RESEARCH ARTICLE

Brain MRI in Progressive Supranuclear Palsy with Richardson's Syndrome and Variant Phenotypes

Mike P. Wattjes, MD, PhD,[1*] Hans-Jürgen Huppertz, MD,[2] Nima Mahmoudi, MD,[1] Sophia Stöcklein, MD,[3] Sophia Rogozinski, MD,[4] Florian Wegner, MD,[4] Martin Klietz, MD,[4] Ivayla Apostolova, MD,[5] Johannes Levin, MD,[6,7,8] Sabrina Katzdobler, MD,[6,7,8] Carsten Buhmann, MD,[9] Andrea Quattrone, MD, PhD,[6,10] Georg Berding, MD,[11] Matthias Brendel, MD,[7,8,12] Henryk Barthel, MD,[13] Osama Sabri, MD,[13] Günter Höglinger, MD,[4,6,7] Ralph Buchert, PhD,[5] and for the Alzheimer's Disease Neuroimaging Initiative

[1]*Department of Neuroradiology, Hannover Medical School, Hannover, Germany*
[2]*Swiss Epilepsy Clinic, Klinik Lengg, Zurich, Switzerland*
[3]*Department of Radiology, University Hospital of Munich, LMU Munich, Munich, Germany*
[4]*Department of Neurology, Hannover Medical School, Hannover, Germany*
[5]*Department of Diagnostic and Interventional Radiology and Nuclear Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*
[6]*Department of Neurology, University Hospital, LMU Munich, Munich, Germany*
[7]*German Center for Neurodegenerative Diseases (DZNE) Munich, Munich, Germany*
[8]*Munich Cluster for Systems Neurology (SyNergy), Munich, Germany*
[9]*Department of Neurology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*
[10]*Institute of Neurology, Department of Medical and Surgical Sciences, University "Magna Graecia" of Catanzaro, Catanzaro, Italy*
[11]*Department of Nuclear Medicine, Hannover Medical School, Hannover, Germany*
[12]*Department of Nuclear Medicine, University Hospital of Munich, LMU Munich, Munich, Germany*
[13]*Department of Nuclear Medicine, University Hospital of Leipzig, Leipzig, Germany*

ABSTRACT: Background: Brain magnetic resonance imaging (MRI) is used to support the diagnosis of progressive supranuclear palsy (PSP). However, the value of visual descriptive, manual planimetric, automatic volumetric MRI markers and fully automatic categorization is unclear, particularly regarding PSP predominance types other than Richardson's syndrome (RS).

*__Correspondence to:__ Dr. Mike P. Wattjes, Department of Diagnostic and Interventional Neuroradiology, Hannover Medical School, Carl-Neuberg-Straße 1, D-30625 Hannover, Germany; E-mail: wattjes.mike@mh-hannover.de

Mike P. Wattjes and Hans-Jürgen Huppertz are first authors contributed equally.

Günter Höglinger and Ralph Buchert are senior authors contributed equally.

Relevant conflicts of interest/financial disclosures: The authors received no specific funding for this study. There were no conflicts of interest relevant to this study to declare. M.P.W. received speaker or consultancy honoraria from Alexion, Bayer Healthcare, Biogen, Biologix, Bristol Myers Squibb, Celgene, Genilac, Imcyse, IXICO, Icometrix, Medison, Merck-Serono, Novartis, Roche, Sanofi-Genzyme. Publication royalties from Springer and Elsevier. H.J.H. has used atlas-based volumetric MRI analysis in industry-sponsored research projects. N.M. has nothing to disclose. S.S. has nothing to disclose. S.R. has nothing to disclose. F.W. has nothing to disclose. M.K. received honoraria for scientific presentations from AbbVie and Ever Pharma. I.A. has nothing to disclose J.L. reports speaker fees from Bayer Vital, Biogen, EISAI, TEVA and Roche, consulting fees from Axon Neuroscience and Biogen, author fees from Thieme medical publishers and W.K. medical publishers and is inventor in a patent "Oral Phenylbutyrate for Treatment of Human 4-Repeat Tauopathies" (EP 23 156 122.6) filed by LMU Munich. In addition, he reports compensation for serving as chief medical officer for MODAG, is beneficiary of the phantom share program of MODAG GmbH and is inventor in a patent "Pharmaceutical Composition and Methods of Use" (EP 22 159 408.8) filed by MODAG, all activities outside the submitted work. S.K. reports travel support from Life Molecular Imaging outside the submitted work. C.B. received a grant from the Hilde-Ulrichs-Stiftung, served as a consultant for Bial, Hormosan Pharma, Merz Pharmaceuticals, and Zambon and received honoraria for scientific presentations from AbbVie, Bial, Stada Pharma, TAD Pharma, UCB Pharma, and Zambon. A.Q. received grant support not related to this study from the Italian Society for Parkinson and movement disorders LIMPE-DISMOV. G.B. has nothing to disclose. M.B. has nothing to disclose. H.B. received reader honoraria from Life Molecular Imaging and speaker honoraria from Novartis/AAA. O.S. received research support from Life Molecular Imaging. G.U.H. served as a consultant for Abbvie, Alzprotect, Aprineua, Asceneuron, Bial, Biogen, Biohaven, Kyowa Kirin, Lundbeck, Novartis, Retrotope, Roche, Sanofi, UCB; received honoraria for scientific presentations from AbbVie, Bayer Vital, Bial, Biogen, Bristol Myers Squibb, Kyowa Kirin, Roche, Teva, UCB, Zambon. R.B. has nothing to disclose.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Received: 24 March 2023; **Revised:** 10 June 2023; **Accepted:** 20 June 2023

Published online 6 August 2023 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.29527

Objectives: To compare different visual reading strategies and automatic classification of T1-weighted MRI for detection of PSP in a typical clinical cohort including PSP-RS and (non-RS) variant PSP (vPSP) patients.
Methods: Forty-one patients (21 RS, 20 vPSP) and 46 healthy controls were included. Three readers using three strategies performed MRI analysis: exclusively visual reading using descriptive signs (hummingbird, morning-glory, Mickey-Mouse), visual reading supported by manual planimetry measures, and visual reading supported by automatic volumetry. Fully automatic classification was performed using a pre-trained support vector machine (SVM) on the results of atlas-based volumetry.
Results: All tested methods achieved higher specificity than sensitivity. Limited sensitivity was driven to large extent by false negative vPSP cases. Support by automatic volumetry resulted in the highest accuracy (75.1% ± 3.5%) among the visual strategies, but performed not better than the midbrain area (75.9%), the best single planimetric measure. Automatic classification by SVM clearly outperformed all other methods (accuracy, 87.4%), representing the only method to provide clinically useful sensitivity also in vPSP (70.0%).
Conclusions: Fully automatic classification of volumetric MRI measures using machine learning methods outperforms visual MRI analysis without and with planimetry or volumetry support, particularly regarding diagnosis of vPSP, suggesting the use in settings with a broad phenotypic PSP spectrum.

