

## Research Report

---

# Detecting Primary Progressive Aphasia Atrophy Patterns: A Comparison of Visual Assessment and Quantitative Neuroimaging Techniques

Stephanie Franczak<sup>a</sup>, Jessica Pommy<sup>a</sup>, Greta Minor<sup>a</sup>, Chandler Zollicoffer<sup>a</sup>, Manav Bhalla<sup>b</sup>, Mohit Agarwal<sup>b</sup>, Andrew Nencka<sup>b</sup>, Yang Wang<sup>b</sup>, Andrew Klein<sup>b</sup>, Darren O'Neill<sup>b</sup>, Jude Henry<sup>a</sup> and Laura Glass Umfleet<sup>a,\*</sup>

<sup>a</sup>*Department of Neurology, Medical College of Wisconsin, Milwaukee, WI, USA*

<sup>b</sup>*Department of Radiology, Medical College of Wisconsin, Milwaukee, WI, USA*

Received 25 May 2022

Accepted 13 July 2022

Pre-press 29 July 2022

Published 5 August 2022

### Abstract.

**Background:** There are now clinically available automated MRI analysis software programs that compare brain volumes of patients to a normative sample and provide *z*-score data for various brain regions. These programs have yet to be validated in primary progressive aphasia (PPA).

**Objective:** To address this gap in the literature, we examined Neuroreader™ *z*-scores in PPA, relative to visual MRI assessment. We predicted that Neuroreader™ 1) would be more sensitive for detecting left > right atrophy in the cortical lobar regions in logopenic variant PPA clinical phenotype (lvPPA), and 2) would distinguish lvPPA ( $n = 11$ ) from amnesic mild cognitive impairment (aMCI;  $n = 12$ ).

**Methods:** lvPPA or aMCI patients who underwent MRI with Neuroreader™ were included in this study. Two neuroradiologists rated 10 regions. Neuroreader™ lobar *z*-scores for those 10 regions, as well as a hippocampal asymmetry metric, were included in analyses.

**Results:** Cohen's Kappa coefficients were significant in 10 of the 28 computations ( $k = 0.351$  to  $0.593$ ,  $p \leq 0.029$ ). Neuroradiologists agreed 0% of the time that left asymmetry was present across regions. No significant differences emerged between aMCI and lvPPA in Neuroreader™ *z*-scores across left or right frontal, temporal, or parietal regions ( $ps > 0.10$ ). There were significantly lower *z*-scores in the left compared to right for the hippocampus, as well as parietal, occipital, and temporal cortices in lvPPA.

**Conclusion:** Overall, our results indicated moderate to low interrater reliability, and raters never agreed that left asymmetry was present. While lower *z*-scores in the left hemisphere regions emerged in lvPPA, Neuroreader™ failed to differentiate lvPPA from aMCI.

Keywords: Logopenic variant PPA, magnetic resonance imaging, mild cognitive impairment, neuroreader

---

\*Correspondence to: Laura Umfleet, PsyD, ABPP-CN, Associate Professor of Neurology, Division of Neuropsychology, Medical College of Wisconsin, 8701 W Watertown Plank Rd, Wauwatosa, WI 53226, USA. E-mail: lumfleet@mcw.edu.

## INTRODUCTION

Primary progressive aphasia (PPA) refers to a group of neurodegenerative disorders characterized by language dysfunction with relative sparing of other cognitive domains in early stages (e.g., amnesic profile) [1, 2]. A clinical diagnosis of PPA involves first obtaining a detailed history of symptoms (e.g., initial primary language concern) and assessment of cognitive functions to identify diagnostic features that may or may not fit with a specific variant of PPA (i.e., semantic, non-fluent/agrammatic, and logopenic). In 2011, an international consensus group of PPA investigators published criteria for the diagnosis and classification of PPA based on both published research and group consensus [3]. The specific pattern of language weaknesses for each of the three main PPA subtypes was outlined and two additional specifiers were included that clinicians could use when appropriate. Namely, when neuroimaging or pathological data was available and consistent with clinical presentation, clinicians could further specify PPA subtypes as: 1) “imaging supported” (based on the pattern of brain atrophy and/or hypometabolism/hypoperfusion detected on neuroimaging, or, 2) “with definite pathology” based on presence of pathology or genetic testing. While pathology and genetic data are likely not readily available in the context of standard clinical care, brain MRI scans are more frequently available. Understanding the utility of MRI findings in the context of clinical care for PPA is of interest. While the pattern of atrophy may help to differentiate PPA variants, the diagnostic potential of MRI assessment within a PPA population has not yet been fully realized.

The logopenic variant of PPA (lvPPA) is typically associated with the greatest atrophy in the parietal and left posterior temporal regions [3–8]. Amyloid deposition is most strongly related to this PPA variant [4, 7] (though there is neuroimaging and clinical heterogeneity in lvPPA) [4–6, 9, 10]. Clinical characteristics of lvPPA include impairments in word retrieval (with intact word knowledge), sentence/phrase repetition, and phonological paraphasias (with initial, relatively preserved grammar and comprehension) [3]. Clinical practice for diagnosing lvPPA includes a cognitive evaluation and neuroimaging to rule out structural lesions to eloquent cortex. Clinical neuroimaging analyses typically rely on visual interpretation by a neuroradiologist. Sajjadi et al. [11] showed the high specificity of imaging patterns for PPA variants on MRI in terms of visual assessment, with the logopenic

group (4 patients originally diagnosed with lvPPA and 12 patients originally diagnosed with mixed PPA, but later reclassified as lvPPA) found to have the highest specificity (95%) though sensitivity was lower (43%).

