analytics Software (PASW, Version 18; IBM, Armonk, New York) and SPM5 were used for statistical analysis of the data. One-way ANOVA with a 2-tailed significance level of .05 was used to compare the demographic and neuropsychological measures. A pair-wise κ was used to assess the intra- and interobserver agreement in the reports of atrophy. Group-level comparisons of the imaging data were made with a 2-sample t test implemented in SPM5, with age and total intracranial volume included as nuisance covariates. Group-level results are reported at a false discovery rate–corrected $P < .01$.

**RESULTS**

Demographic data for all participant groups are summarized in Table 1. Neuropsychological and language assessment results are presented in the On-line Table.

![FIG 1. Group-level patterns of atrophy in identical axial, sagittal, and coronal sections of the brain. Images are displayed in neurologic orientation. Asterisks demonstrate the section most representative for the particular groups. All comparisons were made at false discovery rate–corrected $P < .01$.](image-url)

<table>
<thead>
<tr>
<th>Demographics</th>
<th>svPPA Mean (Range)</th>
<th>nfvPPA Mean (Range)</th>
<th>lvPPA Mean (Range)</th>
<th>AD Mean (Range)</th>
<th>Control Mean (Range)</th>
<th>Omnibus Sig</th>
<th>Omnibus Sig P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at test (yr)</td>
<td>67 (60–79)</td>
<td>68.9 (53–79)</td>
<td>70.8 (60–83)</td>
<td>68 (60–79)</td>
<td>67.5 (53–80)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Disease duration (mo)*</td>
<td>56 (24–108)</td>
<td>38.57 (18–60)</td>
<td>48.7 (24–108)</td>
<td>58 (24–96)</td>
<td>NA</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Education (yr)*</td>
<td>13.6 (10–18)</td>
<td>12.6 (10–20)</td>
<td>11.6 (9–16)</td>
<td>12.5 (10–19)</td>
<td>12.8 (10–20)</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>11 M, 10 F</td>
<td>5 M, 9 F</td>
<td>6 M, 10 F</td>
<td>12 M, 13 F</td>
<td>11 M, 15 F</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>ADL-Q</td>
<td>1 (0–5)</td>
<td>0.53 (0–4)</td>
<td>0.68 (0–4)</td>
<td>2.3 (0–6)</td>
<td>NA</td>
<td>NS</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** ADL-Q indicates Activities of Daily Living Questionnaire; Sig, significant; NA, not applicable; NS, not significant.

*Nonparametric test.
lvPPA, by left posterior temporoparietal atrophy; and typical AD showed bilateral hippocampal and patchy temporoparietal atrophy.

Concerning the single-subject visual reporting outcomes, sensitivity calculations based on the lobar distribution of the abnormalities revealed almost perfect results in the svPPA group (mean sensitivity, 98% ± 2.9%) but low sensitivity of the imaging markers in the other PPA variants and the typical AD group (Table 2). The proposed imaging markers were least sensitive for the diagnosis of nfvPPA (mean, 29% ± 21%). Sensitivity figures based on the prescribed patterns of atrophy revealed slightly lower values compared with the above figures but a similar pattern overall: high sensitivity for svPPA and low values for the other study groups (Table 2). Specificity figures were, however, consistently high for all diagnostic groups with no discernible difference (Table 3).

Tables 4 and 5 provide values for intra- and interobserver agreement for all diagnostic groups. Intraobserver agreement values for the reported lobar distribution of atrophy were consistently above the widely accepted threshold of 0.6 (mean, 0.75 ± 0.18), indicating substantial agreement. The recommendation-based values, however, fell below the 0.6 threshold for 2 of the 3 reporting radiologists (mean, 0.63 ± 0.16). In the pair-wise interobserver agreement values, while $\kappa$ was just below the 0.6 threshold (mean, 0.56 ± 0.08) for the reported lobar distribution of atrophy (first round), it dropped considerably to 0.41 ± 0.09 for the recommendation-based outcomes; 0.4 is generally considered the minimum threshold for moderate agreement.

**DISCUSSION**

This study provides an objective assessment of the utility of the proposed MR imaging markers in supporting the diagnosis of various PPA variants. The group-level atrophy patterns were in precise agreement with the proposed imaging criteria for different PPA subtypes. Assessing the reliability of these measures at a single-subject level is, however, much more relevant—indeed mandatory—for determining the diagnostic utility of the proposed markers.