Key Words: progressive supranuclear palsy; magnetic resonance imaging; hummingbird sign; volumetry; machine learning

Introduction

Progressive supranuclear palsy (PSP) is a four-repeat tauopathy characterized by atrophy of the subthalamic nucleus and brainstem tegmentum and depigmentation of the substantia nigra.[1-3] PSP patients present with postural instability, oculomotor dysfunction, supranuclear gaze palsy, and cognitive/executive symptoms to strongly varying extent, depending on the regional distribution of tau pathology and neurodegeneration.[4,5] Therefore, PSP presents with various clinical phenotypes, the most frequent of which is the "classical" PSP-Richardson's syndrome (PSP-RS). However, the phenotypic spectrum typically encountered in clinical practice is a broad continuum overlapping with other neurodegenerative disorders, rendering early and accurate diagnosis of PSP challenging.[3-5]

The clinical diagnosis and phenotypical characterization of PSP is mainly based on core clinical and supportive features that have been incorporated into the Movement Disorder Society (MDS) diagnostic criteria.[4] Magnetic resonance imaging (MRI)-based supportive features include regional brain atrophy predominantly in the midbrain.[4]

Several descriptive MRI features indicating midbrain atrophy have been proposed and are frequently reported in clinical routine including the hummingbird sign, the Mickey-Mouse sign, and the morning-glory sign.[6-8] Various planimetric measures and combinations of planimetric measures have been validated in PSP, not exclusively focusing on the brain stem, but including adjacent structures such as the cerebellar peduncles.[9-14] More recently, fully automated MRI analysis combining atlas-based volumetry with support vector machine (SVM) classification has been proposed for the (differential) diagnosis of neurodegenerative parkinsonian syndromes including PSP.[15] The added value of this investigator-independent approach beyond descriptive and planimetric measures has not been tested so far. Furthermore, the majority of previous studies on descriptive, planimetric or SVM-based MRI in PSP included cohorts over representing PSP-RS.[1,2,4] Data on their use in (non-RS) variant PSP (vPSP) subtypes is limited.[15]

This retrospective study investigated the diagnostic performance of visual descriptive, manual planimetric, and automatic SVM classification of T1-weighted MRI for detection of PSP in a multicentric cohort representing the broad phenotypic PSP spectrum in clinical practice.

Methods

Patients and Healthy Controls

The study was designed as phase 2 study ("ability to discriminate patients from controls") according to the five-phase framework for biomarker validtion[16] and, therefore, included well-characterized patients with established PSP diagnosis and healthy controls.

Patients were enrolled at the university hospitals of Augsburg (A), Hamburg (HH), Hannover (H), Leipzig (L), and Munich (M) using the following inclusion criteria: (1) clinical PSP diagnosis according to the MDS criteria by a movement disorder specialist; and (2) brain MRI (including T1-weighted and fluid attenuated inversion recovery/T2-weighted sequences) digitally available for retrospective processing. Exclusion criteria were: (1) insufficient magnetic resonance (MR) image quality for visual, manual planimetric or automatic volumetric analysis; and (2) clinically relevant comorbidity visible on brain MRI such as severe

ischemic small vessel disease (grade 3 according to Fazekas' scale[17]), large vessel disease, or brain tumor. Sixty-three patients fulfilling the inclusion criteria were identified. From these, 18 patients (28.6%) were excluded because the T1-weighted sequence was not adequate for volumetric analysis. Four patients (6.3%) were excluded because of severe ischemic small vessel disease. The remaining 41 patients (n = 1/A, 9/HH, 17/H, 5/L, 9/M) were included. The categories of diagnostic certainty were probable/possible/suggestive-of PSP in 35 (85.4%)/1 (2.4%)/5 (12.2%) of these patients.[4] Information on PSP phenotype and clinical presentation is given in Table 1. Patients with vPSP were subcategorized as "cortical" vPSP (predominant corticobasal syndrome or frontal presentation) or "subcortical" vPSP (predominant parkinsonism or progressive gait freezing).[18]

Waiver of informed consent for the retrospective analysis of the anonymized patient data was obtained from the relevant ethics review boards. All procedures were in accordance with the ethical standards of these ethics review boards and with the 1964 Helsinki declaration and its later amendments.

Forty-six sex- and age-matched healthy subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI-1) were included as healthy controls (HC). The first of the two back-to-back 3D T1-weighted scans from the baseline MRI session was used (without any preprocessing by the ADNI imaging corelab). All HC subjects had been cognitively stable for ≥36 months after MRI.

MR Acquisition

MRI had been performed at 3/1.5 Tesla in 12/34 (26.1%/73.9%) of the HC subjects with 34 different scanners from three different manufacturers (41.3%/10.9%/47.8% of the subjects with Siemens Healthineers, Erlangen, Germany/Philips Healthcare, Best, Netherland/GE [GE Healthcare, Chicago, Illinois, USA] scanner).[19] Among the PSP patients, MRIs had been performed at 3/1.5 Tesla in 28/13 patients (68.3%/31.7%) with 15 different scanners from three different manufacturers (92.5%/5.0%/2.5% of the patients with Siemens Healthineers, Erlangen, Germany/Philips Healthcare, Best, Netherlands/Toshiba [Toshiba Medical Systems, Otawara, Japan] scanner).

Manual MRI Planimetry

Manual planimetry of T1-weighted MRI was performed by the same experienced rater in all subjects as described previously.[10,14] The following planimetric measures were obtained: anterior–posterior midbrain diameter, midbrain area (MA), MA to pons area ratio (MAtoPA), middle cerebellar peduncle diameter to superior cerebellar peduncle diameter ratio (MCPtoSCP), third ventricle width to frontal horn width ratio (V3toFH), V3toFH/MAtoPA, MR Parkinson Index (MRPI) = MCPtoSCP/MAtoPA, and MRPI 2.0 = MRPI × V3toFH.