Relatively recent development of clinically available quantitative neuroimaging programs provides qualified practitioners additional clinical data to integrate into their clinical conceptualizations. NeuroQuant™ [12] and Neuroreader™ [13] are FDA-cleared automated segmentation programs that are available for clinical use, whereas programs such as FreeSurfer are used in research [14, 15]. Comparisons between visual ratings and clinically-available automated MRI analysis software in patient populations such as Alzheimer’s disease, traumatic brain injury, and epilepsy have produced mixed findings [16–20]. If automated software programs are implemented into clinical practice, it is reasoned that these programs should be validated in a variety of neurological conditions. To date, there has been no study that has examined the clinical utility of Neuroreader™ to aid in diagnosis of PPA over visual assessment of brain atrophy by neuroradiologists in a clinical setting. Therefore, the present study aimed to 1) establish interrater reliability of blinded neuroradiologist MRI visual assessment, 2) compare automated quantitative volumetric data using Neuroreader™ to neuroradiologist visual MRI assessment, and 3) determine the accuracy of Neuroreader™ data in distinguishing lvPPA from an amnesic mild cognitive impairment (aMCI) group in a clinical sample. We predicted that Neuroreader™ would be more sensitive for detecting left > right atrophy in the cortical lobar regions in logopenic variant PPA and that the automated analyses would distinguish lvPPA from aMCI.

## MATERIALS AND METHODS

### *Participants*

The current retrospective study analysed data, collected as part of a larger clinical database, from a convenience sample comprised of 23 patients (14 men and 9 women) that underwent a comprehensive clinical neuropsychological evaluation and received a neurobehavioral diagnosis of either: MCI most consistent with the lvPPA phenotype (“lvPPA”;  $N=11$ ) or aMCI most consistent with the classic mesial-temporal presentation of AD ( $N=12$ ). Patients groups were of similar age and gender. All patients presented to a clinical visit at a Midwestern

US medical center. At this medical center, dementia specialists within the department of neurology see well over 1,500 cases annually that are diagnosed with neurodegenerative conditions. Clinical diagnoses of lvPPA and aMCI due to possible AD were made by a clinical neuropsychologist as part of clinical care, and were documented in the patients' neuropsychological report in their medical record. Patients were seen in one of two neuropsychol-

ogy clinics: a primary neuropsychology clinic which allows for slightly longer assessments or within a multidisciplinary clinic that included a neuropsychologist, behavioral neurologist, and a clinical social worker. All patients underwent MR Neuroreader™ as part of routine clinical care. Demographic data for lvPPA and aMCI groups are displayed in Table 1. The current study focused solely on patients diagnosed with lvPPA due to limited sample sizes (3 or less) for other phenotypes of PPA (semantic and nonfluent/agrammatic variants). The aMCI group was inspected to ensure that a PPA diagnosis had not been proposed as clinical differential diagnosis in the neuropsychological report (as overarching criteria for PPA diagnosis excludes initial episodic memory impairment, a hallmark of aMCI due to possible/probable AD). Notably, as this study utilizes a clinical sample, there was expected variability in the *specific* neuropsychological measures used across patients; however, standardized scores for neuropsychological measures that were available for more than half the sample were included in Table 2. Measures in Table 2 by no means reflect the entirety of the comprehensive assessment, especially all of the language measures administered. A clinical neuropsychological evaluation includes a detailed clinical interview

Table 1  
Demographics for lvPPA and aMCI Groups

Variable	lvPPA Group N = 11			aMCI Group N = 12		
	Freq.			Freq.		
Gender						
Women	4			5		
Men	7			7		
Race						
White	10			10		
Black	1			2		
Handedness						
RH	9			10		
LH	2			2		
	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
Age	69.64	5.52	60.00–77.00	68.42	7.40	56.00–79.00
Education	15.63	1.75	12.00–18.00	14.17	2.48	10.00–18.00

Table 2  
Means and standard deviations for the lvPPA and aMCI groups on select neuropsychological test variables

Variable	lvPPA Group N = 11			aMCI Group N = 12			All Subjects N = 23		
	N	<i>M</i>	<i>SD</i>	N	<i>M</i>	<i>SD</i>	N	<i>M</i>	<i>SD</i>
WMS-IV Visual Reproduction I standard score	8	8.38	3.25	8	7.63	2.62	16	8.00	2.89
WMS-IV Visual Reproduction II standard score	8	10.13	3.44	8	5.13	2.10	16	7.63	3.77
WRAT-4 Word Reading standard score	10	90.30	13.70	11	97.55	9.60	21	94.10	12.01
Bosting Naming standard score	10	58.50	24.01	12	92.42	17.83	22	77.00	26.69
Category Fluency scaled score	10	4.70	4.69	12	8.92	2.31	22	7.00	4.11
Letter Fluency scaled score	10	6.30	3.71	12	9.58	2.31	22	8.09	3.39
Trails A standard score	11	73.45	25.26	12	103.17	19.46	23	88.96	26.64
Trails B standard score	11	68.82	24.98	12	98.42	19.75	23	84.26	26.59
WAIS-IV Similarities scaled score	8	7.25	3.58	8	9.75	2.43	16	8.50	3.22
WAIS-IV Block Design scaled score	8	10.00	3.30	8	10.00	2.39	16	10.00	2.78
WAIS-IV Digit Span scaled score	11	6.09	3.18	12	9.25	2.53	23	7.74	3.22
WAIS-IV Coding scaled score	11	8.55	3.11	12	10.25	1.71	23	9.43	2.57
HVLT delayed recall standard score	10	77.10	17.23	12	56.25	5.59	22	65.73	16.02
HVLT percent retention standard score	10	98.00	19.52	12	57.58	9.30	22	75.95	25.16
HVLT discrimination index standard score	10	91.60	12.89	12	63.50	11.43	22	76.27	18.57
HVLT total recall standard score	10	69.60	21.06	12	73.58	14.59	22	71.77	17.48