**Tables 2-5**

| Table 2: Sensitivity of the proposed imaging markers for the diagnosis of PPA variants and typical AD based on the lobar distribution of the atrophy and specific consensus recommendations |
|----------------------|------------------|------------------|------------------|------------------|
| Sensitivity based on lobar distribution |
| svPPA | nfvPPA | lvPPA | AD |
| Rater 1 | 100% | 50% | 50% | 36% |
| Rater 2 | 95% | 8% | 57% | 64% |
| Rater 3 | 100% | 29% | 64% | 60% |
| Mean (SD) | 98% (2.9%) | 29% (21%) | 57% (7%) | 53% (15%) |

| Table 3: Specificity of the proposed imaging markers for the diagnosis of PPA variants and typical AD based on the specific consensus recommendations |
|----------------------|------------------|------------------|------------------|------------------|
| Specificity Based on Recommendations |
| svPPA | nfvPPA | lvPPA | AD |
| Rater 1 | 93% | 92% | 95% | 93% |
| Rater 2 | 95% | 89% | 93% | 92% |
| Rater 3 | 93% | 92% | 97% | 91% |
| Mean (SD) | 93% (0.01) | 91% (0.02) | 95% (0.02) | 92% (0.01) |

| Table 4: Intraobserver agreement for the reported lobar distribution of abnormalities and recommendations |
|----------------------|------------------|------------------|------------------|------------------|
| Intraobserver agreement for the lobar distribution of abnormalities and recommendations |
| svPPA | nfvPPA | lvPPA | AD |
| Rater 1 and 2 (κ [SE]) (SE) | 0.61 (0.1) | 0.95 (0.04) | 0.68 (0.1) | 0.75 (0.18) |
| Rater 2 and 3 (κ [SE]) (SE) | 0.5 (0.09) | 0.81 (0.07) | 0.57 (0.1) | 0.63 (0.16) |

| Table 5: Interobserver agreement for the reported lobar distribution of abnormalities and recommendations |
|----------------------|------------------|------------------|------------------|------------------|
| Interobserver agreement for the lobar distribution of abnormalities and recommendations |
| svPPA | nfvPPA | lvPPA | AD |
| Raters 1 and 2 (κ [SE]) (SE) | 0.56 (0.06) | 0.48 (0.06) | 0.64 (0.06) | 0.56 (0.08) |
| Raters 1 and 3 (κ [SE]) (SE) | 0.31 (0.07) | 0.47 (0.07) | 0.44 (0.06) | 0.41 (0.09) |
imaging criteria in further classification of individual patients with PPA. In the absence of reliable automated single-subject statistical measures capable of detecting the abnormalities at an individual level, assessment of the consistency and reliability of neuroradiologists’ reports along with the level of agreement constitutes a suitable substitute. Moreover, it mirrors real-life clinical practice in which visual rating of scans remains the standard reporting method.

The group-level voxel-based morphometry–based gray matter atrophy patterns for each of the PPA variants (Fig 1) were consistent with those in past studies.7–9 This finding was important to confirm because it was precisely this group-level atrophy pattern that led to the recommendations for imaging-supported diagnoses. The results of the visual rating suggested a dichotomized pattern of high sensitivity and specificity of the proposed imaging markers for svPPA, but less reliable outcomes for the other 2 PPA subtypes, with a low sensitivity but rather high specificity. Agreement analyses for the whole group revealed substantial intrarater but only moderate interrater agreement values (mean $\kappa$, 0.63 and 0.41, respectively) for the recommendation-based atrophy patterns. This finding is now discussed in more detail for each variant.

**svPPA**

svPPA is characterized by an amodal loss of knowledge that consistently presents as a reduction of expressive vocabulary and word comprehension.27 Various studies have emphasized the importance of the anterior temporal lobes as hubs of semantic knowledge.28,29 In agreement with the proposed recommendations, group-average voxel-based morphometry analysis of the svPPA participants revealed predominant bilateral anterior temporal lobe atrophy. The high sensitivity and specificity of the proposed imaging markers (means, 0.98% and 93%, respectively) demonstrated that the presence of temporal lobe atrophy offered robust support for the diagnosis of svPPA. This finding is not unexpected because, though not always assessed systematically, atrophy of rostral-inferior temporal structures has been consistently reported in svPPA (also known as semantic dementia) both at a group level and individually.7,10,30–35 Given the uniform pattern of atrophy seen in this consecutively recruited cohort of 21 patients with svPPA, it can be argued that the diagnosis of svPPA should be seriously questioned in the absence of this atrophy pattern. Gil-Navarro et al36 found the same consistent presence of anterior temporal lobe atrophy in a study of 29 patients with PPA that included 5 with svPPA. As already mentioned, most patients with svPPA have frontotemporal lobar degeneration–TDP-43 pathology, but frontotemporal lobar degeneration-tau pathology is found occasionally. Previous work has indicated that the atrophy pattern does not discriminate between these 2 pathologic substrates.31

**nfvPPA**

Clinical features of nfvPPA include effortful, halting speech with sound distortions and/or grammatic errors in language production.6 Degeneration of the left frontal operculum and rostral insula is the culprit lesion in nfvPPA. More recent studies have also highlighted involvement of premotor32 and basal ganglia37 regions. Like svPPA, the group-average voxel-based morphometry findings in our nfvPPA cohort were largely compatible with the proposed diagnostic recommendations. As demonstrated in Fig 1, voxel-based morphometry analysis clearly identified disproportionate left-sided atrophy in the frontal operculum and insula. There was, in addition, evidence of further atrophy in the left basal ganglia region, but no atrophy was visible in the premotor area.

