Atrophy in midbrain & cerebral/cerebellar pedunculi is characteristic for progressive supranuclear palsy – A double-validation whole-brain meta-analysis

Franziska Albrecht, Sandrine Bisenius, Jane Neumann, Jennifer Whitwell, Matthias L. Schroeter

ARTICLE INFO

Keywords: Imaging biomarker Cerebral pedunculi Cerebellar pedunculi Meta-analysis Midbrain Progressive supranuclear palsy Anatomical likelihood estimation Seed-based D mapping

ABSTRACT

Objective: Progressive supranuclear palsy (PSP) is an atypical parkinsonian syndrome characterized by vertical gaze palsy and postural instability. Midbrain atrophy is suggested as a hallmark, but it has not been validated systematically in whole-brain imaging.

Methods: We conducted whole-brain meta-analyses identifying disease-related atrophy in structural MRI. Eighteen studies were identified (N = 315 PSP, 393 controls) and separated into gray or white matter analyses (15/12). All patients were diagnosed according to the National Institute of Neurological Disorders and Stroke and the Society for PSP (NINDS-SPSP criteria, Litvan et al. (1996a)), which are now considered as PSP-Richardson syndrome (Höglinger et al., 2017). With overlay analyses, we double-validated two meta-analytical algorithms: anatomical likelihood estimation and seed-based D mapping. Additionally, we conducted region-of-interest effect size meta-analyses on radiological biomarkers and subtraction analyses differentiating PSP from Parkinson’s disease.

Results: Whole brain meta-analyses revealed consistent gray matter atrophy in bilateral thalamus, anterior insula, midbrain, and left caudate nucleus. White matter alterations were consistently detected in bilateral superior/middle cerebellar pedunculi, cerebral pedunculi, and midbrain atrophy. Region-of-interest meta-analyses demonstrated that midbrain metrics generally perform very well in distinguishing PSP from other parkinsonian syndromes with strong effect sizes. Subtraction analyses identified the midbrain as differentiating between PSP and Parkinson’s disease.

Conclusions: Our meta-analyses identify gray matter atrophy of the midbrain and white matter atrophy of the cerebral/cerebellar pedunculi and midbrain as characteristic for PSP. Results support the incorporation of structural MRI data, and particularly these structures, into the revised PSP diagnostic criteria.

1. Introduction

Progressive supranuclear palsy (PSP) is a gradually progressive, atypical parkinsonian syndrome characterized by vertical gaze palsy and prominent postural instability with backward falls from disease onset (Litvan et al., 1996a). Commonly applied diagnostic criteria were proposed by the National Institute of Neurological Disorders and Stroke and the Society for PSP (NINDS-SPSP criteria) in 1996 (Litvan et al., 1996a, 1996b). To fulfill the diagnostic criteria for possible PSP, patients’ age of onset should be > 40 years. Additionally, postural instability leading to falls, accompanied by slower vertical saccades, or vertical gaze palsy alone is required for diagnosis. Probable PSP is diagnosed, when a patient shows all of the aforementioned features. Certainty of the underlying tauopathy can only be confirmed by histopathology - the remaining gold standard for diagnosis (Litvan et al., 1996b). Despite advances in research, the evaluation of atypical...
parkinsonian syndromes has remained primarily dependent on clinical diagnostics. Disease-specific, brain imaging, biomarkers have been introduced to increase diagnostic validity of neurodegenerative diseases, for example, in frontotemporal lobar degeneration and primary progressive aphasia (Bisienius et al., 2016; Gorno-Tempini et al., 2011; Rascovszky et al., 2011; Schroeter et al., 2014). In terms of potential biomarkers, previous PSP studies have focused on the local boundary shift integral, longitudinal diffusion changes, or measurements of the midbrain/cerebellum (Kato et al., 2003; Longoni et al., 2011; Oba et al., 2005; Paviour et al., 2006; Slowinski et al., 2008; Zhang et al., 2016). The aforementioned metrics of midbrain and cerebellum were mainly investigated in small cohorts. During the analysis phase of the present study, the diagnostic criteria for PSP were updated in an attempt to increase diagnostic confidence by including neuroimaging measures as supportive and exclusion criteria, highlighting the timeliness of our report. Clinical diagnostic criteria were further improved by including more detailed diagnosis and acknowledging the several phenotypes of PSP. The amended criteria propose that former diagnosis according to Litvan et al. (1996a) relates to the phenotype of PSP-Richardson syndrome. Indeed, diagnostic criteria by Litvan et al. (1996a) have been shown to be 95–100% specific, validated by pathological examination (Höglinger et al., 2017). In the new criteria, in particular the use of midbrain atrophy or hypometabolism and/or postynaptic, striatal, dopaminergic degeneration (Höglinger et al., 2017) has been suggested. Whitwell et al. (2017) published a qualitative review of radiological biomarkers, supporting midbrain and cerebellar pedunculi as specific to PSP even in the early phases of disease. However, it is important to prove the reliability of those disease patterns in quantitative whole-brain analyses.