WMS-IV, Wechsler Memory Scale-4<sup>th</sup> Edition Logical Memory subtest assesses immediate and delayed recall of stories; Visual Reproduction is a measure of visual memory across immediate and long delays. Boston Naming Test is a measure of confrontation/picture naming. Letter Fluency requires the examinee to rapidly generate words based on letter cues and Category Fluency is considered a measure of semantic fluency. WRAT-4 Word Reading is a measure of single word reading skills. The Trail Making Test assesses psychomotor processing speed and mental flexibility. Wechsler Adult Intelligence Scale-IV (WAIS-IV) Similarities subtest assesses verbal abstract reasoning, Block Design is a measure of visuoconstruction and planning, Digit Span is an auditory working memory measure. HVLT, Hopkins Verbal Learning Test-Revised is a word list encoding and memory measure. Standard scores have a mean of 100 and standard deviation of 15. Scaled scores have a mean of 10 and standard deviation of 3.

(e.g., onset and course of cognitive symptoms, functioning in activities of daily living, psychosocial history, medical history, and brain imaging result summary), a battery of neuropsychological tests (typically, most tests have been well-validated with available demographically-matched normative data). Neuropsychological assessment of PPA in our clinic often includes a variety of language measures (repetition, confrontation naming, rapid word generation/verbal fluency tasks, phonological decoding, comprehension, reading, vocabulary, praxis, and qualitative characterization of speech fluency), as well as assessment of other cognitive domains (memory, attention, visuospatial functions, executive functioning). As this was a clinical sample and PET testing is not readily covered by insurance, such biomarker data was not available for this sample.

#### *Procedure and materials*

Patient MR images were analyzed retrospectively by two neuroradiologists (blinded reads with rating system as described below) and the quantitative imaging software, Neuroreader™, and all procedures were approved by the Institutional Review Board at a Midwestern US medical center in accordance with the Helsinki Declaration. Data were obtained retrospectively from patient medical records and were entered into a larger clinical database comprised mostly of patients with early-stage neurodegenerative conditions, all of whom completed a neuropsychological evaluation within 1 year of MR Neuroreader™. Patient data were not included in the larger database if the clinician concluded that cognitive impairment was due to a non-neurodegenerative condition (e.g., severe mental illness, seizure, traumatic brain injury), or if there was focal brain pathology (e.g., neoplasm) noted by the clinical neuroradiologist. The clinical sample used for present analyses included patients with MCI most consistent with lvPPA as diagnosed and patients diagnosed with aMCI due to possible AD (without indication of PPA presentation). We selected the aMCI due to possible AD cohort as a comparison group as it is less heterogenous compared to non-amnesic MCI and is a common differential for MCI due to lvPPA. While we would have liked to include biomarker data that would have provided further insight into possible/probable underlying etiology of aMCI (i.e., AD biomarker data), unfortunately these measures were not part of routine clinical care and, therefore, were not in patient medical records and unavailable for analysis.

Table 3  
Brain regions analyzed by MR Neuroreader™  
and rated by neuroradiologists

Brain Regions
Right Frontal Lobe
Left Frontal Lobe
Right Parietal Lobe
Left Parietal Lobe
Right Occipital Lobe
Left Occipital Lobe
Right Temporal Lobe
Left Temporal Lobe
Right Hippocampus
Left Hippocampus

For the present analyses, because we were interested in comparing Neuroreader™ with visual ratings by neuroradiologists, we focused analyses to only brain regions for which there were data captured by both methods: lobar regions and the hippocampus. At this time, Neuroreader™ does not provide more fine-grained metrics for subregions of the hippocampus or cortex (e.g., no gyral level volumes like parahippocampal volume). Listed in Table 3 are the brain regions that were included in analyses for the current study, based on known atrophy patterns in the clinical samples used and availability of both neuroradiologist visual assessment and MR Neuroreader™ volumetric data. Additional volumes generated by Neuroreader™ were excluded from present analyses as they were considered redundant (e.g., additional total volume metric for hippocampus, frontal lobe, parietal lobe, occipital lobe, temporal lobe reflecting sum of left and right volumes), irrelevant for present analysis (e.g., three global measures of volume, four cerebrospinal fluid metrics, three cerebellar volumes, and brainstem volume), or were considered less reliable (i.e., total and bilateral volumes for subcortical regions: putamen, thalamus, ventral diencephalon, pallidum, and caudate).

#### *Radiologist's visual assessment*

Two board-certified, senior neuroradiologists (who are the primary physicians who read scans for neurodegenerative cases) visually rated the MRIs of each patient. Neuroradiology reads scans for over 500 neurodegenerative cases annually, which is approximately 4.3% of total yearly brain MRIs. The raters were provided with the patient's age and gender but were blinded to any clinical information or diagnoses. MRI examinations were performed on any one of the five 1.5T (Tesla) or five 3T MR systems (GE Healthcare Clinical Systems, Wauwatosa,

WI, USA; Siemens Medical Solutions, Erlangen, Germany) across the enterprise. All scans were in alignment with the recommended scan protocols advocated by the Neuroreader™ vender. Axial FLAIR and T2 WI were also available for radiologist review. Each rater rated 10 bilateral brain regions (frontal, parietal, occipital, temporal, and hippocampus) on: 1) volume (age-appropriate versus low volume for age), 2) presence of left-right lobar asymmetry (rated presence of greater left-sided atrophy and greater right-sided atrophy ratings for each region), and 3) lobar sulcal grading for each lobe [21].

