

# UCSF

## UC San Francisco Previously Published Works

### Title

Uncovering spatiotemporal patterns of atrophy in progressive supranuclear palsy using unsupervised machine learning.

### Permalink

<https://escholarship.org/uc/item/3gc4111n>

### Journal

Brain communications, 5(2)

### ISSN

2632-1297

### Authors

Scotton, William J  
Shand, Cameron  
Todd, Emily  
[et al.](#)

### Publication Date

2023
















### DOI

10.1093/braincomms/fcad048

Peer reviewed

# BRAIN COMMUNICATIONS

## Uncovering spatiotemporal patterns of atrophy in progressive supranuclear palsy using unsupervised machine learning

 William J. Scotton,<sup>1</sup>  Cameron Shand,<sup>2</sup>  Emily Todd,<sup>1</sup>  Martina Bocchetta,<sup>1</sup>  David M. Cash,<sup>1</sup>  Lawren VandeVrede,<sup>3</sup>  Hilary Heuer,<sup>3</sup> PROSPECT Consortium, 4RTNI Consortium  Alexandra L. Young,<sup>4</sup>  Neil Oxtoby,<sup>2</sup>  Daniel C. Alexander,<sup>2</sup>  James B. Rowe,<sup>5</sup>  Huw R. Morris,<sup>6,7</sup>  Adam L. Boxer,<sup>3</sup>  Jonathan D. Rohrer<sup>1,†</sup> and  Peter A. Wijeratne<sup>2,†</sup>

† These authors contributed equally to this work.

To better understand the pathological and phenotypic heterogeneity of progressive supranuclear palsy and the links between the two, we applied a novel unsupervised machine learning algorithm (Subtype and Stage Inference) to the largest MRI data set to date of people with clinically diagnosed progressive supranuclear palsy (including progressive supranuclear palsy–Richardson and variant progressive supranuclear palsy syndromes).

Our cohort is comprised of 426 progressive supranuclear palsy cases, of which 367 had at least one follow-up scan, and 290 controls. Of the progressive supranuclear palsy cases, 357 were clinically diagnosed with progressive supranuclear palsy–Richardson, 52 with a progressive supranuclear palsy–cortical variant (progressive supranuclear palsy–frontal, progressive supranuclear palsy–speech/language, or progressive supranuclear palsy–corticobasal), and 17 with a progressive supranuclear palsy–subcortical variant (progressive supranuclear palsy–parkinsonism or progressive supranuclear palsy–progressive gait freezing). Subtype and Stage Inference was applied to volumetric MRI features extracted from baseline structural (T1-weighted) MRI scans and then used to subtype and stage follow-up scans. The subtypes and stages at follow-up were used to validate the longitudinal consistency of subtype and stage assignments. We further compared the clinical phenotypes of each subtype to gain insight into the relationship between progressive supranuclear palsy pathology, atrophy patterns, and clinical presentation.

The data supported two subtypes, each with a distinct progression of atrophy: a ‘subcortical’ subtype, in which early atrophy was most prominent in the brainstem, ventral diencephalon, superior cerebellar peduncles, and the dentate nucleus, and a ‘cortical’ subtype, in which there was early atrophy in the frontal lobes and the insula alongside brainstem atrophy. There was a strong association between clinical diagnosis and the Subtype and Stage Inference subtype with 82% of progressive supranuclear palsy–subcortical cases and 81% of progressive supranuclear palsy–Richardson cases assigned to the subcortical subtype and 82% of progressive supranuclear palsy–cortical cases assigned to the cortical subtype. The increasing stage was associated with worsening clinical scores, whilst the ‘subcortical’ subtype was associated with worse clinical severity scores compared to the ‘cortical subtype’ (progressive supranuclear palsy rating scale and Unified Parkinson’s Disease Rating Scale). Validation experiments showed that subtype assignment was longitudinally stable (95% of scans were assigned to the same subtype at follow-up) and individual staging was longitudinally consistent with 90% remaining at the same stage or progressing to a later stage at follow-up.

In summary, we applied Subtype and Stage Inference to structural MRI data and empirically identified two distinct subtypes of spatiotemporal atrophy in progressive supranuclear palsy. These image-based subtypes were differentially enriched for progressive supranuclear palsy clinical syndromes and showed different clinical characteristics. Being able to accurately subtype and stage progressive supranuclear palsy patients at baseline has important implications for screening patients on entry to clinical trials, as well as tracking disease progression.

Received August 22, 2022. Revised November 22, 2022. Accepted February 27, 2023. Advance access publication March 2, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of the Guarantors of Brain.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

- 1 Dementia Research Centre, Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, London WC1N 3AR, UK
- 2 Centre for Medical Image Computing, Department of Computer Science, University College London, London WC1V 6LJ, UK
- 3 Department of Neurology, Memory and Aging Center, University of California, San Francisco, CA 94158, USA
- 4 Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, UK
- 5 Cambridge University Department of Clinical Neurosciences and Cambridge University Hospitals NHS Trust, Medical Research Council Cognition and Brain Sciences Unit, Cambridge CB2 0QQ, UK
- 6 Department of Clinical and Movement Neurosciences, University College London Queen Square Institute of Neurology, London WC1N 3BG, UK
- 7 Movement Disorders Centre, University College London Queen Square Institute of Neurology, London WC1N 3BG, UK

Correspondence to: Dr William J. Scotton

UCL Institute of Neurology, Department of Neurodegeneration, Dementia Research Centre