Automatic Atlas-Based MRI Volumetry

Automatic atlas-based MRI volumetry was performed as described previously.[15] In brief, the T1-weighted MR images were segmented into grey matter, white matter, and cerebrospinal fluid maps, which then were stereotactically normalized into the anatomical reference space of the Montreal Neurological Institute. Modulation was applied to preserve the total amount of each tissue class. Forty-four different compartments, brain structures, and planes based on masks predefined in the reference space were evaluated.[15] Volumetric measures were scaled to the total intracranial volume (ICV) and then corrected for age based on linear regression of normative values from an independent sample of 73 HC (age, 63.3 ± 7.6 years, range, 48–79 years, 49.3% females) described previously.[15] Next, ICV- and age-corrected measures were transformed to z-scores based on mean and standard deviation in an independent normal database of 30 ADNI-1 HC subjects (74.1 ± 3.8 years, 50.0% females, no overlap with the HC subjects in the test dataset).

For each subject, a bar plot was prepared to represent the ICV- and age-corrected z-scores in a standardized fashion (Fig. S1).

Visual MRI Analysis

The T1-weighted MR images were converted from DICOM to NIFTI format and then back to DICOM format to remove all information from the DICOM headers that might identify a subject as PSP patient or HC. The resulting DICOM images were pseudonymized in randomized order. Only the pseudonymized DICOM images were provided for visual interpretation to three independent readers (M.P.W., S.S., and N.M.) experienced in MRI reading for the diagnosis of movement disorders. All readers used the Visage platform (Visage Imaging, Berlin, Germany) to assess the images. The readers were blinded for clinical data except sex and age.

Three strategies were tested for visual MRI analysis. In the first strategy ("MRI visual only") the readers were asked to first check the images for the presence (yes/no) of the hummingbird sign,[7] the Mickey-Mouse sign,[8] and the morning-glory sign,[6] and then to provide a dichotomic summary assessment as either "PSP patient" or "HC". There was no specific instruction given whether or not the presence of one or more descriptive features should lead to the classification as "PSP patient". In the second strategy ("MRI visual + manual planimetry"), the readers were asked for the same binary assessment based on visual assessment of

TABLE 1 *Demographical and clinical characteristics*

	All PSP patients (n = 41)	PSP-RS (n = 21)	vPSP (n = 20)	Cortical vPSP (n = 11)	Subcortical vPSP (n = 9)	Healthy controls (n = 46)
Age at MRI, years	69.6 ± 7.8 (53.1–81.1)	69.6 ± 7.4 (53.1–81.1)	69.7 ± 8.5 (53.2–80.8)	68.3 ± 8.1 (53.2–79.9)	71.4 ± 9.1 (57.3–80.8)	72.2 ± 4.7 (59.8–79.9)
Sex (% females)	48.8	47.6	50.0	63.6	33.3	39.1
Disease duration at MRI [y]	3.1 ± 2.6 (0.3–12.2)	3.2 ± 3.1 (0.4–12.2)	2.9 ± 2.0 (0.3–6.8)	2.4 ± 1.4 (0.9–5.0)	3.6 ± 2.5 (0.3–6.8)	–
PSP rating scale scores[a]	29.8 ± 14.1 (9–86)	**35.1 ± 15.9** (15–86)[d]	**24.3 ± 9.5** (9–47)[d]	26.3 ± 11.4 (9–47)	21.9 ± 6.4 (14–31)	–
MoCA scores[b]	21.4 ± 4.4 (10–29, n = 32)	20.9 ± 4.2 (11–29, n = 17)	21.9 ± 4.7 (10–28, n = 15)	**19.0 ± 5.1** (10–24, n = 6)[e]	**23.9 ± 3.3** (19–28)[e]	–
GDS scores[c]	5.7 ± 2.6 (1–12, n = 28)	5.1 ± 2.5 (1–10, n = 14)	6.2 ± 2.7 (3–12, n = 14)	7.1 ± 3.1 (3–12, n = 7)	5.3 ± 2.0 (3–9, n = 7)	–

Note: The metric parameters were compared (1) between all PSP patients and healthy controls; (2) between PSP-RS and vPSP; and (3) between cortical vPSP and subcortical vPSP using the homo- or heteroscedastic unpaired *t* test depending on the result of Levene's test of homogeneity of variance. The distribution of sex was compared between groups using the χ^2 test. A group difference was considered significant if two-sided $P < 0.05$. No correction for multiple testing was performed. If a parameter was not available in all subjects, the number of subjects for that parameter is given in parentheses.

[a]According to PSP rating scale.[30]

[b]According to MoCA.[31]

[c]According to GDS.[32]

[d]PSP-RS versus vPSP <0.05. The *p*-value for the difference between PSP-RS and vPSP with respect to the PSP rating scale score was 0.013.

[e]Cortical syndrome versus subcortical syndrome $P < 0.05$.

Abbreviations: PSP, progressive supranuclear palsy; PSP-RS, PSP with Richardson's syndrome; vPSP, (non-RS) variant PSP; MRI, magnetic resonance imaging; y, year; MoCA, Montreal Cognitive Assessment; GDS, Geriatric Depression Scale. The *p*-value for the difference between cortical vPSP and subcortical vPSP with respect to the MoCA score was 0.042.

T1-weighted MRI supported by the planimetric measures. For cutoffs on the planimetric measures for the differentiation of PSP from HC, readers were referred to previous studies.[10,11,14] There was no specific instruction given whether pathological values of one or more planimetric measures should lead to the classification as "PSP patient". In the third strategy ("MRI visual + automatic volumetry"), the readers were asked for the same binary summary assessment based on visual inspection of T1-weighted MRI supported by the bar plot of the ICV- and age-corrected z-scores from automatic volumetry (Supplementary Fig. S1). There was no specific instruction whether specific z-scores or patterns of z-scores should lead to the classification as "PSP patient".

To assess intra-reader variability, each strategy was used twice by each reader. The mean time between the two reading sessions for a given strategy was 14 ± 4 weeks (range, 9–20 weeks). Randomization of the images was different for each session.