### Neuroreader

Participants' MRI scans were processed using the Neuroreader™ software program. Briefly, each participant's scan was segmented into absolute volumes which were then compared against an age- and gender-matched normative sample, resulting in *z*-scores for each of the 10 brain regions. The segmentation algorithm and normative database for Neuroreader™ are components of the closed-source FDA-cleared software and were not available, but the reader is referred to methods described previously by Ahdidan and colleagues [22]. At this time, NeuroReader™ calculates one asymmetry index (AI) for hippocampal volume. The value is computed by this formula: Volume of left – volume of right/volume of left + volume of right × 100. An AI of 0 means *L*=*R*, a lower AI means decreased asymmetry and higher AI means increased asymmetry. Raw structural volumes are used for computation without normalization to intracranial volume. We included the hippocampal asymmetry index, but primarily defined asymmetry in the present analyses based on *z*-score comparisons.

### Statistical analysis

Cohen's Kappa coefficients were calculated to determine interrater reliability between neuroradiologist visual assessment of volume loss, asymmetry, and lobar sulcal grade. Mann-Whitney U tests were calculated to examine group differences between Neuroreader™ *z*-score values, and Wilcoxon Signed Rank tests were calculated to compare Neuroreader™ *z*-scores across different regions within groups. In order to control for false positives associated with multiple comparisons, a false-discovery rate (FDR) correction, the Benjamini–Hochberg procedure, was performed.

## RESULTS

### Neuroradiologist ratings and interrater reliability

Cohen's Kappa analyses were run to establish interrater agreement on volume loss, lobar sulcal grade, and asymmetry across the 10 brain regions in a sample of lvPPA patients and aMCI patients. A complete list of Kappa coefficients across brain regions is presented in Table 4. The ratings of presence or absence of atrophy and asymmetry for each region by each rater are also depicted in Table 5. Interrater agreement was strongest and most significant between raters when rating presence of brain region volume loss, compared to agreement on brain region asymmetry and lobar sulcal grade. There were no significant differences in agreement between groups for any of the metrics examined, therefore, results for interrater agreement will be presented for the combined group.

Table 4  
Interrater Reliability (Cohen's Kappa Coefficients)

Brain Region	Cohen's Kappa Statistic	<i>p</i>
Volume		
Right Frontal	0.427	0.012
Left Frontal	0.319	ns
Right Parietal	0.485	0.013
Left Parietal	0.485	0.013
Right Occipital	0.589	0.005
Left Occipital	0.506	0.015
Right Temporal	0.633	0.003
Left Temporal	0.654	0.001
Right Hippocampal	0.482	0.024
Left Hippocampal	0.465	0.025
Asymmetry		
Right Frontal	0.000	ns
Left Frontal	0.000	ns
Right Parietal	0.000	ns
Left Parietal	0.000	ns
Right Occipital	0.000	ns
Left Occipital	0.000	ns
Right Temporal	0.000	ns
Left Temporal	0.075	ns
Right Hippocampal	0.330	ns
Left Hippocampal	0.623	0.001
Lobar Sulcal Grade		
Right Frontal	0.288	ns
Left Frontal	0.288	ns
Right Parietal	0.323	ns
Left Parietal	0.349	0.036
Right Occipital	0.041	ns
Left Occipital	0.135	ns
Right Temporal	0.061	ns
Left Temporal	0.069	ns