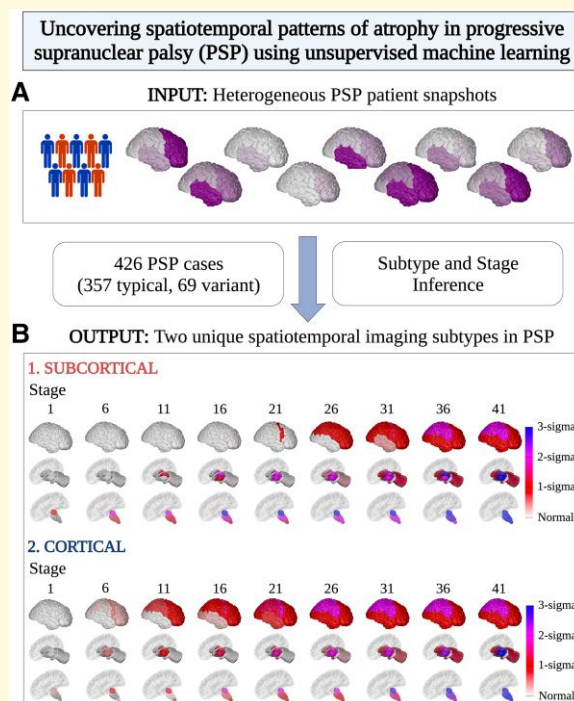
First Floor, 8-11 Queen Square, WC1N 3AR, London, UK

E-mail: w.scotton@ucl.ac.uk

**Keywords:** Subtype and Stage Inference; disease progression; progressive supranuclear palsy; biomarkers; machine learning

**Abbreviations:** AD = Alzheimer's disease; CBD = cortico-basal degeneration; CVIC = cross-validation information criteria; DAV = davunetide; DC = diencephalon; DRC = Dementia Research Centre; EBM = event-based model; FTLDNI = frontotemporal lobar degeneration neuroimaging initiative; GIF = geodesic information flow algorithm; GGT = globular glial tauopathy; GP = globus pallidus; LWENC = Leonard Wolfson Experimental Neurology Centre; MAX = multiple allocation extinction; MCMC = Markov Chain Monte Carlo; MDS = Movement Disorder Society; MMSE = Mini-Mental State Examination; MOCA = Montreal Cognitive Assessment; MS = multiple sclerosis; NA = nucleus accumbens; NINDS-SPSP = National Institute of Neurological Disorders and Stroke and Society for Progressive Supranuclear Palsy; PROSPECT = PROgressive Supranuclear Palsy CorTico-Basal Syndrome Multiple System Atrophy Longitudinal Study; PSP = progressive supranuclear palsy; PSP-CBS = PSP-corticobasal syndrome; PSP-F = PSP-frontal; PSP-P = PSP-parkinsonism; PSP-PGF = PSP-primary gait freezing; PSP-RS = PSP-Richardson; PSP-SL = PSP-speech/language syndrome; PVD = positional variance diagram; 4R = four repeat; 4RTNI1 = 4R Tauopathy Imaging Initiative Cycle 1; 4RTNI2 = 4R Tauopathy Imaging Initiative Cycle 2; ROI = region of interest; SAL = salsalate; SCP = superior cerebellar peduncles; SD = standard deviation; SEADL = Schwab and England Activities of Daily Living; SPM = statistical parametric mapping; SuStaIn = Subtype and Stage Inference; TEBM = temporal event based model; TIV = total intracranial volume; UCL = University College London; UPDRS = Unified Parkinson Disease Rating Scale; vPSP = variant PSP; YP = young plasma

## Graphical Abstract



## Introduction

Many neurodegenerative diseases are complicated by poor clinico-pathological correlation, with the underlying pathology often manifesting as a range of different, and often overlapping, clinical syndromes. Defining disease phenotypes based on common underlying biological mechanisms, as opposed to the clinical phenotype, is an important step towards enriching clinical trials with patients that are most likely to benefit from the medicine being investigated, especially as therapeutics increasingly target these biological mechanisms. Recent advances in machine learning have enabled analysis of multidimensional data to classify and stage groups with similar data-driven features (such as spatio-temporal atrophy patterns on MRI<sup>1</sup>) rather than just on common clinical features, providing new tools to tackle the problem of clinical heterogeneity.

Progressive supranuclear palsy (PSP), a neurodegenerative disease defined pathologically by the aggregation and spread of 4-repeat tau protein in neurons, astrocytes, and oligodendrocytes,<sup>2</sup> shows significant differences in severity and neuro-anatomical distribution of pathology,<sup>3</sup> resulting in a range of clinical phenotypes involving language, behaviour, and movement abnormalities.<sup>4</sup> No effective disease-modifying treatment has yet been proven for PSP, despite increasingly available clinical trials.<sup>5,6</sup> Clinical progression appears to be dependent on progressive spreading of the four-repeat (4R) tau pathology within the brain, with a recent pathology staging system<sup>3</sup> defining six sequential stages of progression for the most common clinical phenotype PSP–Richardson syndrome (PSP-RS), starting in the pallido–nigro–luisian system and spreading rostrally via the striatum and amygdala to the cerebral cortex (frontal > temporo-parietal > occipital) and caudally to the medulla oblongata, pons, and cerebellum.<sup>7</sup> Although the molecular pathogenic basis for clinical variation is still poorly understood, this study suggests that differences in tau burden and different tau cytopathologies may distinguish clinical subtypes.