Automatic SVM Classification of the T1-Weighted MRI Dataset

For automatic classification of T1-weighted MRI as "PSP patient" or "healthy control", an SVM was used that had been trained previously using a fully independent dataset.[15] The input to the SVM consists of the 44 different features from automatic volumetry, all corrected for ICV and age and scaled to a range of 0 to 1. The output of the SVM consists in an SVM-PSP score that ranges between 0 (most likely "healthy control") and 1 (most likely "PSP patient"). The 2016 pre-trained SVM was used without changes.

Statistical Analyses

Cross tables, sensitivity, specificity, accuracy, balanced accuracy, the percentage of discrepant cases, and Cohen's κ were used to characterize accuracy and/or intra- and inter-reader variability of the visual MRI analysis, separately for each reading strategy, each reader (or each pair of readers), and for the majority read.

Pearson's correlation analysis was used to test manual planimetric measures and the SVM-PSP score for impact of healthy aging and, if meaningful, ICV. This was done in the independent normal database of 30 ADNI1 HC. Correction for age (and ICV) of a given parameter in the test set was performed by computing residuals with respect to the linear regression of this parameter in the independent normal database.

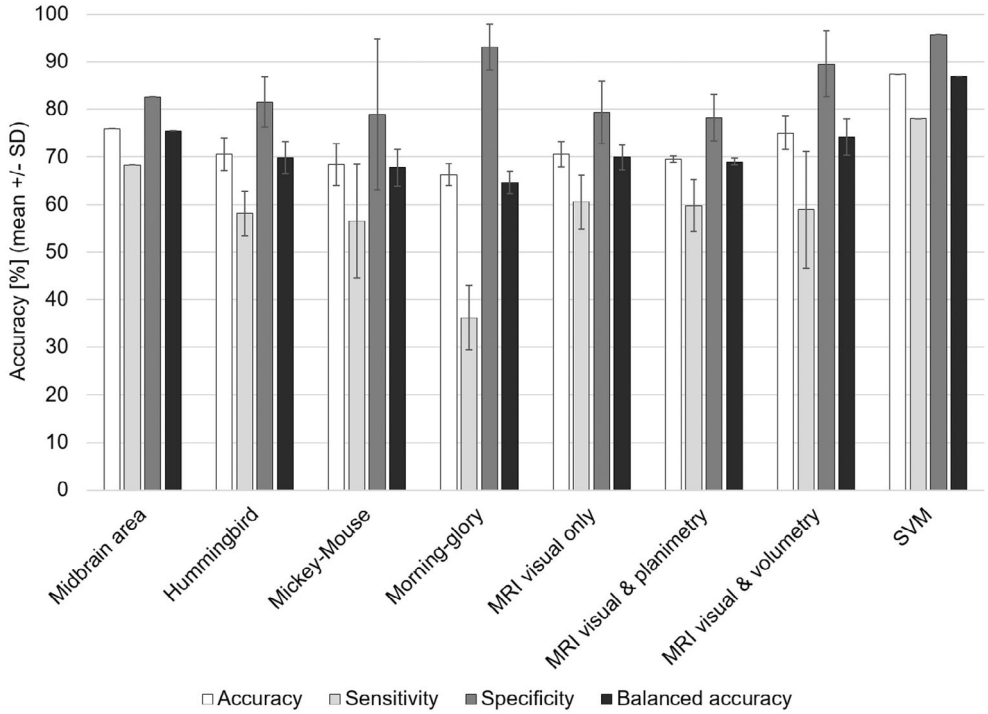

FIG. 1. Mean accuracy of the binary visual interpretation in the whole sample (n = 87) with respect to the visual detection of progressive supranuclear palsy (PSP) based on the hummingbird sign, the Mickey-Mouse sign, the morning-glory sign, and according to the three tested visual magnetic resonance imaging (MRI) reading strategies. Mean value and standard deviation were computed across the three readers and across the two reading sessions. Sensitivity and specificity of the binary visual interpretation in the whole sample for each individual reader are given in Supplementary Figure S2. The accuracy of the majority read of the three readers is given in Supplementary Figure S3. The accuracy estimates for the dichotomized manual midbrain area and for the fully automatic support vector machine (SVM) classification using cutoffs derived from the corresponding receiver operating characteristic curves for detection of PSP in the whole sample are also given.

Receiver operating characteristic (ROC) analysis was used to characterize the use of planimetric measures and the SVM-PSP score for the detection of PSP. DeLong's test was used to compare the area under the (correlated) ROC curves between different measures.[20] Cutoffs for dichotomization were obtained from ROC curves using Youden's criterion.[21]

Statistical analyses were performed in the whole sample and in the following subgroups: PSP-RS, vPSP, cortical vPSP, and subcortical vPSP. IBM SPSS (version 27) was used. Statistical significance was assumed if two-sided $P \leq 0.05$.

Results

Diagnostic Performance of Visual MRI Analysis

Performance of the visual analysis in the whole sample is given in Figure 1 (exact values in Supplementary Table S1) and Supplementary Figures S2 and S3. Among the descriptive features, the hummingbird sign had the highest accuracy (70.5% ± 3.4%, mean ± standard deviation across readers and reading sessions). Among the visual strategies, "MRI visual + automatic volumetry" showed the best accuracy (75.1% ± 3.5%).

Specificity was considerably higher than sensitivity for all descriptive features and all strategies. The highest specificity was provided by the morning-glory sign (93.1% ± 4.8%).

The percentage of misclassified cases was considerably higher in vPSP for all descriptive features and all visual reading strategies (Fig. 2).

Supplementary Figure S4 compares cognitive performance in different domains between true positive and false negative cases.