Table 5  
Neuroreader™ z-scores for lvPPA and aMCI patients

		Right hippocampus	Left hippocampus	Right frontal lobe	Left frontal lobe	Right parietal lobe	Left parietal lobe	Right temporal lobe	Left temporal lobe	Right occipital lobe	Left occipital lobe	Hippocampal Asymmetry Index
Logopenic N = 11	1	0.42	<u>0.22</u>	<u>-0.15</u>	<u>-0.53</u>	<u>0.24</u>	<u>-0.01</u>	<u>0.25</u>	<u>-0.31</u>	-0.16	-0.20	-0.44
	2	1.24	0.95	0.07	-0.15	-1.01	<b>-1.53</b>	-0.51	<b>-1.73</b>	0.12	-0.54	-0.29
	3	-0.10	-0.08	<u>-0.09</u>	<u>-0.03</u>	<u>-0.91</u>	<u>-1.02</u>	<u>-1.06</u>	<b>-1.50</b>	<u>-0.43</u>	<u>-0.33</u>	0.05
	4	0.15	-0.09	-0.79	-0.99	-0.78	-1.34	0.27	-0.76	-0.17	-1.25	-0.44
	5	0.27	0.04	<u>-0.60</u>	<u>-0.70</u>	<u>-0.41</u>	<u>-0.47</u>	<u>-0.23</u>	<u>-0.74</u>	<u>0.25</u>	<u>-0.19</u>	-0.40
	6	0.29	<u>-0.23</u>	<u>-0.66</u>	<u>-1.34</u>	<u>-0.74</u>	<b>-2.03</b>	<u>-0.59</u>	<b>-1.99</b>	<u>-0.12</u>	<u>-0.02</u>	-1.04
	7	<u>-0.41</u>	<u>-0.56</u>	<u>-0.84</u>	<u>-0.56</u>	<u>-1.24</u>	<b>-1.51</b>	<b>-1.92</b>	<b>-1.60</b>	<u>-0.70</u>	<u>-0.96</u>	-0.28
	8	<u>0.06</u>	<u>0.39</u>	<u>-0.66</u>	<u>-0.82</u>	<u>-0.46</u>	<u>-1.12</u>	<u>-0.12</u>	<u>-1.36</u>	<u>0.25</u>	<u>-0.39</u>	0.63
	9	0.26	0.24	-0.30	-0.23	-0.07	-0.37	-0.19	-0.08	0.21	0.07	-0.03
	10	0.20	<u>-0.41</u>	<u>-1.24</u>	<u>-1.04</u>	<b>-1.57</b>	<u>-1.23</u>	-1.14	<u>-1.07</u>	-0.69	-0.43	-1.21
	11	0.42	0.26	<u>-0.14</u>	<u>-0.20</u>	0.07	<u>-0.12</u>	0.64	<u>-0.52</u>	0.99	0.06	-0.32
	Mean	0.2545	0.0664	-0.4909	-0.5991	-0.6255	-0.9773	-0.4182	-1.0600	-0.0409	-0.3800	-0.34
aMCI N = 12	1	0.04	0.11	<u>-1.08</u>	<u>-0.93</u>	<u>-1.22</u>	<u>-1.32</u>	-0.65	-0.65	-0.80	-0.70	0.12
	2	<b>-1.58</b>	<b>-1.59</b>	<u>-1.00</u>	<u>-0.98</u>	<u>-1.00</u>	<u>-1.07</u>	<u>-1.45</u>	<u>-1.39</u>	<u>-0.33</u>	<u>-0.79</u>	-0.41
	3	<u>-1.18</u>	-0.61	<u>-0.73</u>	<u>-0.68</u>	<u>-0.68</u>	<u>-0.71</u>	<u>-0.33</u>	-0.48	0.10	0.29	1.47
	4	<u>-1.29</u>	<u>-1.40</u>	<u>0.10</u>	<u>0.05</u>	0.05	-0.39	<u>-0.47</u>	<u>-0.38</u>	0.05	-0.58	-0.15
	5	-0.65	-0.65	0.24	0.23	0.08	-0.18	-0.25	0.08	0.71	0.48	-0.02
	6	<u>-1.02</u>	<u>-1.13</u>	<u>-0.40</u>	<u>-0.53</u>	<u>-0.93</u>	<u>-1.04</u>	<u>0.19</u>	<u>-0.30</u>	<b>-1.64</b>	<u>-0.67</u>	-0.20
	7	<u>-0.96</u>	<u>-1.27</u>	<u>-0.72</u>	<u>-0.69</u>	<u>-0.68</u>	<u>-0.64</u>	<u>-0.67</u>	<u>-0.69</u>	<u>-0.56</u>	<u>-0.48</u>	-0.77
	8	<u>-0.72</u>	-1.31	<u>0.15</u>	<u>-0.20</u>	0.44	-0.16	<u>-1.10</u>	<b>-2.04</b>	0.43	<u>-0.36</u>	-1.56
	9	-0.23	<u>-1.30</u>	<u>-0.63</u>	<u>-0.54</u>	<u>-0.26</u>	<u>-0.43</u>	<u>-0.58</u>	<b>-1.52</b>	0.45	0.19	-2.65
	10	-0.21	-0.78	-0.32	<u>-0.39</u>	-0.11	-0.22	0.24	-0.53	-0.24	-0.25	-1.29
	11	-1.32	-0.64	<u>-0.56</u>	<u>-0.28</u>	<u>-0.42</u>	<u>-1.07</u>	-1.39	-1.27	0.02	-0.38	1.38
	12	-0.68	-0.69	0.44	0.22	0.26	0.04	0.12	-0.23	0.01	-0.13	-0.07
Mean	-0.8167	-0.9383	-0.3758	-0.3933	-0.3725	-0.5992	-0.5283	-0.7833	-0.1500	-0.2817	-0.35	

Gray highlighted text denotes low volume for age as rated by rater 1. Underlined text denotes low volume for age as rated by rater 2. Neuroreader z-scores that were 1.5 standard deviations below the normative mean were bolded.

### Brain region volume loss

Cohen's Kappa coefficients demonstrated moderate to substantial agreement between raters regarding the presence of right-sided atrophy for: frontal ( $K_R FV = 0.427$ ,  $p = 0.012$ ), parietal ( $K_R PV = 0.427$ ,  $p = 0.012$ ), hippocampal ( $K_R HV = 0.482$ ,  $p = 0.024$ ), occipital ( $K_R OV = 0.589$ ,  $p = 0.005$ ), and temporal ( $K_R TV = 0.633$ ,  $p = 0.003$ ) brain regions. Similarly, Cohen's Kappa coefficients again revealed moderate to substantial agreement between raters regarding the presence of left-side atrophy across regions: hippocampal ( $K_L HV = 0.465$ ,  $p = 0.025$ ), parietal ( $K_L PV = 0.485$ ,  $p = 0.013$ ), occipital ( $K_L OV = 0.506$ ,  $p = 0.015$ ), and temporal ( $K_L TV = 0.654$ ,  $p = 0.001$ ) brain regions.

### Lobar sulcal grade

Interrater reliability analyses yielded no significant agreement for presence of right-sided lobar sulcal

grade between raters ( $K = 0.000 - 0.330$ ,  $p > 0.05$ ). In contrast to the right hemisphere, interrater reliability analyses yielded significant agreement between raters for lobar sulcal grade of left parietal ( $K_{LPG} = 0.349$ ,  $p = 0.036$ ) and left hippocampal brain regions ( $K = 0.623$ ,  $p = 0.001$ ). Interrater reliability analyses were not significant for left lobar sulcal grade for frontal, occipital, or temporal regions ( $K = 0.069 - 0.288$ ,  $p > 0.05$ ).