The Movement Disorder Society 2017 PSP diagnostic criteria<sup>8</sup> were introduced to try and account for variant PSP clinical phenotypes (vPSP) and increase the sensitivity and specificity for diagnosis of early PSP pathology. The criteria categorize symptomatology into four clinical domains (ocular motor dysfunction, postural instability, akinesia, and cognitive dysfunction), with differing combinations of these symptoms defining a range of clinical PSP syndromes. Although the most common clinical presentation of PSP is Richardson syndrome (PSP-RS), vPSP may account for up to 50% of individuals with PSP pathology.<sup>4,9,10</sup> The vPSP syndromes include subcortical variants [including PSP–parkinsonism (PSP-P) and PSP–primary gait freezing (PSP-PGF)] and cortical variants [PSP–frontal (PSP-F), PSP–corticobasal syndrome (PSP-CBS), and PSP–speech/language syndrome (PSP-SL)]. These new criteria have higher sensitivity than the previous National Institute of Neurological Disorders and Stroke and Society for

Progressive Supranuclear Palsy criteria (87.9% versus 45.5%),<sup>11,12</sup> and the ‘suggestive of PSP’ clinical category significantly increases the sensitivity for early identification of patients with PSP pathology.<sup>13</sup> One potential issue with the Movement Disorder Society diagnostic criteria as first defined was that patients can be assigned multiple phenotypes according to clinical symptomatology.<sup>14</sup> The introduction of the multiple allocation extinction rules<sup>13</sup> helps to allocate patients to one phenotype, though its application to clinically diagnosed PSP patients appears to lead to an over-representation of PSP-RS versus vPSP syndromes such as PSP-P and PSP-PGF.<sup>15–17</sup> This has important implications for clinical trials given that the subcortical variants (PSP-P and PSP-PGF) have better survival with longer disease durations (PSP-P 9 years and PSP-PGF 13 years versus PSP-RS 6–7 years) and slower rates of disease progression.<sup>9,17–23</sup> Improved quantification of the progression of pathological brain changes across the PSP phenotypic spectrum in living patients will be essential to the success of future therapeutic trials.<sup>7,24</sup>

A probabilistic data-driven modelling approach (event-based modelling) has been used to characterise the in-vivo sequence of brain atrophy in PSP-RS,<sup>25</sup> where the order of regional atrophy broadly mirrors the sequential spread of tau pathology proposed by Kovacs *et al.*<sup>3</sup> However, the event-based modelling assumes sample homogeneity, making it unsuitable to investigate the full spectrum of PSP phenotype heterogeneity. Subtype and Stage Inference (SuStaIn), an unsupervised machine learning algorithm,<sup>1</sup> has been developed to identify data-driven disease subtypes with distinct temporal progression patterns and can do so using only cross-sectional data. The trained model can then be used to subtype and stage new individuals. This ability to disentangle both ‘phenotypic’ and ‘temporal’ heterogeneity from cross-sectional biomarkers distinguishes SuStaIn from traditional approaches that focus on either one or the other. The SuStaIn subtypes account for temporal heterogeneity, enabling more accurate subtype assignment than traditional clustering algorithms. This algorithm has been successfully applied to Alzheimer’s disease,<sup>1,26</sup> multiple sclerosis,<sup>27</sup> and genetic frontotemporal dementia,<sup>1,28</sup> providing important insights into distinct data-driven subtypes of disease progression. This type of disease progression modelling approach is ideally suited to disentangling the clinical and pathological heterogeneity of PSP.

In this study, we applied the SuStaIn algorithm to cross-sectional MRI data from a large international cohort of clinically diagnosed PSP patients (including PSP-RS and vPSP syndromes), to identify imaging subtypes with distinct sequences of atrophy progression. We validated the observed subtypes and stages using a subset of longitudinal imaging data. Finally, we characterized the clinical features of each subtype to gain insight into the relationship between PSP pathology, atrophy patterns, and clinical presentation.

In this paper, we use ‘syndrome’ when referring to the PSP clinical syndrome/phenotype (as defined in the Movement

Disorder Society 2017 diagnostic criteria<sup>8</sup>) and ‘subtype’ when referring to MRI-based subtypes identified by SuStaIn.

## Materials and methods

### Participants and clinical data collected

Clinical and MRI data from individuals with a clinical diagnosis of ‘possible’ or ‘probable’ PSP (including PSP-RS and vPSP syndromes) as per the Movement Disorder Society 2017 PSP diagnostic criteria<sup>8</sup> were collected from seven main sources; the 4R Tauopathy Imaging Initiative Cycle 1 (4RTNI1; ClinicalTrials.gov: NCT01804452),<sup>29,30</sup> the 4R Tauopathy Imaging Initiative Cycle 2 (4RTNI2; ClinicalTrials.gov: NCT02966145), the davunetide randomized control trial (ClinicalTrials.gov: NCT01056965),<sup>31</sup> the salsalate clinical trial (ClinicalTrials.gov: NCT02422485),<sup>32</sup> the young plasma clinical trial (ClinicalTrials.gov: NCT02460731),<sup>32</sup> the PROgressive Supranuclear Palsy CorTico-Basal Syndrome Multiple System Atrophy Longitudinal Study (PROSPECT; ClinicalTrials.gov: NCT02778607),<sup>33</sup> and the University College London Dementia Research Centre FTD cohort. Controls for z-scoring imaging data were collected from three sources with available cross-sectional volumetric MRI: the Frontotemporal Lobar Degeneration Neuroimaging Initiative data set (FTLDNI; <http://4rtni-ftldni.ini.usc.edu/>), PROSPECT, and the University College London Dementia Research Centre FTD cohort. Controls were defined as no known diagnosis of a neurological or neurodegenerative condition and no known history of memory complaints. Further details on the 4RTNI2 cohort are included in the [Supplementary material](#), and details of demographics and clinical information by cohort are provided in [Supplementary Table 1](#). Information pertaining to recruitment, diagnostic criteria, and MRI scanner acquisition protocols for the other cohorts has been provided previously.<sup>25</sup> Appropriate ethics was applied for and approved via each of the individual trial and research ethics committees. To be included in this study, all participants needed to have, as a minimum, a clinical diagnosis of PSP (PSP-RS or vPSP), a baseline T<sub>1</sub> volumetric MRI on a 1.5 or 3 tesla scanner, and basic demographic data (gender and age at time of scan). Clinical rating scale scores [PSP rating scale, Unified Parkinson Disease Rating Scale (UPDRS), Schwab and England Activities of Daily Living scale (SEADL), and Montreal Cognitive Assessment (MoCA) or Mini-mental State Examination (MMSE) at baseline and follow-up], pathology at autopsy, and follow-up scans were also included if available. As detailed in a previous work,<sup>25</sup> original trial analyses failed to show any treatment effect (including no change in volumetric MRI measurements) in the salsalate, young plasma,<sup>32</sup> and davunetide<sup>31</sup> trials, so data were combined from each study’s treatment and placebo arms. Longitudinal data were used to validate

the consistency of SuStaIn’s subtype and stage assignments at follow-up.