Manual Planimetric Measures and Automatic Classification

The results of the ROC analyses of the manual planimetric measures and the automatic SVM-PSP score for the discrimination of PSP from HC are given in Table 2. Among the manual planimetric measures, the midbrain area provided the largest area under the ROC curve in the whole sample. The difference was significant compared to all other planimetric measures except midbrain diameter ($P = 0.057$) and midbrain area to pons area ratio ($P = 0.147$). A cutoff on the midbrain area determined from the ROC curve by the Youden's criterion (107 mm^2) achieved overall accuracy,

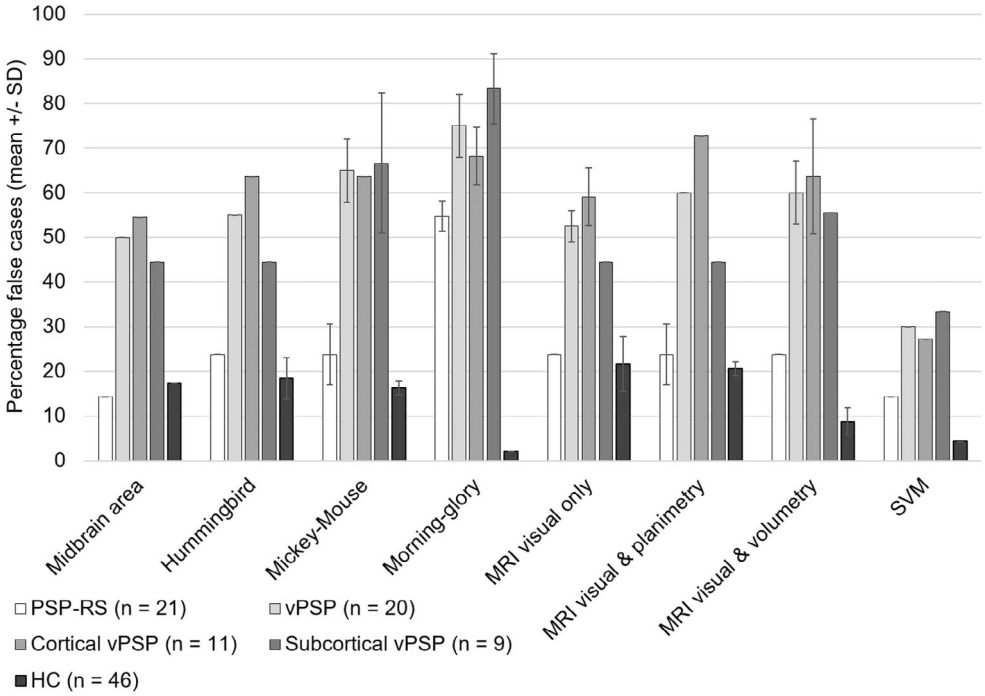

FIG. 2. Percentage of misclassified cases by the majority read of the three readers in the different subgroups. Mean value and standard deviation were computed across the two reading sessions. The percentage of cases misclassified by the dichotomized manual midbrain area and by the automatic support vector machine classification using cutoffs derived from the corresponding receiver operating characteristic curves for detection of progressive supranuclear palsy (PSP) in the whole sample are given for comparison. The false positive rate, that is, the proportion of healthy controls (HC) subjects misclassified as PSP by visual magnetic resonance imaging (MRI) analysis, separately for each descriptive sign and reading strategy and separately for each of the three readers and each of the two reading sessions, is shown in Supplementary Figure S9.

TABLE 2 *Diagnostic performance of the manual planimetric measures and the automatic SVM classification in the differentiation of the PSP patients (all or different subgroups) from the 46 normal controls in the test set*

	All PSP patients (n = 41)	PSP-RS (n = 21)	vPSP (n = 20)	Cortical vPSP (n = 11)	Subcortical vPSP (n = 9)
Morphometric analysis					
MD	0.730 (0.057)	0.887 (0.046)	0.565 (0.084)	0.561 (0.111)	0.569 (0.119)
MA	0.808 (0.047)	0.924 (0.034)	0.686 (0.073)	0.667 (0.097)	0.709 (0.097)
MAtoPA	0.758 (0.053)	0.891 (0.045)	0.618 (0.078)	0.567 (0.102)	0.681 (0.102)
MCPtoSCP	0.486 (0.063)	0.523 (0.082)	0.448 (0.080)	0.462 (0.097)	0.430 (0.117)
V3toFH	0.683 (0.058)	0.732 (0.069)	0.633 (0.075)	0.628 (0.090)	0.638 (0.110)
V3toFH/MAtoPA	0.732 (0.053)	0.827 (0.055)	0.633 (0.075)	0.626 (0.096)	0.640 (0.101)
MRPI	0.679 (0.059)	0.818 (0.067)	0.533 (0.080)	0.545 (0.110)	0.517 (0.098)
MRPI 2.0	0.701 (0.056)	0.813 (0.061)	0.584 (0.075)	0.587 (0.098)	0.580 (0.090)
Automatic volumetry					
SVM-PSP score	0.910 (0.032)	0.950 (0.030)	0.868 (0.052)	0.832 (0.082)	0.913 (0.047)

Note: The table gives the area under the ROC curve, the standard error of the area is given in parentheses.
Abbreviations: SVM, support vector machine; PSP, progressive supranuclear palsy; PSP-RS, PSP with Richardson's syndrome; MD, anterior–posterior midbrain diameter; MA, midbrain area; MAtoPA, MA to pons area ratio; MCPtoSCP, middle cerebellar peduncle diameter to superior cerebellar peduncle diameter ratio; V3toFH, third ventricle width to frontal horn width ratio; MRPI, Magnetic Resonance Parkinson Index = MCPtoSCP/MAtoPA, MRPI 2.0 = MRPI × V3toFH); ROC, receiver operating characteristic.