We investigated the neural correlates of PSP to validate pathomonic signs and further identify possible imaging biomarkers with systematic and quantitative meta-analyses; a thorough and powerful data-driven approach (Bisienius et al., 2016; Schroeter et al., 2014). The meta-analysis focused on whole-brain voxel-based morphometry (VBM) studies that applied structural magnetic resonance imaging (MRI) in PSP. By including only whole-brain studies, we prevented possible vicious circles if only expected brain regions are examined. We investigated disease-associated atrophy separately in gray and white matter in PSP compared with controls. Two widely acknowledged meta-analytical approaches, anatomical likelihood estimation (ALE) (Eickhoff et al., 2012) and anisotropie effect size seed-based D mapping (SDM) (Radua and Mataix-Cols, 2009; Radua et al., 2014), were applied and double-validated against each other. Overlay analyses were used to confirm the findings of the individual approaches, by assessing the degree of overlap. This provided reliable results and enabled us to identify a robust, disease-characteristic imaging biomarker. To investigate disease-specificity we run subtraction analyses to compare PSP with related diseases, i.e. Parkinson’s disease (Albrecht et al., 2018). We hypothesized atrophy in PSP of white matter in the midbrain and gray matter in the midbrain, thalamus, and insulae (Shao et al., 2014; Shi et al., 2013; Yu et al., 2015). Additionally we performed a supportive region-of-interest effect size meta-analysis on the disease-specific imaging marker studies identified by (Whitwell et al., 2017).

2. Materials and methods

2.1. General study selection

The meta-analysis was conducted according to the PRISMA guidelines (www.prisma-statement.org) and Müller et al. (2018). Two authors (FA, SB) independently searched the PubMed database with the following search strategy: (“progressive supranuclear palsy” OR “Steele-Richardson-Olszewski syndrome”) AND (voxel* OR gray matter OR VBM). Studies were included if they: (1) were peer-reviewed, (2) used established diagnostic criteria, (3) were original work, (4) made comparisons with age-matched healthy controls, (5) reported results normalized to a stereotactic space (Talairach or the Montreal Neurological Institute’s (MNI)), and (6) took a whole-brain approach. Region-of-interest analyses, small volume corrected peaks, and case studies were discarded to prevent any regional a priori assumptions. As we did not find enough [18F]-fluorodeoxyglucose positron emission tomography (FDG-PET) or diffusion tensor imaging (DTI) studies fulfilling our criteria, we limited analysis to structural MRI, investigating gray and/or white matter atrophy. The literature search was performed between February and May 2016, reviewing all studies regardless of their publication date. We contacted corresponding authors to request local maxima not reported in their original publications and obtained data (Burciu et al., Premi et al., and Hughes et al.). The corresponding author of Burciu et al. confirmed by email that they used the NINDS-SPS criteria for diagnosis.

2.2. Statistical analyses

2.2.1. Whole-brain ALE & SDM meta-analyses

To obtain robust and reliable neural correlates of PSP, we applied two separate meta-analysis methods, enabling confirmation of the results of each individual technique. Our methods included the widely used approaches of ALE and SDM (Eickhoff et al., 2012; Radua and Mataix-Cols, 2009). For in depth details and differences between the methods see Table e-1. In brief, we first applied SDM (v4.31) (Radua and Mataix-Cols, 2009; Radua et al., 2014). SDM involves rebuilding effect size and variance maps for each study using t-scores and the extracted local maxima (i.e. points of difference between patients and controls). Each map consists of voxels with assigned values depending on their proximity to reported maxima coordinates. Voxels close to reported coordinates have higher values. If voxels lie in the proximity of more than one coordinate, the values are summed. Effect sizes from peaks close to each other are calculated by a weighted average. From these study-specific effect size and variance maps, SDM calculates a mean meta-analytical map, taking between study variability into account. The between study variability is defined as the weighted inverse of the variance of the difference of the studies. The resulting map is tested, voxel-wise, against a null distribution of the meta-analytical values. Analyses were performed on 500 permutations with an anisotropic kernel of 1.0. Results are reported and displayed at p < 0.001, uncorrected for multiple comparisons, because SDM does not correct the brain mask for the number of statistical tests by default. However, to correct the resulting values for multiple comparisons, probabilities were put into the web-converter of SDM (false discovery rate (FDR) correction) (Hochberg and Benjamini, 1990) and are reported in the Supplement (Table e-3). Note that this only provides a peak-based FDR correction for multiple comparisons. To examine the robustness of the main meta-analytical output, jack-knife analyses were carried out by removing one study at a time and repeating the analysis. Clusters present in > 90% of iterations are marked as such in Table e-3.

Secondly, we ran meta-analyses using ALE (GingerALE 2.3.6) (2012). This technique transforms the study-specific extracted peaks, of difference between patients and controls, into Gaussian probability distributions surrounding the coordinates. Coordinates reported in the Talairach stereotactic space were transformed into the stereotactic MNI space using the Lancaster transform implemented in GingerALE (Lancaster et al., 2007). The estimation of the width of these Gaussian probability distributions is adapted for each study using the number of included subjects based on empirical estimates of between-subject and between-template variability. The resulting ALE maps are then combined, across studies, and tested against the null hypothesis of a random spatial distribution between the modeled maps. We first analyzed the data at p < 0.001, uncorrected for multiple comparisons, to ensure comparability with SDM. Thereafter, we corrected the ALE results with FDR, as was done with SDM (Figure e-2) (Laird et al., 2005). FDR is the only method implemented in ALE and SDM, but they use different algorithms.
performed on studies reporting three metrics: midbrain area, midbrain volume, and total brain size meta-analyses identifying the neural correlates of progressive supranuclear palsy with structural magnetic resonance imaging.