### Right-left asymmetry ratings

Interrater reliability analyses yielded no significant agreement for presence of right-sided ( $K = 0.000 - 0.330$ ,  $p > 0.05$ ) or left sided asymmetry ratings ( $K = 0.000 - 0.075$ ,  $p > 0.05$ ) between raters. Rather, neuroradiologists were more likely to agree on the absence of asymmetry. More specifically, within the lvPPA patients, neuroradiologists agreed 91% of the time that left asymmetry was absent in frontal

and parietal lobes (9% of the time disagreed) and 64% agreed left atrophy was absent in temporal lobes (36% of the time raters disagreed). Across aMCI patients, raters agreed 100% of the time that left asymmetry was absent in frontal and parietal lobes, agreed 83% of the time no left asymmetry in temporal lobes (17% of the time they disagreed). Given that the raters never agreed that left asymmetry was present (more frequently agreed asymmetry was absent), sensitivity and specificity statistics for detecting left-side asymmetry were not calculated.

### MR Neuroreader™

Results revealed significant differences (uncorrected and corrected) in Neuroreader™  $z$ -scores in aMCI compared to lvPPA patients for the hippocampus (right:  $U=4.0$ ,  $p<0.000$ ; left:  $U=6.0$ ,  $p<0.000$ ; Benjamini-Hochberg  $p$ -value= $<0.000$ ). There were no significant between group differences for cortical regions examined ( $ps>0.10$  for both left and right frontal, temporal, and parietal regions). Neuroreader™  $z$ -scores for lvPPA and aMCI individual cases are listed in Table 5.

Within the lvPPA group, Wilcoxon signed rank tests revealed significantly lower (all significant when uncorrected for multiple comparisons but not for all corrected comparisons)  $z$ -scores in the left compared to right hemisphere for the parietal ( $T=7.0$ ,  $z=-2.312$ ,  $p=0.021$ ; Benjamini-Hochberg  $p$ -value= $0.052$ ) and temporal cortices ( $T=6.0$ ,  $z=-2.401$ ,  $p=0.016$ ; Benjamini-Hochberg  $p$ -value= $0.052$ ). Marginally significantly lower  $z$ -scores emerged in the left compared to the right hemisphere for the hippocampus ( $T=10.5$ ,  $z=-2.001$ ,  $p=0.045$ ; Benjamini-Hochberg  $p$ -value= $0.056$ ) and occipital ( $T=10.5$ ,  $z=-2.002$ ,  $p=0.045$ ; Benjamini-Hochberg  $p$ -value= $0.056$ ) cortices, though failed to reach statistical significance after controlling for multiple comparisons. There was no significant difference between frontal  $z$ -scores ( $p=0.247$ ).

In the aMCI group, Wilcoxon signed rank tests revealed significantly lower  $z$ -scores in the left compared to right hemisphere for the parietal lobes ( $p=0.004$ ; Benjamini-Hochberg  $p$ -value= $0.01$ ) and the hippocampus ( $p=0.002$ ; Benjamini-Hochberg  $p$ -value= $0.01$ ). There were no significant differences (i.e., asymmetry) for the temporal ( $p=0.109$ ), occipital ( $p=0.189$ ), or frontal cortices ( $p=0.844$ ).

### Neuroreader versus rater

A primary goal of this study was to compare visual assessment to the automated software for the

diagnosis of lvPPA. Unfortunately, because visual assessment failed to yield agreement between raters on the presence of left > right atrophy (as lvPPA is associated with greatest left-sided atrophy, primarily in parietotemporal regions), this could not be fully examined. Rather, raters were more likely to agree on the absence of atrophy. Rater agreement for presence of volume loss by region for each patient is presented in Table 4 along with the Neuroreader™  $z$ -scores (Rater 1 indicated with a gray box and Rater 2 indicated with underlined text). Neuroreader identified 2 patients as having left parietal and left temporal lobe atrophy in the lvPPA group and identified 1 subject as having hippocampal volume loss within the aMCI group.

## DISCUSSION

The current study found that while interrater reliability results evidenced moderate to substantial agreement on the presence or absence of low volume for age across most lobar regions, agreement on left-right asymmetry ratings in lvPPA by blinded neuroradiologists was absent. Given the diagnosis is supported by left-sided atrophy in temporal and parietal regions, this finding raises concern for the ability to reliably detect lvPPA with visual assessment. For both the lvPPA and aMCI groups, neuroradiologists never agreed that left asymmetry was present in examined regions. Interrater reliability indicated moderate to substantial agreement on the presence or absence of low volume for age across most lobar regions. The reader can examine Table 5 for cases in which raters concluded that atrophy was abnormal for age along with corresponding Neuroreader™  $z$ -score data for each patient ( $z$ -score values that were 1.5 SD below the mean or more). This table shows there were instances where rater consensus (both raters agreed atrophy was abnormal or normal for age) was in conflict with  $z$ -scores generated by Neuroreader™. Neuroreader™  $z$ -scores did not significantly differ between lvPPA and aMCI groups across left or right frontal, temporal, or parietal regions. However, in the lvPPA group, Wilcoxon signed rank tests revealed significantly lower  $z$ -scores in the left compared to right hemisphere for the parietal and temporal lobes. Within the aMCI group, Wilcoxon signed rank tests revealed significantly lower  $z$ -scores in the left compared to right hemisphere for the hippocampus and parietal lobes. Overall, our results indicate that both visual reads and Neuroreader™ analyses cannot reli-

ably differentiate clinically diagnosed lvPPA and aMCI. This is possibly, in part, related to the complexities of diagnosing PPA. Specifically, clinical diagnosis of PPA is challenging due to overlap in symptoms across the variants, clinical variability in presentation (i.e., lvPPA and nfvPPA) [23, 24], and tests used to assess speech and language functions. These factors can affect the relationship between clinical diagnosis and neuroimaging biomarkers.