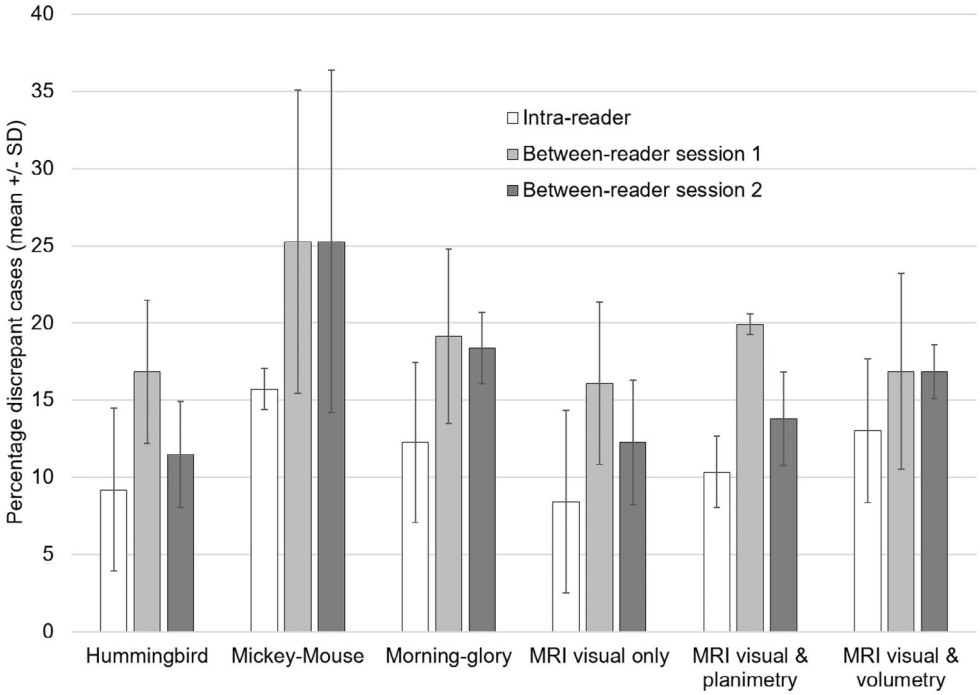

FIG. 3. Mean percentage of intra- and inter-reader discrepant cases in the visual reads in the whole sample (n = 87) with respect to the detection of the hummingbird sign, the Mickey-Mouse sign, and the morning-glory sign, and with respect to the detection of progressive supranuclear palsy according to the three tested magnetic resonance imaging (MRI) reading strategies. For intra-reader agreement, mean value and standard deviation were computed across the three readers. For inter-reader agreement, mean value and standard deviation were computed across the three pairs of readers. Inter-reader agreement was assessed separately for the two reading sessions (to identify possible learning effects). Intra- and inter-reader discrepant cases for each individual reader and for each single pair of readers are shown in Supplementary Figure S6. Cohen's κ values of intra- and inter-reader agreement are shown in Supplementary Figure S7.

sensitivity, specificity, and balanced accuracy of 75.9%, 68.3%, 82.6%, and 75.5%, respectively (Fig. 1, Supplementary Table S1).

Linear regression of the manual midbrain area in the normal database with $ICV^{2/3}$ (to match units) and age as predictors revealed a trend toward a significant effect of normal ageing (standardized regression coefficient ß = −0.359, $P = 0.054$). The manual midbrain area was not associated with $ICV^{2/3}$ (ß = 0.114, $P = 0.528$). When the manual midbrain area was corrected for age and $ICV^{2/3}$ by computing residuals with respect to the regression line in the normal database, the area under the ROC curve in the whole sample increased from 0.808 (standard error, 0.047) to 0.836 (0.043) without statistical significance ($P = 0.320$).

The SVM-PSP score was not correlated with age in the independent normal database (Pearson correlation coefficient $R = -0.277$, $P = 0.138$). It achieved a significantly larger area under the ROC curve in the whole sample than the midbrain area (0.910 vs. 0.808, $P = 0.019$). With the cutoff determined from the ROC curve in the whole sample by Youden's criterion (0.563), the SVM-PSP score provided overall accuracy, sensitivity, specificity, and balanced accuracy of 87.4%, 78.0%, 95.7%, and 86.9%, respectively (Fig. 1, Supplementary Table S1). Using the "natural" cutoff 0.5 on the SVM-PSP score for dichotomization resulted in a small loss of specificity (from 95.7% – 91.3%), sensitivity remained unchanged.

Regarding subgroups, the area under the ROC curve was larger for the SVM-PSP score than for the midbrain area in all subgroups, most pronounced in the vPSP groups (Table 2). Accordingly, the proportion of misclassified (false negative) cases was similar for the SVM-PSP score compared to the midbrain area in the PSP-RS group (14%), but it was clearly lower for the SVM-PSP score in the vPSP group (30% vs. 50%) (Fig. 2). This was also reflected by the median SVM-PSP score being clearly above the cutoff (of 0.563) not only in PSP-RS, but also in vPSP, whereas the median of the midbrain area was clearly below the cutoff (of 107 mm^2) only in PSP-RS and very close to this cutoff in vPSP (Supplementary Fig. S5).

Inter- and Intra-Rater Variability

Results regarding inter- and intra-rater agreement are presented in Figure 3, Supplementary Figures S6–S8, Supplementary Table S2.

Discussion

Descriptive and manual planimetric MRI features are useful to support the diagnosis of PSP-RS.[13] This study

examined descriptive features and planimetric measures in a cohort representing the broad phenotypic PSP spectrum in clinical practice. The cohort included ≈50% vPSP patients allowing direct comparison of MRI performance between PSP-RS and vPSP. In addition, descriptive features and planimetric measures were compared with fully automatic MRI classification by an SVM previously trained in an independent dataset.[15]

Regarding descriptive features, the sensitivity for the detection of PSP-RS was similar for the hummingbird sign and the Mickey-Mouse sign (≈75%), in line with previous studies.[22] It was considerably lower for the morning-glory sign (≈45%). Visual summary assessment of the MRI with and without support by manual planimetry or automatic volumetry did not improve the sensitivity for the detection of PSP-RS compared to the hummingbird sign or the Mickey-Mouse sign alone.

Higher sensitivity for the detection of PSP-RS (≈85%) was achieved by the midbrain area, the best manual planimetric measure. The SVM-PSP score provided about the same (85%) sensitivity for the detection of PSP-RS, in good to excellent agreement with previous studies using the same[15] or a different[23] SVM.

Sensitivity for the detection of vPSP was ≤50% for all visual methods indicating only very limited use of visual MRI analysis in vPSP, particularly in cortical vPSP (sensitivity ≤40%). This is in line with a multicenter study on tau-positron emission tomography (PET) with ^{18}F-PI-2620 and structural MRI for image-based detection of PSP in a sample with ≈40% vPSP that reported 63% accuracy for MRI and 76% accuracy for ^{18}F-PI-2620-PET.[24]

The manual midbrain area did not provide clear improvement in the detection of vPSP compared to the visual methods. In contrast, the automatic SVM-PSP score achieved ≈70% sensitivity for the detection of both cortical and subcortical vPSP, clearly superior to all other methods.