Table 1

| Study                  | PSP (N) | HC (N) | Age (years) | Gender (male/female) | Disease Duration (years) | MMSE | UPDRS-III | Notes | Criteria |
|------------------------|---------|--------|-------------|----------------------|--------------------------|------|-----------|-------|----------|
| Boxer et al., 2006     | 15      | 80     | 70.9 ± 6.9  | 9/6                  | 4.8 ± 1.7                | 24.0 ± 3.2 | NR        | ab,     | Litvan   |
| Brenneis et al., 2004  | 12      | 12     | 67.5 ± 6.6  | NR                   | 2.7 ± 0.9                | NR   | 38.9 ± 10.9 | ab     | Litvan   |
| Buruci et al., 2015    | 20      | 20     | 67.8 ± 7.1  | 10/10                | 2.6 ± NR                 | NR   | 39.0 ± 14.5 | a,b,*   | Litvan   |
| Cordato et al., 2005   | 21      | 23     | 70.3 ± 6.4  | 14/7                 | 4.0 ± 2.8                | 25.4 ± 2.2 | 23.1 ± 10.1 | a,b     | Litvan   |
| Ghosh et al., 2012     | 23      | 22     | 71.1 ± 8.6  | 14/9                 | 2.5                      | NR   | 33.8 ± 15.7 | a,b,*   | Litvan   |
| Giordano et al., 2013  | 15      | 15     | 68.9 ± 1.2  | 8/7                  | 3.2 ± 1.3                | 21.2 ± 1.2 | 38.3 ± 4.0 | a,b,*   | Litvan   |
| Kamiyama et al., 2013  | 16      | 21     | 71.4 ± 6.0  | 10/6                 | NR                       | NR   | NR        | b       | Litvan   |
| Lagarde, 2015          | 21      | 18     | 65.5 ± 6.5  | 14/7                 | 4.4 ± 1.7                | 25.8 ± 2.7 | NR        | a,b     | Litvan   |
| Leheucet et al., 2010  | 10      | 9      | 66.9 ± 6.4  | 4/6                  | 4.3 ± 1.0                | 27.0  | 30.0      | a,b     | Litvan   |
| Padovani et al., 2006  | 14      | 13     | 73.0 ± 5.7  | 7/6                  | 3.1 ± 1.0                | 25.8 ± 2.7 | 22.1 ± 8.9 | a,b     | Litvan   |
| Piattella et al., 2015 | 16      | 16     | 68.1 ± 5.9  | 9/7                  | 3.1 ± NR                 | 24.3 ± 3.9 | 27.0 ± 3.9 | a,b,*   | Litvan   |
| Premi et al., 2016     | 32      | 32     | 73.5 ± 6.9  | 17/15                | 6.9 ± 3.5                | 24.9 ± 3.9 | NR        | b       | Litvan   |
| Sandhya et al., 2014   | 10      | 8      | NR          | 9/1                  | NR                       | NR   | NR        | a,b     | Litvan   |
| Takahashi et al., 2011 | 16      | 20     | 64.6 ± 6.4  | 11/5                 | NR                       | 21.0 ± 4.4 | NR        | b       | Litvan   |
| Whitwell et al., 2013  | 16      | 20     | 72.1 ± 4.6  | 8/8                  | 4.1 ± NR                 | 25.8 ± 2.7 | 52.9 ± 12.6 | b,*     | Litvan   |
| Total GM               | 2257    | 3334   | 69.4 ± 2.8  | 140/105              | 3.9 ± 1.2                | 24.5 ± 2.0 | 33.9 ± 9.7 |        |          |

Gender, age, disease duration, MMSE and UPDRS scores are specified for patients (mean ± standard deviation). All MRI studies used 1.5 T (except for * = 3 T). Disease duration mean scores were calculated without Gosh et al. because they reported data as median. Abbreviations: GM gray matter, HC healthy controls, MMSE mini-mental state examination, N number of subjects, NR not reported, PSP progressive supranuclear palsy, UPDRS-III motor score of unified Parkinson’s disease rating scale, WM white matter. a corrected for multiple comparisons; b modulated for patients (mean ± standard deviation).

Thirdly, separate conjunction analyses were performed for gray and white matter by superimposing the results of SDM and ALE, revealing the most consistent atrophic regions. To overcome the potential bias the different FDR algorithms may introduce, we used the results at p < 0.001 uncorrected.

Finally, to investigate disease-specificity of the results, we ran subtraction analyses comparing the present results with the findings of our recently published meta-analysis on Parkinson’s disease (For further information please refer to Albrecht et al., 2018). We used the results at p < 0.001 uncorrected from the present analysis and the white and gray matter MRI cohort of the Parkinson’s disease meta-analysis (cohort with Parkinson’s disease only). Data were analyzed in GingerALE with 10,000 randomizations at p < 0.001 uncorrected. Again, we corrected the ALE results applying FDR afterwards. Results are shown for the contrasts of PSP - Parkinson’s disease for gray and white matter.

To reveal differences in age, gender, disease duration, Mini-Mental-State Examination (MMSE), and the motor score of the Unified Parkinson’s Disease Rating Scale (UPDRS-III) between the gray and white matter cohorts, we performed Pearson’s $\chi^2$ and unpaired t-tests in R with a significance threshold at $p < 0.05$.

2.2.2. Region-of-interest effect size meta-analyses

As supportive analyses, we performed an effect size meta-analysis on the biomarker studies provided by the systematic review of Whitwell et al., 2017. To include recently published studies, two authors searched the Pubmed database applying the same search tags as Whitwell and colleagues (October 2017–March 2018). Analyses were performed on studies reporting three metrics: midbrain area, midbrain-pons area ratio, and the MR Parkinsonism Index (MRPI), a midbrain topons and middle to superior cerebellar pedunculus ratio. Data were analyzed with metafor in R (Viechtbauer, 2010) and normalized by calculating the standardized mean difference (Hedges’ $g$), with 95% confidence intervals as an effect size estimate. A random-effects model was applied to account for both within- and between-study variance.