Given that the PROSPECT and 4RTNI2 trials only assessed cognitive function using the MOCA (as opposed to the MMSE for the other trials), raw MOCA scores were converted to MMSE scores using the validated method used by Lawton *et al.*<sup>34</sup> For missing data in clinical scales, an adjusted mean score was used if at least 80% of the assessment was complete.<sup>33</sup>

### MRI acquisition and image processing

The acquisition and postprocessing procedures have been described previously.<sup>25</sup> Briefly, cortical and subcortical structures were parcellated using the geodesic information flow algorithm (GIF),<sup>35</sup> which automatically extracts regions based on the Neuromorphometrics atlas (Neuromorphometrics Inc.), using an atlas propagation and label fusion strategy.<sup>36,37</sup> Subregions of the cerebellum were extracted using GIF based on the Diedrichsen atlas.<sup>38</sup> The medulla, pons, superior cerebellar peduncles (SCP), and midbrain were subsequently segmented using a customised version of a module available in FreeSurfer to accept the GIF parcellation of the whole brainstem as input.<sup>39</sup> Volumes for 24 grey matter regions were calculated; four brainstem (medulla, pons, SCP, and midbrain), three cerebellar (cerebellar cortex, dentate nucleus, and vermis), nine subcortical (thalamus, globus pallidus, caudate, putamen, ventral diencephalon, thalamus, hippocampus, amygdala, and nucleus accumbens), and eight cortical (basal forebrain, cingulate, frontal anterior, frontal posterior, insula, temporal, parietal, and occipital) regions. Regions that had a right and left label were combined. A list of the GIF subregions included in each cortical region is included in [Supplementary Table 2](#). The total intracranial volume was calculated using SPM12 v6225 (Statistical Parametric Mapping, Wellcome Trust Centre for Neuroimaging, London, UK) running under MATLAB R2012b (Math Works, Natick, MA, USA) (Scotton 2022; 40). All segmentations were visually inspected to ensure accurate segmentation.

All volumes were corrected for total intracranial volume, scanner field strength (1.5 T or 3 T), scanner manufacturer, age at baseline scan, and sex, using linear regression on the control population and then propagating this model to the PSP population, to generate covariate-adjusted regional volumes. To confirm that age effects on regional brain volumes had been successfully regressed out, linear models were fit to assess for any residual association between individual covariate-adjusted regional volumes and age at scan.

### Biomarker selection and data preparation

A biomarker in this study was defined as an image-based regional volume extracted using GIF. To reduce dimensionality, we carried out pairwise comparisons between healthy volunteers and patients at baseline visit and selected MRI regions whose differences between groups were associated

with a moderate to large effect size (Cohen's  $d$  effect size of  $\geq 0.6$  for standardized mean differences between the cases and controls). Applying this threshold to the regional volumes segmented by GIF (see 'MRI acquisition and image processing') resulted in selection of 20 regions of interest (ROI) that were then included for downstream analysis (Supplementary Table 3). Adjusted regional volumes for the ROI were converted into  $z$ -scores relative to the control group by subtracting the mean of the control group from each patient's ROI volume and dividing by the standard deviation of the control group. Given that regional brain volumes decrease with disease progression, the  $z$ -scores become negative as the disease progresses;  $z$ -scores were therefore multiplied by  $-1$ , to give positive  $z$ -scores that increase with disease progression. The  $z$ -scored data was then used as input to SuStaIn.

## Subtype and Stage Inference

SuStaIn is a probabilistic machine learning algorithm that simultaneously clusters individuals into groups (subtypes) and infers a trajectory of change associated with each group; that trajectory defines the disease stage (degree of disease progression within a subtype) of each individual within the corresponding group. SuStaIn requires only cross-sectional data as input, although it can exploit longitudinal data for training if available. The details of the algorithm have been published previously<sup>1</sup> and applied to a range of different diseases.<sup>26-28,40</sup> Each subtypes' progression pattern is described using a piecewise linear  $z$ -score model, expressing a trajectory with a series of stages that each correspond to a single biomarker (regional brain volume in this case) reaching a new  $z$ -score. Importantly, the number of SuStaIn stages is determined by the number of biomarkers (the product of the number of ROIs and number of  $z$ -score thresholds per ROI) provided as input. SuStaIn optimizes both the subtype membership and the ordering in which different biomarkers reach different  $z$ -scores in each subtype (for example one, two or three standard deviations away from the control mean for that ROI) using a data likelihood function.

Only clinically diagnosed PSP cases were used to fit SuStaIn. Supplementary Table 4 provides a summary of the  $z$ -score settings, Markov Chain Monte Carlo iterations and number of random starting sequences used for the SuStaIn algorithm. Model uncertainty was estimated using 100 000 Markov Chain Monte Carlo iterations, and in the single-cluster expectation maximization procedure, the single-cluster sequence was optimized from 24 different random starting sequences to find the maximum likelihood solution. The optimal number of subtypes was determined using information criteria calculated through ten-fold cross-validation (cross-validation information criteria), to balance internal model accuracy with model complexity.<sup>1</sup> Where there was no strong evidence for an additional subtype in the model using the cross-validation information criteria, we assessed the average log-likelihood across folds for the additional subtype and, if there was no improvement, selected the

most parsimonious model, i.e. the model with fewer subtypes.