Regarding specificity, the proportion of misclassified (false positive) HC was lowest for the morning-glory sign (in line with its low sensitivity), the automatic SVM-PSP score, and for visual MRI analysis supported by automatic volumetry (Fig. 2, Supplementary Fig. S9). Therefore, support by the bar plots of the ICV- and age-corrected z-scores of automatic volumetric methods improved the specificity of the visual analysis (without change of sensitivity). This might be explained by the fact that the bar plot included many subcortical and cortical brain structures in addition to the midbrain. This might have made visual analysis (and SVM-based classification) more robust with respect to (minor) midbrain alterations. For all other methods (with strict focus on the midbrain), including the manual midbrain area, the proportion of false positive HC was ≈20%.

Taken together, automatic classification of T1-weighted MRI using SVM on automatic volumetric measures not only was the only method that provided clinically useful sensitivity (≈70%) for the detection of vPSP, but also performed equally or better than all other methods with respect to the detection of PSP-RS (≈85% sensitivity) at very high specificity (≈95%). This might be because of the multivariate SVM approach essentially taking into account the whole brain. The findings support the combination of automatic MRI volumetry and SVM classification to complement visual reading and planimetry/volumetry in the etiological diagnosis of clinical uncertain suspicion of PSP.

The good SVM-based detection of vPSP is remarkable, as the SVM was trained for PSP-RS.[15,25] The SVM was not optimized for the current dataset to avoid overly optimistic performance estimates because of overfitting. Therefore, the observed performance estimates should be representative for SVM use in clinical practice. The performance might be further improved by more heterogeneous training samples or combination of different SVM (eg, one for each PSP subtype).

All methods showed considerably larger specificity than sensitivity. This was most pronounced for the morning-glory sign that provided ≈35% sensitivity for the detection of PSP (all types), but was falsely detected in only 2% of the HC subjects, in line with previous studies.[26] Therefore, the morning-glory sign is a particularly strong indicator of PSP.

Among the manual planimetric measures, the midbrain area provided the best discriminative power, in line with a previous multicenter study without harmonization of MRI sequences.[10] Post-mortem studies demonstrated ante-mortem MRI-based volumetric midbrain measures to be useful surrogates of tau pathology in subcortical and brainstem regions in primary four-repeat tauopathies including PSP.[27] The rather low sensitivity of midbrain area alterations to detect PSP in the current study is most likely because midbrain atrophy is not consistently detected in vPSP.[28] The composite MR Parkinson indices, MRPI and MRPI 2.0, performed clearly worse than the midbrain area for both, vPSP and PSP-RS. The area under the ROC curve for vPSP was close to chance level for both indices suggesting that additional physiological inter-subjects variability and additional measurement errors caused by adding regions outweighs the potential benefit from PSP-effects in the additional regions. Operator dependence of manual planimetry might have caused loss of performance, more pronounced for the composite MRPI and MRPI 2.0 than for the individual measures such as the midbrain area. To test this hypothesis, we retrospectively performed fully automatic planimetry using the algorithms of Quattrone and colleagues[29] (Supplementary Figs. S10–S13). Automatic measurement did not have a relevant impact on the performance of MRI planimetry. In particular, the midbrain area outperformed MRPI and

MRPI 2.0 in the current dataset also when using the automatic algorithms.

Inter-reader variability of the visual methods was ≈15% to 25% in the current study, larger than reported previously.[24,26] This was despite the fact that all readers were experienced in MRI reading in movement disorders. The reason for this difference is unclear. In particular, it most likely cannot be fully explained by the large fraction of vPSP in the current study, because inter-reader variability was not systematically larger in vPSP than in PSP-RS. Intra-reader variability was lower than inter-reader variability, but still sizeable (≈10%–15%). Therefore, a relevant fraction of the rather large inter-reader variability can be explained by intra-reader variability indicating that visual reading of T1-weighted MRI for the diagnosis of PSP is difficult. Among the descriptive MRI features, the detection (or exclusion) of the hummingbird sign showed the best intra- and inter-reader stability, particularly in PSP-RS. Interestingly, support by manual planimetry or automatic volumetry did not improve intra- and inter-reader stability. This might be explained by the fact that no specific instructions regarding the use of the additional information were given. Therefore, for use in clinical routine, specific instructions should be provided.

Inter-scanner variability of the MR image characteristics (eg, grey-to-white matter contrast) might also have contributed to inter-reader variability. It probably also impacted on manual planimetry and automatic volumetry. This might be considered a strength of the study, because the lack of strict standardization of MR image acquisition reflects clinical routine in the tertiary setting, as many patients referred to a movement disorder specialist have already received the MRI scan in different institutions. Inter-scanner variability (including differences between PSP patients and HC subjects with respect to the distribution of scanner manufacturers) most likely did not affect the comparison of the different MRI analysis methods. In particular, the fully automatic SVM analysis used the same volumetric data as provided to the readers in the "visual + automatic volumetry" strategy.

There was no eligibility criteria with respect to PSP subtype in this study. Therefore, we assume that the composition of the cohort with respect to PSP-RS and vPSP reflects the tertiary referral setting.

Limitations of this study include the use of the clinical diagnosis without histopathological confirmation as gold standard and the rather small number of PSP patients. Almost a third of the eligible patients had to be excluded because the T1-weighted MRI was not adequate for automatic volumetric analyses. Future guidelines on the diagnosis of parkinsonian syndromes should recommend that structural MRI is performed with sufficient quality to allow reliable automatic volumetry. Another limitation relates to the retrospective and observational character of this study. It cannot be excluded that, at least in some patients, the MRI information was considered in establishing the clinical diagnosis. This would artificially increase the diagnostic accuracy of the MRI analyses.

In conclusion, fully automatic classification of volumetric measures using conventional machine learning methods such as SVMs clearly outperforms (manual and automatic) planimetry as well as visual analysis of T1-weighted MRI (with and without support by planimetry or volumetry) for the detection of PSP in clinical settings with a broad phenotypic spectrum and no harmonization of MR acquisition sequences. We, therefore, recommend the use of such techniques to complement visual and planimetric analyses. This requires that T1-weighted MRI is acquired with adequate contrast between tissue classes and with adequate (more or less isotropic) spatial resolution to allow reliable automatic volumetry. ■

Acknowledgments: We thank Prof. Xiaoqi Ding and Prof. Karl-Titus Hoffmann for their help with the MRI data acquisition. Open Access funding enabled and organized by Projekt DEAL.

Ethical Compliance Statement

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this work is consistent with those guidelines. Waiver of informed consent for the retrospective analyses of this study was obtained from the relevant ethics review boards.