2.3. Potential sources of bias and error

To avoid biasing the results toward specific brain areas, only quantitative and automated whole-brain studies were included for the whole-brain meta-analyses; region-of-interest studies were excluded. To eliminate age as a confounding variable, we only included comparisons with age-matched controls. Studies comparing PSP with other neurodegenerative diseases were further excluded. In line with the PRISMA recommendations, two investigators performed the literature search independently and discussed discrepancies (FA, SB). SDM and ALE algorithms are able to balance analyses by taking the sample size of each included study into account, thereby equalizing the contributions of each study. It may be the case that only studies finding atrophy in PSP patients were published and others finding null results were not, leading to a publication bias. While we cannot rule this out completely it does seem unlikely as the effect sizes in the published studies tend to be large and the proportion of patients showing quite severe atrophy has been extremely high.
2.4. Data availability statement

Data and analysis methods are shared at request for purposes of replicating procedures and results.

3. Results

3.1. Whole-brain meta-analyses

3.1.1. Study characteristics

A total of 18, gray and white matter, MRI studies, with 315 patients and 393 healthy controls, were included (Table 1, Figure e-1). Some studies reported both measures, leading to the inclusion of 12 white matter and 15 gray matter patient cohorts consisting of 257 and 210 patients respectively. Unpaired Student’s t-tests indicated no significant difference between the mean values of the white and gray matter samples concerning age \( t = 0.28, p = 0.78 \), disease duration \( t = -0.04, p = 0.97 \), and severity of clinical symptoms using MMSE scores \( t = -0.84, p = 0.41 \), and UPDRS-III scores \( t = -0.02, p = 0.99 \). Furthermore, Pearson’s \( \chi^2 \) test showed no significant differences in gender distribution \( \chi^2 = 0.41, p = 0.52 \). There were also no significant differences in standard deviations between the samples regarding age \( t = -0.83, p = 0.42 \), disease duration \( t = -0.44, p = 0.67 \), and severity of clinical symptoms using MMSE scores \( t = -0.26, p = 0.80 \), and UPDRS-III scores \( t = -0.35, p = 0.74 \).

All included studies reported decreased gray/white matter volume or density. There were no reports of gray/white increases found in the literature. Moreover, all studies used exactly the same clinical diagnostic criteria (Litvan et al., 1996a), avoiding a potential vicious circle by the Litvan criteria for PSP. The results of the two separate approaches meta-analyzing PSP (Fig. 2). The results for gray matter revealed that clusters in the bilateral thalamus, bilateral anterior insulae, midbrain, and left caudate nucleus were consistently identified as affected by PSP. The results for white matter analysis revealed consistent converged atrophy bilaterally in the superior and middle cerebellar pedunculi, cerebral pedunculi, midbrain, and right thalamus.

3.1.2. Seed-based D mapping analysis

The upper part of Fig. 1 and Table e-2 illustrate the consolidated neural correlates of PSP identified by SDM meta-analysis. The analysis across structural MRI studies revealed regional gray matter atrophy convergence bilaterally in the inferior and middle frontal and superior temporal gyrus, adjacent to the insulae, in the putamen, caudate nuclei, midcingulate cortex, thalamus, and midbrain. White matter convergence occurred bilaterally in the corpus callosum, thalamus, cerebellar pedunculi, and right middle cerebellar pedunculus.

3.1.3. Anatomical likelihood analysis

The bottom part of Fig. 1 and Table e-3 display the results of the ALE meta-analysis in PSP. The analysis revealed gray matter atrophy convergence in the paracingulate and anterior cingulate gyri, insulae, inferior frontal gyrus, superior parietal lobule, thalamus, caudate nuclei, and cerebellum. White matter regions consistently found were the midbrain, pyramid, thalamus, and cerebellar and superior cerebellar pedunculi. FDR correction of the ALE meta-analysis confirmed involvement of the anterior insula and thalamus for gray matter and midbrain for white matter (see Supplement, Figure e-2).

3.1.4. Overlap analysis

Finally, we performed an overlap analysis by superimposing the results of the two separate approaches meta-analyzing PSP (Fig. 2). The results for gray matter revealed that clusters in the bilateral thalamus, bilateral anterior insulae, midbrain, and left caudate nucleus were consistently identified as affected by PSP. The results for white matter analysis revealed consistent converged atrophy bilaterally in the superior and middle cerebellar pedunculi, cerebral pedunculi, midbrain, and right thalamus.

3.1.5. Subtraction analysis

We performed a subtraction analysis to directly compare disease-specificity of the aforementioned results by integrating white and gray matter atrophy of PSP and Parkinson’s disease patients. The Parkinson’s disease cohort consist of 809 patients from 29 gray matter studies and 75 patients from four white matter studies (Fig. 3, Table e-5). Gray matter atrophy differed between PSP and Parkinson’s disease in the thalamus bilaterally as well as the left insula and claustrum, both surviving multiple comparisons correction. Concerning white matter, subtracting Parkinson’s disease from PSP revealed the midbrain. Note that the white matter analysis did not survive FDR correction, which may have been due to underpowered analysis.