Finally, the fitted SuStaIn model was used to calculate the probability that each individual falls at each stage of each subtype, and individuals were assigned to their maximum likelihood stage and subtype based on their baseline scan (as described in Young *et al.*<sup>1</sup>). The subtype progression patterns identified by SuStaIn were visualized using BrainPainter software,<sup>41</sup> modified to include the brainstem segmentations.

## Assigning individuals to subtypes and stages

Individuals' stage was computed based on their average stage, weighted by the probability of belonging to each stage of each subtype. Individuals that were assigned to either SuStaIn stage 0 (i.e. no atrophy on imaging compared to controls) or stage 41 (end stage, i.e. all ROI maximum atrophy) were labelled 'no subtype'. All other individuals were labelled as 'subtypable' and were then assigned to their most probable subtype.

## Statistical analyses

For all analyses, PSP cases were stratified into PSP-RS, PSP-subcortical, and PSP-cortical groups based on their baseline clinical diagnosis. The PSP-subcortical group includes individuals with PSP-P and PSP-PGF; the PSP-cortical group includes cases with PSP-SL, PSP-F, and PSP-CBS.

### Clinical phenotype and baseline characteristics

Pairwise comparisons of baseline characteristics were performed between all PSP cases and controls, PSP syndrome (PSP-RS, PSP-cortical, and PSP-subcortical) versus all PSP cases, and each PSP syndrome against each other, using  $t$ -tests for continuous variables and  $\chi^2$  tests for categorical variables. Statistical significance was reported at a level of  $P < 0.05$  and at the Bonferroni corrected level of  $P < 0.001$  to correct for multiple comparisons (44 items).

### Association between subtype assignment and covariates

To assess for any residual association between covariates (total intracranial volume, scanner field strength, scanner manufacturer, age at baseline scan, sex, cohort, and SuStaIn stage) and the SuStaIn subtype, a logistic regression model was fit to the data using the 'lm()' function from the 'stats' package (version 3.6.2).

### Subtype characterisation

Overall differences between subtypes were first assessed independently of stage, with individuals classified as 'no subtype' (i.e. stage 0 or 41) excluded from analysis. To compare whether there were any differences between subtypes, we performed  $t$ -tests for continuous variables and  $\chi^2$  tests for categorical variables (and *post hoc* pairwise

comparisons for clinical syndrome versus the SuStaIn subtype using the ‘chisq.multicomp()’ function from the ‘RVAideMemoire’ R package version 0.9-81-2).

To test for associations between clinical scores (PSP rating scale, UPDRS, SEADL, and MMSE) and subtype, we accounted for the SuStaIn stage, age, and sex by fitting a linear model (clinical test score~subtype + stage + age + sex) for each clinical test score. Statistical significance was reported uncorrected at a level of  $P < 0.05$  and at the Bonferroni corrected level of  $P < 0.005$  for demographic variables (11 items) and for clinical scores (10 items), to account for multiple comparisons.

To assess the average stage by clinical syndrome by the SuStaIn subtype, a one-way ANOVA was performed (mean stage~PSP syndrome + SuStaIn baseline subtype) with the ‘aov()’ function the ‘stats’ package (version 3.6.2). Tukey *post hoc* significant differences were then calculated to identify the level of significance.

All statistical analyses were performed either in R (version 4.0.5) or Python (version 3.7.6).

## Longitudinal validation

The SuStaIn model fitted on the cross-sectional baseline data was used to assign the maximum likelihood stage and subtype to all follow-up scans (at all time points). These scans were used to validate the stability of subtypes and to assess stage progression, based on the hypothesis that individuals should remain in the same subtype but should advance to higher stages over time (or at least remain at the same stage). Subtype stability was defined as the proportion of individuals assigned to the same subtype at follow-up(s) or progressed from stage 0 into a subtype, compared to the total number of individuals. Stage progression was assessed by comparing the SuStaIn stage at baseline and follow-up(s) for all individuals. Specifically, we calculated the proportion of individuals that advanced or stayed at the same stage at follow-up scan.

## Results

### Participants

**Table 1** summarizes the key baseline clinical features for individuals included in the study. For a breakdown of this data by each contributing cohort, please refer to [Supplementary Table 1](#). We collected a large imaging cohort of PSP cases; 1083 MRI images were included (after quality control) from a total of 716 individuals: 426 with a clinical diagnosis of PSP (with 367 follow-up scans) and 290 controls. Of the PSP cases, 357 (84%) were diagnosed with PSP-RS, 52 (12%) with a PSP-cortical syndrome (PSP-SL, PSP-F, or PSP-CBS), and 17 (4%) with a PSP-subcortical syndrome (PSP-P or PSP-PGF). After coming to post-mortem, 31 (7%) of the PSP cases had a pathological diagnosis, of which 29 (94%) showed 4R tau pathology consistent with PSP, whereas two cases that presented with PSP-RS had

non-PSP tau pathology (one corticobasal degeneration and one globular glial tauopathy). Given that the focus of this study was understanding the clinical heterogeneity of clinically diagnosed PSP, both cases were included in the analysis.