Data Availability Statement

Visual and support vector machine scores are available from the corresponding author

References

1. Dickson DW, Ahmed Z, Algom AA, Tsuboi Y, Josephs KA. Neuropathology of variants of progressive supranuclear palsy. Curr Opin Neurol 2010;23(4):394–400.
2. Roemer SF, Grinberg LT, Crary JF, et al. Rainwater charitable foundation criteria for the neuropathologic diagnosis of progressive supranuclear palsy. Acta Neuropathol 2022;144(4):603–614.
3. Stamelou M, Respondek G, Giagkou N, Whitwell JL, Kovacs GG, Hoglinger GU. Evolving concepts in progressive supranuclear palsy and other 4-repeat tauopathies. Nat Rev Neurol 2021;17(10):601–620.
4. Hoglinger GU, Respondek G, Stamelou M, et al. Clinical diagnosis of progressive Supranuclear palsy: the Movement Disorder Society criteria. Movement Disord 2017;32(6):853–864.
5. Respondek G, Stamelou M, Kurz C, et al. The phenotypic Spectrum of progressive Supranuclear palsy: a retrospective multicenter study of 100 definite cases. Movement Disord 2014;29(14):1758–1766.
6. Adachi M, Kawanami T, Ohshima H, Sugai Y, Hosoya T. Morning glory sign: a particular MR finding in progressive supranuclear palsy. Magn Reson Med Sci 2004;3(3):125–132.
7. Kato N, Arai K, Hattori T. Study of the rostral midbrain atrophy in progressive supranuclear palsy. J Neurol Sci 2003;210(1–2):57–60.

8. Page I, Gaillard F. Descriptive neuroradiology: beyond the hummingbird. Pract Neurol 2020;20(6):463–471.
9. Heim B, Mangesius S, Krismer F, et al. Diagnostic accuracy of MR planimetry in clinically unclassifiable parkinsonism. Parkinsonism Relat D 2021;82:87–91.
10. Moller L, Kassubek J, Sudmeyer M, et al. Manual MRI morphometry in parkinsonian syndromes. Movement Disord 2017;32(5): 778–782.
11. Picillo M, Tepedino MF, Abate F, et al. Midbrain MRI assessments in progressive supranuclear palsy subtypes. J Neurol Neurosur Ps 2020;91(1):98–103.
12. Quattrone A, Antonini A, Vaillancourt DE, et al. A new MRI measure to early differentiate progressive Supranuclear palsy from De novo Parkinson's disease in clinical practice: an international study. Movement Disord 2021;36(3):681–689.
13. Quattrone A, Morelli M, Bianco MG, et al. Magnetic resonance Planimetry in the differential diagnosis between Parkinson's disease and progressive Supranuclear palsy. Brain Sci 2022;12(7):949.
14. Quattrone A, Morelli M, Nigro S, et al. A new MR imaging index for differentiation of progressive supranuclear palsy-parkinsonism from Parkinson's disease. Parkinsonism Relat D 2018;54:3–8.
15. Huppertz HJ, Moller L, Sudmeyer M, et al. Differentiation of neurodegenerative parkinsonian syndromes by volumetric magnetic resonance imaging analysis and support vector machine classification. Mov Disord 2016;31(10):1506–1517.
16. Garibotto V, Herholz K, Boccardi M, et al. Clinical validity of brain fluorodeoxyglucose positron emission tomography as a biomarker for Alzheimer's disease in the context of a structured 5-phase development framework. Neurobiol Aging 2017;52:183–195.
17. Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. AJR Am J Roentgenol 1987;149(2):351–356.
18. Jabbari E, Holland N, Chelban V, et al. Diagnosis across the Spectrum of progressive Supranuclear palsy and Corticobasal syndrome. JAMA Neurol 2020;77(3):377–387.
19. Jack CR, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J Magn Reson Imaging 2008;27(4):685–691.
20. Delong ER, Delong DM, Clarkepearson DI. Comparing the areas under 2 or more correlated receiver operating characteristic curves-a nonparametric approach. Biometrics 1988;44(3):837–845.
21. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3(1): 32–35.
22. Massey LA, Micallef C, Paviour DC, et al. Conventional magnetic resonance imaging in confirmed progressive Supranuclear palsy and multiple system atrophy. Movement Disord 2012;27(14):1755–1762.
23. Mueller K, Jech R, Bonnet C, et al. Disease-specific regions outperform whole-brain approaches in identifying progressive Supranuclear palsy: a multicentric MRI study. Front Neurosci 2017;11:100.
24. Messerschmidt K, Barthel H, Brendel M, et al. (18)F-PI-2620 tau PET improves the imaging diagnosis of progressive Supranuclear palsy. J Nucl Med 2022;63(11):1754–1760.
25. Eckert T, Tang CK, Ma YL, et al. Abnormal metabolic networks in atypical parkinsonism. Movement Disord 2008;23(5):727–733.
26. Mueller C, Hussl A, Krismer F, et al. The diagnostic accuracy of the hummingbird and morning glory sign in patients with neurodegenerative parkinsonism. Parkinsonism Relat Disord 2018;54:90–94.
27. Carlos AF, Tosakulwong N, Weigand SD, et al. Histologic lesion type correlates of magnetic resonance imaging biomarkers in four-repeat tauopathies. Brain Commun 2022;4(3):fcac108.
28. Sakurai K, Tokumaru AM, Shimoji K, et al. Beyond the midbrain atrophy: wide spectrum of structural MRI finding in cases of pathologically proven progressive supranuclear palsy. Neuroradiology 2017;59(5):431–443.
29. Quattrone A, Bianco MG, Antonini A, et al. Development and validation of automated magnetic resonance parkinsonism index 2.0 to distinguish progressive Supranuclear palsy-parkinsonism from Parkinson's disease. Movement Disord 2022;37(6):1272–1281.
30. Golbe LI, Ohman-Strickland PA. A clinical rating scale for progressive supranuclear palsy. Brain 2007;130:1552–1565.
31. Nasreddine ZS, Phillips NA, BÃ©dirian VÃ©, et al. The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005;53(4):695–699.
32. Sheikh JI, Yesavage JA. Geriatric depression scale (GDS): recent evidence and development of a shorter version. Clinical Gerontologist: The Journal of Aging and Mental Health 1986;5:165–173.

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.