Fig. 1. Brain regions consistently associated with progressive supranuclear palsy. Analysis of gray matter (GM, dark blue) included 257 patients contrasted to 334 healthy subjects. White matter analysis (WM, light blue) included 210 patients contrasted to 265 healthy subjects. Upper image shows effect size estimates (seed-based D mapping, SDM) and bottom image anatomical likelihood estimate (ALE) meta-analyses. Abbreviation: L left. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
3.2. Region-of-interest effect size meta-analyses

3.2.1. Study characteristics

Our literature search identified one study (Kim et al., 2017) that we included additionally to the studies already summarized by (Whitwell et al., 2017) in their systematic review. Study details and demographics are specified in Table-e6 in the Supplement. In total, 15 studies, including 246 patients and 173 healthy controls reported midbrain metrics in three neurodegenerative diseases: PSP (15 cohorts, age = 69.8 ± 3.6 years, disease duration = 3.3 ± 0.8 years), Parkinson’s disease (14 cohorts, age = 66.6 ± 6.7, disease duration = 6.3 ± 2.4), and multiple system atrophy (5 cohorts, age = 64.7 ± 4.3, disease duration = 5.6 ± 1.6). The cohort of PSP patients in the region-of-interest effect size meta-analyses did not differ in age (t = 0.50, p = 0.62) or gender distribution (χ² = 0.77, p = 0.38) compared to the PSP cohort of the whole-brain meta-analyses. Note that disease duration (t = −2.10, p = 0.04) was slightly higher in the PSP cohort of the whole-brain meta-analyses.

3.2.2. Effect size meta-analyses

All random-effects models revealed significant results, highlighting that midbrain metrics generally perform very well in distinguishing PSP from multiple system atrophy and Parkinson’s disease with strong effect sizes (Fig. 4). Note that inter-study heterogeneity was overall quite large. Midbrain area measurements yielded smallest between-study heterogeneity in our study as well as largest effect size (I² = 73.8%, Hedges’ g = −2.79). The highest between-study heterogeneity but still high effect size was observed in midbrain-pons area ratio studies (I² = 90.7%, Hedges’ g = −1.64). Notably, even multiple system atrophy and PSP were distinguished by using midbrain-pons ratio with a high effect size (Hedges’ g = −2.37). MRPI also yielded a large effect size but a high between-study heterogeneity (I² = 89.9%, Hedges’ g = 2.10) in distinguishing PSP from Parkinson’s disease. These results emphasize the importance of our work in validating and replicating these patterns in whole brain with quantitative statistical methods.
Fig. 4. Effect size (Hedges' $g$) meta-analyses of radiological biomarkers of progressive supranuclear palsy (PSP) based on Whitwell et al. (2017) and additional literature search. (1) Measurement using 1.5 Tesla, (2) measurement using 3 Tesla, (3) pathologically confirmed, and (4) clinically diagnosed groups. Abbreviation: CI confidence interval, df degrees of freedom, RE random effects, PD Parkinson’s disease, MSA multiple system atrophy, MRPI Magnetic Resonance Parkinsonism Index.
4. Discussion

To our knowledge, we report the largest most powerful data-driven whole-brain meta-analyses of the neural correlates of PSP to date. The analyses included 315 patients and 393 healthy controls. To help ensure high validity and specificity, all included patients were diagnosed by the same clinical diagnostic criteria, by Litvan et al. (1996a). In the aforementioned criteria imaging biomarkers are not implemented, which avoids a biased and circular design. According to the new PSP criteria (Höglinger et al., 2017), these patients would be referred to as PSP-Richardson syndrome. Note, however, that time criteria are different between both diagnostic systems, that our cohort might have included a minority of subjects with PSP-Parkinsonism additionally to PSP-Richardson syndrome (see Results, 3.1.1), and that one has to be generally very careful in translating old into new clinical criteria see Bisenius et al. (2016 for a related discussion on primary progressive aphasia).

Including only whole-brain studies in our meta-analysis guaranteed a data-driven approach. Notably, the overlap analysis between the two meta-analytical methods revealed gray matter atrophy in four regions in PSP: the bilateral thalamus, bilateral anterior insulae, midbrain, and left caudate nucleus. Concerning white matter atrophy, an overlap of the two methods was observed in the bilateral superior/middle cerebral and cerebellar pedunculi, midbrain, and right thalamus in PSP. Subtraction analysis differentiating PSP from Parkinson’s disease confirmed the midbrain as specific for PSP.

Although previous meta-analyses of gray and white matter atrophy have been conducted for PSP, studies investigated either gray or white matter alterations alone and relied on a single meta-analytical algorithm (Shao et al., 2014; Shi et al., 2013; Yang et al., 2014; Yu et al., 2015). Remarkably, these meta-analyses may have been biased by systematic errors introduced by flawed multiple comparison correction analysis tools, as has been stated recently by (Eickhoff et al., 2016) for former GingerALE software versions. Despite using distributions that contained errors in the FDR calculation, the results are still quite consistent with our own (Yang et al., 2014; Yu et al., 2015).

The present meta-analysis improves on previous studies in several ways. First, we used a novel approach, combining and double-validating two independent meta-analytical techniques to increase validity of our biomarker findings. Second, we included the largest cohort of PSP patients studied to date. Third, we investigated changes in gray and white matter with exactly the same methods to enable comparability of results. Finally, our meta-analyses support region-specific atrophy as an imaging biomarker for PSP – a move toward the inclusion of MRI in the diagnostic workflow for PSP.