Overall, the PSP cases at baseline had an older average age compared to controls {68.5 years [standard deviation (SD)  $\pm$  6.8] versus 62.5 years (SD 62.5),  $P < 0.001$ }, though they were matched for gender. We confirmed that despite the difference in age between cases and controls, age effects had been effectively regressed out of the regional covariate-adjusted volumes for both groups: cases ([Supplementary Fig. 1](#) and [Supplementary Table 5](#)) and controls ([Supplementary Fig. 2](#) and [Supplementary Table 6](#)). There were significant differences in baseline clinical scores between the different clinical PSP phenotypes. The highest PSP rating scale score (measure of motor predominant disease burden) was seen in PSP-RS [37.2 (SD  $\pm$  13.2)], followed by PSP subcortical syndromes [26.3 (SD  $\pm$  10.7)], with PSP cortical syndromes being the least impaired [20.6 (SD  $\pm$  20)]. There was, however, a large variation in this score for the cortical syndromes [PSP-SL 11.9 (SD  $\pm$  10.8), PSP-F 19.5 (SD  $\pm$  13.2), and PSP-CBS 56.4 (SD  $\pm$  11.9)]. In keeping with the increased motor predominant disease burden (higher PSP rating scale score) in the PSP-RS and PSP-subcortical cases, the UPDRS was significantly higher in these cases versus PSP-cortical cases [PSP-subcortical 36.2 (SD  $\pm$  12.9), PSP-RS 29.3 (SD  $\pm$  14), PSP-cortical 18.8 (SD  $\pm$  7.6),  $P < 0.05$  for each comparison]. The PSP-subcortical cases had a better MMSE score on average compared to the PSP-RS and PSP-cortical syndromes [22.4 (SD  $\pm$  1.2) versus 25.7 (SD  $\pm$  3.7) and 24.7 (SD  $\pm$  4.6),  $P < 0.001$  for each comparison]. There was no significant difference in MMSE between the latter two syndromes.

### Spatiotemporal subtypes of PSP

SuStaIn was fit using PSP cases only, based on the rationale that PSP is a rare disease, and it is very unlikely for our cohort of controls to have asymptomatic PSP. Indeed, it is more likely that the controls would have a more common neurodegenerative disorder such as Alzheimer’s disease rather than PSP, and we did not want this to confound the Subtype and Stage Inference estimation hence the exclusion.

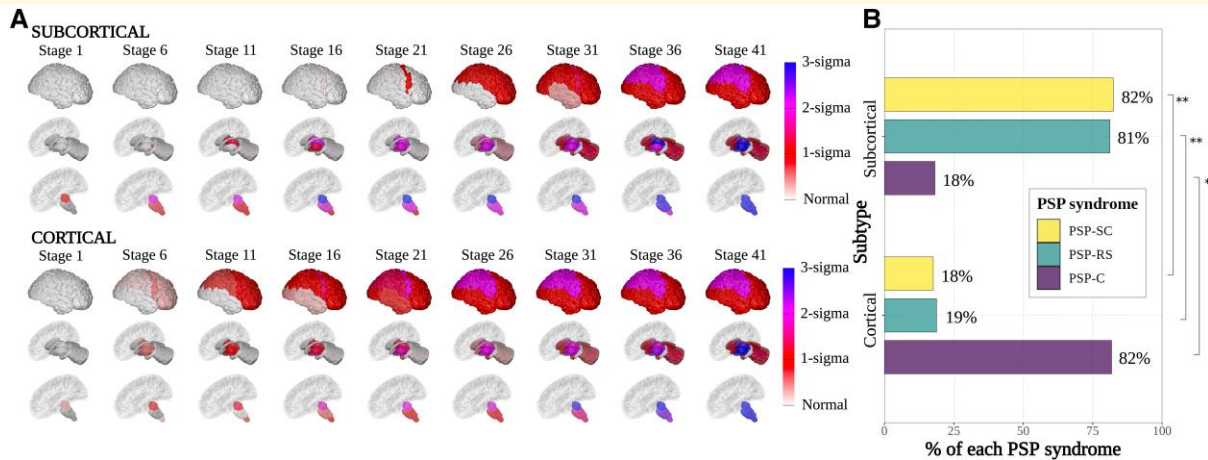
SuStaIn identified two imaging subtypes with distinct patterns of regional atrophy evolution ([Fig. 1A](#) and [Supplementary Fig. 3](#) for the positional variance diagrams). [Supplementary Fig. 4](#) shows the log-likelihoods after 10-fold cross validation with the associated cross-validation information criteria demonstrating that the two-subtype model was the most parsimonious. Based on the earliest MRI abnormalities seen in the SuStaIn-defined trajectories, we labelled the first the ‘subcortical’ subtype and the second the ‘cortical’ subtype. The ‘subcortical’ subtype (75% of the cases) has atrophy in the midbrain followed by the other brainstem structures (medulla, pons, and SCP) and the ventral diencephalon at early SuStaIn stages. The atrophy then progresses caudally to the dentate nucleus of the cerebellum