Concerning the single-subject outcomes, none of the patients with nfvPPA had unanimous reports of the prescribed atrophy pattern by all 3 neuroradiologists. In fact, “left posterior frontal and insular atrophy” was only the third most commonly observed report in this group with “no atrophy” and “left posterior perisylvian atrophy” being the first and second, respectively (data not shown). “No atrophy” was reported by at least one of the neuroradiologists in 10 of 14 (71%) individuals with nfvPPA and in 22 of the total 42 (14 × 3) reports (52%). Low sensitivity values (mean, 21% ± 7%) further corroborated the above findings. Given the abundance of no-atrophy reports, the most plausible explanation for this result seems to be that atrophy in patients with nfvPPA is often very subtle. Also, clinical heterogeneity inherent in the recommended features of nfvPPA (ie, requiring the presence of either abnormal speech or agrammatism) and more white than gray matter burden are other potential explanations for the observed discrepancies. High specificity values, however, indicated the potential utility of the prescribed pattern of atrophy in the diagnosis of nfvPPA if present.

Further evidence for the inconsistency of imaging findings in nfvPPA comes from previous single case studies reporting widely discrepant findings, ranging from no atrophy40–43 to left hemispheric atrophy43 to left frontotemporal atrophy,14 bifrontal atrophy,40 and generalized atrophy.41 Even group-level findings, using parametric analysis techniques such as SPM, have been inconsistent, with different studies showing evidence of: left-sided inferior frontal and insular atrophy42 and hypometabolism43; atrophy in a wide distribution comprising the left inferior frontal, superior temporal, and inferior parietal areas, with7,9 and without additional atrophy in the premotor area; and finally abnormalities in the premotor cortex and left basal ganglia.8,32,44 One previous study reported a considerably higher sensitivity for the MR imaging–defined atrophy pattern of nfvPPA (76%).36 The discrepancy, however, likely relates to the study design in that the raters had to expressly classify scans for the 3 proposed atrophy patterns and in a group comprising only patients with PPA (there were neither healthy controls nor controls with dementia). This difference is important, given the high prevalence of no-atrophy reports in our nfvPPA group (see above). Considering that the main challenge in the diagnosis of degenerative aphasia is at the mildest stages when it is difficult to distinguish degenerative aphasia from a normal variation, inclusion of scans with normal findings makes our study a closer reflection of real-life situations.

**LvPPA**

LvPPA is the PPA variant that is highly associated with Alzheimer pathology.7 Impaired single-word retrieval in spontaneous speech and impaired sentence repetition are the recommended features. While a number of studies have failed to demonstrate the utility of these features in the diagnosis of LvPPA,17,45 atrophy of the left
temporoparietal lobe has been consistently emphasized in Alzheimer disease—related aphasia. In the “Materials and Methods” section, we used the lvPPA label to the group of patients with PPA whom we had previously reported as having mixed PPA. Turning to the individual visual reporting outcomes, we found low-to-moderate sensitivity (mean, 49% ± 7%) for the prescribed atrophy patterns. This was consistent with the results of the only previous study looking at the same metrics that found a sensitivity of 57% for MR imaging atrophy. Given the high specificity value (mean, 95% ± 2%), it can be inferred that while the presence of the prescribed pattern of “left posterior peri-Sylvian or parietal” atrophy is highly suggestive of lvPPA, its absence does not exclude the diagnosis of lvPPA. In addition, assessment of individual reports revealed that contrary to what might be expected, a typical AD atrophy pattern (ie, medial temporal or parietal atrophy) was reported in only 8% of the ratings of patients with lvPPA; an lvPPA atrophy pattern was the most common (38%), while an nfvPPA atrophy pattern was the second most frequently reported outcome (20%). This radiologic finding resonates with the previously reported difficulties in distinguishing nfv- and lvPPA variants on clinical and neuropsychological grounds. None of the radiologists had reported global atrophy for the n = 16 lvPPA cohort. This is an important negative, given possible concerns about the severity of dementia in this group.

CONCLUSIONS
This study provides an objective assessment of the utility of the proposed MRI recommendations for supporting the diagnoses of 3 PPA variants. Our findings are largely compatible with the only previous study on the subject. Moreover, to our knowledge, this article is the first to report the intra- and interrater agreement of the reporting radiologists and the specificity of the MR imaging markers for the diagnosis of PPA variants. Our study provides compelling evidence for the utility of the proposed imaging recommendations for the diagnosis of svPPA. On the basis of the findings of the current and previous studies, it could even be argued that lack of anterior temporal lobe atrophy should exclude the diagnosis of svPPA. The results were less consistent in the other groups. While high specificity values observed in all groups indicate the potential utility of the recommendations for patients in whom the atrophy patterns can be identified, low sensitivity and modest agreement values suggest that absence of the proposed atrophy patterns is common in the nonsemantic PPA subtypes.


REFERENCES
and structural MR image analysis and implementation as FSL. *Neurolmage* 2004;23(suppl 1):S208–19 CrossRef Medline
29. Mion M, Patterson K, Acosta-Cabronero J, et al. What the left and right anterior fusiform gyri tell us about semantic memory. *Brain* 2010;133:3256–68 CrossRef Medline