4.1. Validating pathognomonic imaging markers for PSP with meta-analyses

Thinning of regions of the midbrain has been suggested as a pathognomonic marker of PSP; previously identified in small cohorts by precisely measuring the diameters of the pons, midbrain, and superior/middle cerebellar pedunculi (Kato et al., 2003; Longoni et al., 2011; Oba et al., 2005; Slowinski et al., 2008). A recent multi-centric volumetric and manual morphometry study has endorsed this view, demonstrating that the midsagittal area and the anterior-posterior diameter of the midbrain distinguishes between PSP and other parkinsonian syndromes (Moller et al., 2017). Previous findings describe three different pathognomonic signs, typically found in PSP patients. First, the ‘hummingbird sign’ reflects a thinning of the anterior part of the midbrain tegmentum (Kato et al., 2003). The second characteristic feature called the ‘penguin silhouette sign’ relates to an
atrophy in the shapes of midbrain tegmentum and pons (Oba et al., 2005). The hummingbird and the penguin silhouette signs are apparent in mid-sagittal MRI scans. Another feature, the ‘Mickey Mouse sign’, describes selective atrophy of the midbrain tegmentum, with relative preservation of the tectum and cerebral pedunculi (Schott, 2007). Axial MRI scans of PSP patients reveal this sign. We were able to lend support to these pathognomonic markers with a large cohort (Fig. 5) by revealing white matter loss in the midbrain (specify the tectum and parts of the tegmentum such as cerebellar pedunculi, red nucleus, and substantia nigra). The Mickey Mouse sign seems less specific, according to our data, because atrophy spread to the cerebral pedunculi (Fig. 5). Due to the algorithm, assumptions concerning the exact neuroanatomical substrates of meta-analyses should be interpreted with caution. Nevertheless, our confidence is increased by the fact that both meta-analytical algorithms across whole-brain studies revealed midbrain atrophy. By replicating previously observed atrophy patterns in a large meta-analytical study, using multiple approaches, we provide further evidence supporting the use of these signs as a diagnostic tool for PSP.

4.2. Atrophy in the midbrain and cerebral/cerebellar pedunculi is disease-characteristic for PSP as compared to other neurodegenerative diseases

Through the conjunctive overlay of the two meta-analytic approaches, we identified gray matter atrophy in PSP patients in the midbrain, thalamus, insulae, and caudate nuclei compared to controls. White matter atrophy was found in the midbrain, thalamus, cerebral pedunculus and superior/middle cerebellar pedunculus. To assess the potential disease-specificity of these patterns we compared our findings to other whole-brain meta-analyses of neurodegenerative diseases qualitatively (Table 2) and quantitatively (Table e-5, Fig. 4). A direct statistical comparison of PSP with another neurodegenerative disease, i.e. Parkinson’s disease (Albrecht et al., 2018), revealed that PSP differs in white matter midbrain atrophy. This highlights again the importance of the midbrain in PSP, especially in distinguishing closely related diseases. Gray matter was significantly impaired in PSP compared to Parkinson’s disease in the thalamus and insula. Note that in qualitative comparisons, gray matter loss in the insulae, caudate nuclei, and thalamus has been recently reported in several other neurodegenerative diseases. Hence, these patterns are neither specific to PSP nor other neurodegenerative diseases.

As also illustrated in Table 2 white matter atrophy in the midbrain has not been observed in other neurodegenerative diseases and thus is suggested as a pathognomonic signature of PSP. PSP shows also severe gray matter atrophy distributed over the whole midbrain indicating disease-relatedness in this comparison. Although gray matter atrophy was detected in ALE meta-analyses in the red nucleus in multiple system atrophy (Shao et al., 2015), it was restricted to this brain region. Moreover, white matter atrophy in the cerebral or superior/middle cerebellar pedunculi seems to be characteristic for PSP as we did not find any other evidence in the literature. Missing effects in the cerebral and cerebellar pedunculi in our meta-analytical subtraction analysis might be related to the small number of white matter studies investigating Parkinson’s disease. Although, one might assume cerebral pedunculi atrophy in amyotrophic lateral sclerosis/motor neuron disease, a recent meta-analysis of DTI data rejects such an assumption (Li et al., 2012b). White matter atrophy in the thalamus/adjacent white matter structures seem to be characteristic for PSP as compared to other meta-analyses of Parkinson’s disease and Alzheimer’s disease (Albrecht et al., 2018; Li et al., 2012a; Yin et al., 2015). As studies on white matter are rare, greatly limiting meta-analytical assessment, disease-specificity should be interpreted with care. Nonetheless, our meta-analyses suggest that gray matter loss in the midbrain and white matter atrophy in the midbrain, cerebral pedunculus, superior/middle cerebellar pedunculi, and thalamus may serve as a disease-characteristic signature for PSP, quantitatively and qualitatively compared to related diagnoses. Further meta-analyses might apply the same methods and