**Table 1 Clinical phenotypes and baseline characteristics**

	Study group		PSP by subgroup							
			PSP-cortical			PSP-subcortical				
	Controls	All	PSP-RS	PSP-SL	PSP-F	PSP-CBS	All cortical	PSP-P	PSP-PGF	All subcortical
Baseline, n (fu visits)	290	426 (367)	357 (329)	35 (23)	10 (2)	7 (0)	52 (25)	7 (8)	10 (5)	17 (13)
Sex, % female	57%	49%	49%	46%	40%	71%	48%	14%	60%	41%
Age at first scan, y	62.5 (9.4)	68.5 (6.8) <sup>d</sup>	68.1 (6.6)	68.6 (7.7)	68.8 (8.6)	77.6 (2.7)	69.9 (7.9)	71.5 (7.7)	71.6 (7.1)	71.5 (7.1)
Age at first symptom, y <sup>a</sup>		64.1 (7.5)	63.5 (7.2)	64.1 (8.1)	64.6 (7.7)	72.7 (3.8)	65.4 (8.0)	64.5 (10.3)	66.6 (5.4)	65.7 (7.6)
Disease duration, y <sup>a,b</sup>		4.7 (3.2)	4.6 (3.3)	4.5 (2.8)	4.19 (3.0)	4.9 (2.0)	4.5 (2.7)	6.9 (3.5)	5.1 (2.9)	5.8 (3.2)
Pathology, % PSP		31 (94%) <sup>c</sup>	26 (92%)	-	1 (100%)	1 (100%)	2 (100%)	4 (100%)	1 (100%)	5 (100%)
PSP rating scale score		34.9 (15.0)	37.2 (13.2) <sup>a,h,i</sup>	11.9 (10.8)	19.5 (13.2)	56.4 (11.9)	20.6 (20.0) <sup>f,h</sup>	33.8 (7.89)	23 (10.4)	26.3 (10.7) <sup>e,i</sup>
History		7.8 (3.9)	8.3 (3.6) <sup>h,i</sup>	3.4 (3.3)	5 (2.8)	10.7 (3.4)	4.8 (4.2) <sup>h</sup>	8.5 (3.9)	4.4 (3.1)	5.7 (3.7) <sup>i</sup>
Mentation		3.4 (2.6)	3.5 (2.3) <sup>i</sup>	1.8 (1.7)	5.2 (4.2)	4.8 (2.6)	2.8 (2.2)	4 (2.6)	1.0 (0.9)	1.9 (1.3) <sup>e,i</sup>
Bulbar		2.5 (1.8)	2.7 (1.8) <sup>h</sup>	1.8 (1.7)	1.2 (1.4)	2.3 (1.9)	1.8 (1.7) <sup>e,h</sup>	2.7 (2.5)	1.6 (1.1)	1.9 (1.7)
Ocular motor		7.6 (4.0)	8.4 (3.5) <sup>a,h,i,j</sup>	1.3 (0.6)	2.4 (1.9)	10.4 (2.4)	2.9 (2.2) <sup>f,h,j</sup>	5.8 (3.3)	4.6 (3.5)	4.9 (3.4) <sup>e,h,j</sup>
Limb motor		4.3 (2.9)	4.5 (2.7) <sup>i</sup>	2.4 (2.2)	1.2 (1.0)	10.9 (2.9)	3.5 (3.2)	3.5 (1.3)	2.9 (1.4)	3.1 (1.3) <sup>e,i</sup>
Gait and midline		9.2 (5.2)	9.9 (4.7) <sup>h</sup>	2.2 (2.1)	3.6 (3.2)	17.3 (3.0)	4.8 (3.5) <sup>f,h,k</sup>	9.5 (2.1)	8.6 (3.5)	8.9 (3.1) <sup>k</sup>
SEADL		54.8 (24.8)	53.8 (24.0) <sup>i</sup>	66.2 (33.1)	51.4 (13.5)	26.6 (13.5)	58.5 (30.2)	57.5 (20.6)	68.9 (14.5)	65.4 (16.6) <sup>e,i</sup>
UPDRS		27.3 (18.0)	29.3 (14.7) <sup>g</sup>	9.3 (8.9)	11.5 (10.9)	65.6 (12.2)	18.8 (17.6) <sup>g,k</sup>	35.5 (9.3)	36.4 (14.7)	36.2 (12.9) <sup>e,k</sup>
MIMSE		25.5 (3.8)	25.7 (3.7) <sup>i</sup>	27 (2.6)	21.9 (5.3)	19.4 (4.0)	24.7 (4.6) <sup>l</sup>	22 (0.8)	22.6 (1.3)	22.4 (1.2) <sup>f,i,l</sup>

Values are mean (SD), apart from gender, % female; baseline n (n follow-up visits); and pathology n (% PSP). Pairwise comparisons between groups were performed using t-tests for continuous variables and  $\chi^2$  tests for categorical variables. CBD, cortico-basal degeneration; GGT, globular glial tauopathy; MMSE, Mini-Mental State Examination; PSP-C, PSP-cortical (includes PSP-frontal, PSP-speech/language disorder, and PSP-corticobasal syndrome); PSP-RS, PSP-Richardson syndrome; PSP-SC, PSP-subcortical (includes PSP-parkinsonism and PSP-progressive gait freezing); SEADL, Schwab and England Activities of Daily Living; UPDRS, Unified Parkinson's Disease Rating Scale. <sup>a</sup>Note incomplete data for disease duration/age at first symptom. <sup>b</sup>Time from first symptom to first scan. <sup>c</sup>Two cases of not PSP pathology (1 CBD, 1 GGT). <sup>d</sup>PSP all versus controls. Statistically significant at  $P < 0.05$ , corrected for multiple comparisons (44 comparisons,  $P < 0.00114$ ). <sup>e</sup>PSP (subgroup) versus PSP-all. Statistically significant at  $P < 0.05$ , uncorrected for multiple comparisons. <sup>f</sup>PSP (subgroup) versus PSP-all. Statistically significant at  $P < 0.05$ , corrected for multiple comparisons (44 comparisons,  $P < 0.00114$ ). <sup>g</sup>PSP-RS versus PSP-C. Statistically significant at  $P < 0.05$ , uncorrected for multiple comparisons. <sup>h</sup>PSP-RS versus PSP-C. Statistically significant at  $P < 0.05$ , corrected for multiple comparisons (44 comparisons,  $P < 0.00114$ ). <sup>i</sup>PSP-RS versus PSP-SC. Statistically significant at  $P < 0.05$ , uncorrected for multiple comparisons. <sup>j</sup>PSP-RS versus PSP-SC. Statistically significant at  $P < 0.05$ , corrected for multiple comparisons (44 comparisons,  $P < 0.00114$ ). <sup>k</sup>PSP-RS versus PSP-SC. Statistically significant at  $P < 0.05$ , uncorrected for multiple comparisons. <sup>l</sup>PSP-C versus PSP-SC. Statistically significant at  $P < 0.05$ , corrected for multiple comparisons (44 comparisons,  $P < 0.00114$ ).