While the current study is the first to examine the clinical utility of Neuroreader™ in a lvPPA sample of patients diagnosed with comprehensive neuropsychological testing, there are several limitations. First, we had a small sampling of neuroimaging readers, and the patient sample sizes were small, which impacted power to detect differences. We attempted to address this by limiting our sample to two relatively homogenous samples, logopenic variant of PPA and aMCI; and by limiting our variables of analysis to only the most relevant volumes (i.e., focused on the cortical and hippocampal volumes). Second, as we limited our analyses of PPA to the lvPPA phenotype (sample sizes were 3 or smaller for svPPA and nfvPPA), we were unable to draw conclusions about PPA more broadly. With a larger sample size, future work examining the use of Neuroreader™ with the agrammatic/nonfluent and semantic PPA variants would be of great interest. Third, current analyses did not include a typically-aging control group. This would be of great interest in future work to better characterize using these methods to differentiate normal aging from atypical aging. Also, this is a cross-sectional study and longitudinal data are needed. Finally, as the data in present analyses was collected from a clinical database, valuable amyloid biomarker data frequently included in AD research was not available, as these variables were not part of routine clinical care for our samples.

Limitations of Neuroreader™ include z-score data for only regional volumes and not for atrophy in specific gyri such as the superior temporal gyrus. Prior neuroimaging studies in lvPPA report that left posterior peri-Sylvian or parietal atrophy is specific to this variant but absence of atrophy in these regions does not rule out this variant; indeed, one study demonstrated that in the consideration of two distinct and common neuroimaging biomarkers for each of the variants of PPA, patients with lvPPA were less likely to show both positive neuroimaging biomarkers [11, 25]. In order to improve clinical studies that incorporate neuroimaging data, Neuroreader™ and

similar clinically-available automated segmentation programs may need to include data at the gyral and sulcal level. Future research should also compare volumetric data obtained via Neuroreader™ to volumetric data obtained via widely used open source brain MRI software program (i.e., FreeSurfer) to ensure its continued value.

### Conclusions

Our findings suggest that while clinical neuroradiologists generally reached agreement on the presence of atrophy across lobar regions, identifying the left-sided asymmetry characteristic in lvPPA was significantly more challenging. This highlights the need for more fine-grained approaches to detecting asymmetrical atrophy patterns in conditions like lvPPA. While Neuroreader™ z-scores captured greater left compared to right volume loss in hippocampus, parietal, occipital, and temporal cortices, similar left-sided atrophy was observed using neuroreader for hippocampus and parietal lobes in the aMCI group, as well. Unfortunately, Neuroreader™ z-scores could not differentiate aMCI from lvPPA when group comparisons were made across the key cortical regions affected in lvPPA. Given NeuroReader's development based on hippocampal segmentation [22], hippocampal z-scores were expectedly lower in the aMCI group than in lvPPA. Clearly, Neuroreader™ continues to have utility in identifying hippocampal volume loss; however, its value in aiding in the clinical conceptualization of more complex cortical neurodegenerative patterns has not yet been realized. The analyses presented here suggest the need for using caution if one is considering employing Neuroreader™ to aid in the differential diagnosis of neurodegenerative conditions. Undoubtedly, clinicians should continue to consider the entire, comprehensive clinical picture in the diagnosis of PPA variants (consistent with FDA-clearance guidance). Identifying early onset neurodegenerative processes is imperative to the field's understanding and ultimate prevention and treatment of these conditions. Unfortunately, identifying asymmetric patterns of atrophy proved formidable for both clinical neuroradiologists and the software package Neuroreader™. Further research is needed to better understand the clinical utility of software packages such as Neuroreader™ in the diagnosis of more complex neurodegenerative conditions, such as the three PPA variants and other neurodegenerative conditions (e.g., frontotemporal degeneration).

## ACKNOWLEDGMENTS

Data used in the current study was obtained from clinical records.

## FUNDING

The authors have no funding to report.

## CONFLICT OF INTEREST

The authors have no conflict of interest to report.