| Neurodegenerative disease                      | Study                                      | Thalamus | Midbrain | Insula | Cerebellar & cerebral pedunculi | Caudate nucleus |
|-----------------------------------------------|--------------------------------------------|----------|----------|--------|--------------------------------|----------------|
| Gray matter                                   |                                            |          |          |        |                                |                |
| Parkinson’s disease                           | (Shao et al., 2014; Shao et al., 2015, Albrecht et al., 2018) | –        | –        | –      | –                              |                |
| Atypical Parkinson’s syndromes                |                                            |          |          |        |                                |                |
| Corticobasal degeneration & syndrome          | (Yu et al., 2015; Albrecht et al., 2017)   | +        | –        | –      | –                              | +              |
| Progressive supranuclear palsy                | (Present study, Shao et al., 2014; Yu et al., 2015; Shi et al., 2013) | +        | +        | –      | –                              | +              |
| Multiple system atrophy                       | (Shao et al., 2015)                        | +        | –        | +      | –                              | –              |
| Lewy body dementia                            | (Zhong et al., 2014)                       | –        | +        | –      | –                              | –              |
| Alzheimer’s disease                           | (Schroeter and Neumann, 2011; Schroeter et al., 2009; Yang et al., 2012a) | +        | +        | –      | –                              | –              |
| Behavioral variant frontotemporal dementia    | (Schroeter et al., 2014; Schroeter and Neumann, 2011; Schroeter et al., 2008; Schroeter et al., 2007; Pan et al., 2012) | –        | –        | +      | –                              | +              |
| Primary progressive aphasias                  |                                            |          |          |        |                                |                |
| Nonfluent/agrammatic variant                  | (Bisenius et al., 2016; Schroeter and Neumann, 2011; Schroeter et al., 2007) | –        | –        | +      | –                              | –              |
| Semantic variant (semantic dementia)          | (Bisenius et al., 2016; Schroeter and Neumann, 2011; Schroeter et al., 2007; Yang et al., 2012b) | –        | –        | –      | –                              | –              |
| Logopenic variant (logopenic aphasia)         | (Bisenius et al., 2016)                    | –        | –        | –      | –                              | –              |
| White matter                                  |                                            |          |          |        |                                |                |
| Parkinson’s disease                           | (Albrecht et al., 2018)                    | –        | –        | –      | –                              | +              |
| Atypical Parkinson’s syndromes                |                                            |          |          |        |                                |                |
| Progressive supranuclear palsy                | (Present study, Yang et al., 2014)         | +        | –        | –      | –                              | –              |
| Alzheimer’s disease                           | (Yin et al., 2015; Li et al., 2012a)       | –        | –        | –      | –                              | –              |

Meta-analyses were conducted by calculating either anatomical likelihood estimates (ALE) or effect-size signed differential mapping (SDM) in the gray or white matter.
parameters to compare several diseases and assess disease-specificity quantitatively by comparing disease patterns with subtraction analyses.

During the course of the present investigation the diagnostic criteria were revised by the Movement Disorder Society-endorsed, P.S.P. Study Group (Höglinger et al., 2017). In this context, (Whitwell et al., 2017) reviewed radiological biomarkers. Our argumentation is in line with their qualitative review. The researchers identified the MRPI, a midbrain to pons and middle to superior cerebellar pedunculus ratio, measured from structural MRI scans, as a supportive criterion even for early clinical diagnosis. Our region-of-interest effect sizes analyses based on Whitwell et al.’s review demonstrate large effect sizes for distinguishing PSP from Parkinson’s disease and multiple system atrophy. Our dual-approach meta-analytic assessment with a large PSP sample is consistent with these recent developments, supporting the use of these particular regions in diagnosing PSP.

It is clearly necessary to validate the applicability of these disease-related MRI biomarkers for individual diagnosis in single patients before being transferable to clinical routine, for example by support vector machine classification for single-patient neurodegenerative disease discrimination (Bisenius et al., 2017; Dukart et al., 2013; Meyer et al., 2017). In an atlas-based MRI volumetry study, high classification accuracy for PSP versus other parkinsonian syndromes (86%) was indeed driven by atrophy in the midbrain and cerebellar pedunculi (Huppertz et al., 2016). Another MRI study suggests that disease-characteristic regions extracted from meta-analyses can outperform whole-brain approaches in identifying PSP by increasing classification accuracy from 80% to 85% (Mueller et al., 2017). Furthermore, combining structural MRI with DTI in support vector machine classification can increase classification results to 100% accurate (Cherubini et al., 2014). One may conclude that our meta-analytically extracted disease-specific atrophy pattern is ready for application as an imaging biomarker for PSP.

4.3. Study limitations

Our study identified consistent findings of atrophy in PSP, but we recognize some limitations. MRI studies on autopsy-proven cases of PSP are extremely rare (Table 1), hence, we could not validate clinical syndromes with histopathology (Albrecht et al., 2017). Furthermore, we could not disentangle the clinical syndromes seen in PSP as too few studies reported those data. Our meta-analyses had to be limited to MRI studies, because insufficient number of studies applied FDG-PET or DTI. We recommend including other imaging modalities and longitudinal data in the future (Bisenius et al., 2016; Schroeter et al., 2014; Schroeter and Neumann, 2011). The accuracy of the disease-specific patterns needs further validation in multi-centric, independent patient cohorts, which provide clinical imaging data, surrogate markers for histopathology from serum/cerebrospinal fluid or post-mortem examination. Researchers have already applied this successfully in Alzheimer’s disease (Klöppel et al., 2008). Due to the lack of combined white matter and gray matter meta-analyses, direct statistical investigation of the results was only possible between PSP and Parkinson’s disease. Note that SDM results are only peak-based FDR corrected for multiple comparisons, which should be preferably replaced by cluster- or voxel-based corrections in the future like it is already available for GingerALE.

4.4. Conclusion

We conducted a comprehensive, systematic, and quantitative meta-analysis investigating the neural correlates of PSP. Patients were all diagnosed applying the same diagnostic criteria of Litvan et al. (1996a), which is now referred to as PSP-Richardson syndrome according to Höglinger et al. (2017). By combining two commonly used meta-analytical algorithms, we identified consistent results validating well-known pathognomonic signs in whole-brain imaging to support the diagnosis of PSP. Results suggest gray matter atrophy in the midbrain and white matter atrophy in the cerebral pedunculus, superior/middle cerebellar pedunculi, and midbrain as imaging biomarkers for diagnosis of PSP ante-mortem. Further, subtraction analyses of meta-analyses of related neurodegenerative diseases could strengthen the disease-specificity of midbrain atrophy by quantitatively comparing atrophy patterns. Taken together, our results are completely consistent with the recent move toward the inclusion of MRI data in the diagnostic workflow for PSP.

Conflict of interest

The authors declare that they have no conflict of interest. All financial disclosures are mentioned in the Acknowledgements.

Acknowledgements

This study has been supported by the Parkinson's Disease Foundation (PDF-IRG-1307), the Michael J Fox Foundation (MJFF-11362), the German Federal Ministry of Education and Research (BMBF; FKZ 01GI1007A; German FTLD consortium), the German Research Foundation (DFG, SCHR 774/5-1), the National Institutes of Health (NIH, R01-NS89757), and the International Max Planck Research School (IMPRS) NeuroCom by the Max Planck Society. Jane Neumann is supported by the Federal Ministry of Education and Research Germany (BMBF, FKZ: 01EO1001), and the German Research Foundation (DFG-SFB1052).

Authors’ roles

1. Research project:
   A. Conception: F. Albrecht, M. Schroeter.
   B. Organization: F. Albrecht, M. Schroeter
   C. Execution: F. Albrecht, S. Bisenius, J. Neumann, M. Schroeter

2. Statistical Analysis
   A. Design: F. Albrecht, J. Neumann, M. Schroeter.
   B. Execution: F. Albrecht, S. Bisenius.
   C. Review and Critique: F. Albrecht, J. Neumann, J. Whitwell, M. Schroeter

3. Manuscript Preparation:
   A. Writing of the first draft: F. Albrecht.
   B. Review and Critique: F. Albrecht, S. Bisenius, J. Neumann, J. Whitwell, M. Schroeter

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2019.101722.

References

Albrecht, F., Bisenius, S., Morales Schaack, R., Neumann, J., Schroeter, M.L., 2017. Disentangling the Neural Correlates of Corticobasal Syndrome and Corticobasal Degeneration with Systematic and Quantitative ALI Meta-Analyses. (npj Parkinson’s Disease 3, 12).
Albrecht, F., Ballarini, T., Neumann, J., Schroeter, M.L., 2018. FDG-PET hypometabolism is more sensitive than MRI atrophy in Parkinson's disease: a whole-brain multimodal imaging meta-analysis. NeuroImage Clin Epub ahead of print.
Bisenius, S., Neumann, J., Schroeter, M.L., 2016. Validating new diagnostic imaging criteria for primary progressive aphasia via anatomical likelihood estimation meta-analyses. Eur. J. Neurol. 23, 704–712.
Bisenius, S., Mueller, K., Diehl-Schmid, J., Fanzbender, K., Glimmer, T., Jessen, F., Kastubek, J., Kornhuber, J., Landwehrmeyer, B., Ludolph, A., Schneider, A., Anderl-Straub, S., Stuke, K., Danek, A., Otto, M., Schroeter, M.L., group, F.T.s., 2017. Predicting primary progressive aphasia with support vector machine approaches in structural MRI data. Neuroimage Clin. 14, 334–343.
Boxer, A.L., Geschwind, M.D., Belfor, N., et al., 2006. Patterns of brain atrophy that
differentiate cortical degeneration syndrome from progressive supranuclear palsy. Arch. Neurol. 63 (1), 81–86.

Brenneis, C., Seppi, K., Schocke, M., Benke, T., Wenning, G.K., Poeke, W., 2004. Voxel based morphometry reveals a distinct pattern of frontobasal atrophy in progressive supranuclear palsy. J. Neurol. Neurosurg. Psychiatry 75, 246–249.

Burciu, R.G., Ofori, E., Planetta, P.J., Snyder, A.F., Okun, D.S., Hass, C.J., 2017. Implementation errors in the GingerALE software: description and recommendations. Hum. Brain Mapp. 38 (1), 7–17.

Dukart, J., Muller, K., Barthel, H., Villringer, A., Sabri, O., Schroeter, M.L., Alzheimer’s Disease Neuroimaging Initiative, 2014. Meta-analysis based SVM classification enables accurate detection of Alzheimer’s disease across different clinical centers using FDG-PET and MRI. Psychiatry Res. 212, 230–236.

Eickhoff, S.B., Zilles, K., 2018. Ten simple rules for neuroimaging meta-analysis. Neuroimage. 174, 234–250.

Eickhoff, S.B., Laird, A.R., Fox, P.T., Lancaster, J.L., Fox, P.T., 2016. Implementation errors in the GingerALE software: description and recommendations. Hum. Brain Mapp. 38 (1), 7–17.
G.X., Li, H.Y., 2013. Gray matter atrophy in progressive supranuclear palsy: meta-analysis of voxel-based morphometry studies. Neurol. Sci. 34, 1049–1055.

Slowinski, J., Imamura, A., Uitti, R.J., Pooley, R.A., Strongosky, A.J., Dickson, D.W., Broderick, D.F., Wszolek, Z.K., 2008. MR imaging of brainstem atrophy in progressive supranuclear palsy. J. Neurol. 255, 37–44.

Takahashi, R., Ishii, K., Kakigi, T., Yokoyama, K., Mori, E., Murakami, T., 2011. Brain alterations and mini-mental state examination in patients with progressive supranuclear palsy: voxel-based investigations using f-fluorodeoxyglucose positron emission tomography and magnetic resonance imaging. Dement. Geriatr. Cogn. Dis. Extra 1 (1), 381–392.

Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36, 1–48.

Whitwell, J.L., Duffy, J.R., Strand, E.A., et al., 2013. Neuroimaging comparison of primary progressive apraxia of speech and progressive supranuclear palsy. Eur. J. Neurol. 20 (4), 629–637.

Albrecht, F., et al., 2019. F. Albrecht, et al. NeuroImage: Clinical 22 (2019) 101722.