## REFERENCES

- [1] Mesulam MM (1982) Slowly progressive aphasia without generalized dementia. *Ann Neurol* **11**, 592-598.
- [2] Mesulam MM (2001) Primary progressive aphasia. *Ann Neurol* **49**, 425-32.
- [3] Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, Ogar JM, Rohrer JD, Black S, Boeve BF, Manes F, Dronkers NF, Vandenberghe R, Rascovsky K, Patterson K, Miller BL, Knopman DS, Hodges JR, Mesulam MM, Grossman M (2011) Classification of primary progressive aphasia and its variants. *Neurology* **76**, 1006-1014.
- [4] Botha H, Duffy JR, Whitwell JL, Strand EA, Machulda MM, Schwarz CG, Reid RI, Spychalla AJ, Senjem ML, Jones DT, Lowe V, Jack CR, Josephs KA (2015) Classification and clinicoradiologic features of primary progressive aphasia (PPA) and apraxia of speech. *Cortex* **69**, 220-36.
- [5] Butts AM, Machulda MM, Duffy JR, Strand EA, Whitwell JL, Josephs KA (2015) Neuropsychological profiles differ among the three variants of primary progressive aphasia. *J Int Neuropsychol Soc* **21**, 429-35.
- [6] Krishnan K, Machulda MM, Whitwell JL, Butts AM, Duffy JR, Strand EA, Senjem ML, Spychalla AJ, Jack CR, Lowe VJ, Josephs KA (2017) Varying degrees of temporoparietal hypometabolism on FDG-PET reveal amyloid-positive logopenic primary progressive aphasia is not a homogeneous clinical entity. *J Alzheimers Dis* **55**, 1019-1029.
- [7] Leyton CE, Villemagne VL, Savage S, Pike KE, Ballard KJ, Piguet O, Burrell JR, Rowe CC, Hodges JR (2011) Subtypes of progressive aphasia: Application of the International Consensus Criteria and validation using beta-amyloid imaging. *Brain* **134**, 3030-3043.
- [8] Marshall CR, Hardy CJD, Volkmer A, Russell LL, Bond RL, Fletcher PD, Clark CN, Mummery CJ, Schott JM, Rossor MN, Fox NC, Crutch SJ, Rohrer JD, Warren JD (2018) Primary progressive aphasia: A clinical approach. *J Neurol* **265**, 1474-1490.
- [9] Sajjadi SA, Patterson K, Nestor PJ (2014) Logopenic, mixed, or Alzheimer-related aphasia? *Neurology* **82**, 1127-1131.
- [10] Whitwell JL, Duffy JR, Strand EA, Machulda MM, Senjem ML, Schwarz CG, Reid R, Baker MC, Perkerson RB, Lowe VJ, Rademakers R, Jack CR Jr, Josephs KA (2015) Clinical and neuroimaging biomarkers of amyloid-negative logopenic primary progressive aphasia. *Brain Lang* **142**, 45-53.
- [11] Sajjadi SA, Sheikh-Bahaei N, Cross J, Gillard JH, Scoffings D, Nestor PJ (2017) Can MRI visual assessment differentiate the variants of primary-progressive aphasia? *AJNR Am J Neuroradiol* **38**, 954-960.
- [12] NeuroQuant. Accessed December 14, 2016, <https://www.advancedradiology.com/our-services/mri/neuroquant%C2%AE>
- [13] (2016) *NeuroReader*. Brainreader.
- [14] Fischl B (2012) FreeSurfer. *Neuroimage* **62**, 774-781.
- [15] Tanpitukpongse TP, Mazurowski MA, Ikkena J, Petrella JR (2017) Predictive utility of marketed volumetric software tools in subjects at risk for Alzheimer disease: Do regions outside the hippocampus matter? *AJNR Am J Neuroradiol* **38**, 546-552.
- [16] Azab M, Carone M, Ying SH, Yousem DM (2015) Mesial temporal sclerosis: Accuracy of NeuroQuant versus neuro-radiologist. *AJNR Am J Neuroradiol* **36**, 1400-1406.
- [17] Persson K, Barca ML, Cavallin L, Brækhus A, Knapsgog AB, Selbæk G, Engedal K (2018) Comparison of automated volumetry of the hippocampus using NeuroQuant® and visual assessment of the medial temporal lobe in Alzheimer's disease. *Acta Radiol* **59**, 997-1001.
- [18] Ross DE, Ochs AL, Seabaugh JM, Shrader CR (2013) Man versus machine: Comparison of radiologists' interpretations and NeuroQuant(R) volumetric analyses of brain MRIs in patients with traumatic brain injury. *J Neuropsychiatry Clin Neurosci* **25**, 32-39.
- [19] Westman E, Cavallin L, Muehlboeck JS, Zhang Y, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Spenger C, Lovestone S, Simmons A, Wahlund LO (2011) Sensitivity and specificity of medial temporal lobe visual ratings and multivariate regional MRI classification in Alzheimer's disease. *PLoS One* **6**, e22506.
- [20] Louis S, Morita-Sherman M, Jones S, Vegh D, Bingaman W, Blumcke I, Obuchowski N, Cendes F, Jehi L (2020) Hippocampal sclerosis detection with NeuroQuant compared with neuroradiologists. *AJNR Am J Neuroradiol* **41**, 591-597.
- [21] Pasquier F, Leys D, Weerts JG, Mounier-Vehier F, Barkhof F, Scheltens P (1996) Inter- and intraobserver reproducibility of cerebral atrophy assessment on MRI scans with hemispheric infarcts. *Eur Neurol* **36**, 268-272.
- [22] Ahdidan J, Raji CA, DeYoe EA, Mathis J, Noe KØ, Rimestad J, Kjeldsen TK, Mosegaard J, Becker JT, Lopez O (2016) Quantitative neuroimaging software for clinical assessment of hippocampal volumes on MR imaging. *J Alzheimers Dis* **49**, 723-732.
- [23] Sajjadi SA, Patterson K, Arnold RJ, Watson PC, Nestor PJ (2012) Primary progressive aphasia: A tale of two syndromes and the rest. *Neurology* **78**, 1670-1677.
- [24] Wicklund MR, Duffy JR, Strand EA, Machulda MM, Whitwell JL, Josephs KA (2014) Quantitative application of the primary progressive aphasia consensus criteria. *Neurology* **82**, 1119-1126.
- [25] Gil-Navarro S, Lladó A, Rami L, Castellví M, Bosch B, Bargalló N, Lomeña F, Reñé R, Montagut N, Antonell A, Molinuevo JL, Sánchez-Valle R (2013) Neuroimaging and biochemical markers in the three variants of primary progressive aphasia. *Dement Geriatr Cogn Disord* **35**, 106-117.