**Figure 1 Subtype progression patterns of PSP atrophy identified by Subtype and Stage Inference (SuStaln). (A)** Spatial distribution and severity of atrophy at each SuStaln stage by subtype. Each row (subcortical top, cortical bottom) represents a subtype progression pattern identified by SuStaln consisting of a set of stages at which brain volumes in PSP cases reach different z-scores relative to controls. **(B)** Assignment of PSP clinical syndromes to each SuStaln subtype. Size of bar (x-axis) represents percentage of cases labelled with that PSP syndrome assigned to that SuStaln subtype (y-axis). A Pearson chi-square test was performed with *post hoc* pairwise comparisons for clinical syndrome versus the SuStaln subtype using the `chisq.multcomp()` function from the `RVAideMemoire` R package version 0.9-81-2 [ $\chi^2(2, N = 406) = 81.8, P = 2.2 \times 10^{-16}$ ]. \*Statistically significant at  $P < 0.05$ , uncorrected for multiple comparisons. \*\*Statistically significant at  $P < 0.05$ , corrected for multiple comparisons. Visualizations in **A** were generated using the BrainPainter software,<sup>41</sup> modified to include the brainstem segmentations.

**Table 2 Comparison of demographics, clinical diagnosis, and test scores between subtypes**

	Subcortical subtype	Cortical subtype	P-value
All scans, <i>n</i>	321 (75.4)	105 (24.6)	
Subtypable scans, <i>n</i>	302 (74.4)	104 (25.6)	0.07 <sup>a</sup>
Average subtype probability <sup>b</sup>	0.94 (0.1)	0.85 (0.2)	<0.005 <sup>c</sup>
Sex, % female	49%	48%	0.82
Age at first scan, <i>y</i>	68.1 (6.3)	70.0 (8.1)	0.02
Age at first symptom, <i>y</i> <sup>d</sup>	63.8 (6.9)	65.5 (8.3)	0.14
Disease duration, <i>y</i> <sup>d,e</sup>	4.4 (3.1)	5.0 (3.2)	0.26
PSP syndrome, <i>n</i>			<0.005 <sup>c</sup>
PSP-RS	280 (81%)	65 (19%)	–
PSP-C	8 (18%)	36 (82%)	–
PSP-SC	14 (82%)	3 (18%)	<0.005 <sup>c</sup>
PSP rating scale	37.0 (13.6)	30.3 (16.9)	<0.005 <sup>c</sup>
SEADL	53.7 (23.8)	56.3 (26.6)	0.39
UPDRS	30.0 (16.2)	22.8 (21.0)	<0.005 <sup>c</sup>
MMSE	25.5 (3.6)	24.9 (4.2)	0.20

Values are mean (SD) or *n* (%), apart from sex = % female. Pairwise comparisons between groups were performed using *t* tests for continuous variables and  $\chi^2$  tests for categorical variables. PSP-C, PSP-cortical (includes PSP-frontal, PSP-predominant speech/language disorder, and PSP-predominant corticobasal syndrome); PSP-RS, PSP-Richardson syndrome; PSP-SC, PSP-subcortical (includes PSP-parkinsonism and PSP-progressive gait freezing); SEADL, Schwab and England Activities of Daily Living; UPDRS, Unified Parkinson's Disease Rating Scale; MMSE, Mini-Mental State Examination. <sup>a</sup>All scans versus subtypable scans. <sup>b</sup>Subtype probability = the probability of assignment for an individual case to the given subtype. <sup>c</sup>Statistically significant at  $P < 0.05$ , corrected for multiple comparisons (11 comparisons,  $P$ -value < 0.005). <sup>d</sup>Note incomplete data for disease duration/age at first symptom. <sup>e</sup>Time from first symptom to first scan.

and rostrally to the thalamus and lentiform nucleus (globus pallidus and putamen) before spreading to the cortex (after stage 13). Cortical atrophy progresses in an anterior to

posterior direction, beginning in the insula and posterior frontal lobe, before spreading to the temporal, parietal, and finally the occipital lobe. The 'cortical' subtype (25% of cases) has more generalized atrophy in the early SuStaln stages, with the midbrain and insula affected first and then the frontal lobes (posterior > anterior), thalamus, ventral diencephalon, and the basal ganglia all affected at a similar time (before stage 13). Interestingly, the end stage atrophy pattern is similar for both subtypes.

Overall, 20 of the 426 scans (5%) were not subtypable at baseline and so were excluded from subtype *post hoc* analyses. Three of these individuals had a clinical diagnosis of PSP-RS and were at stage 41 and 17 were at stage 0 [9 PSP-RS and 8 PSP-cortical (all PSP-SL)].

A logistic regression model was fit to assess for any residual association between the SuStaln subtype, regressed covariates, and SuStaln stage (SuStaln subtype~SuStaln stage + TIV + age at first scan + sex + scanner + scanner field strength + scanner manufacturer). This demonstrated a remaining association between the SuStaln subtype and age at first scan ( $z = 2.8, P = 6 \times 10^{-3}$ ), General Electric 3T scanner ( $z = -3.0, P = 3 \times 10^{-3}$ ), Phillips 3T scanner ( $z = -2.5, P = 0.01$ ), and the 4RTN12 cohort ( $z = 3.6, P = 3 \times 10^{-4}$ ). There was no dependency of the subtype on stage ( $z = -0.1, P = 0.91$ ) with a similar distribution of stages across each subtype (Supplementary Fig. 5A).

## Association between PSP clinical syndromes and subtype

We compared the subtype assignments ('subcortical' versus 'cortical') for clinical PSP syndromes (PSP-RS versus PSP-

cortical and PSP–subcortical). Figure 1B and Table 2 show the percentage of each of these clinical syndromes assigned to each subtype. The ‘subcortical’ SuStaIn subtype was significantly enriched for PSP-RS and PSP–subcortical syndromes; 81% of PSP-RS cases ( $P = 2 \times 10^{-6}$ ) and 82% of PSP–subcortical cases ( $P = 0.007$ ) were assigned to the SuStaIn ‘subcortical’ subtype. The ‘cortical’ SuStaIn subtype was enriched for the PSP–cortical syndromes; 81% of PSP–cortical syndromes ( $P = 2 \times 10^{-5}$ ) were assigned to the SuStaIn ‘cortical’ subtype.

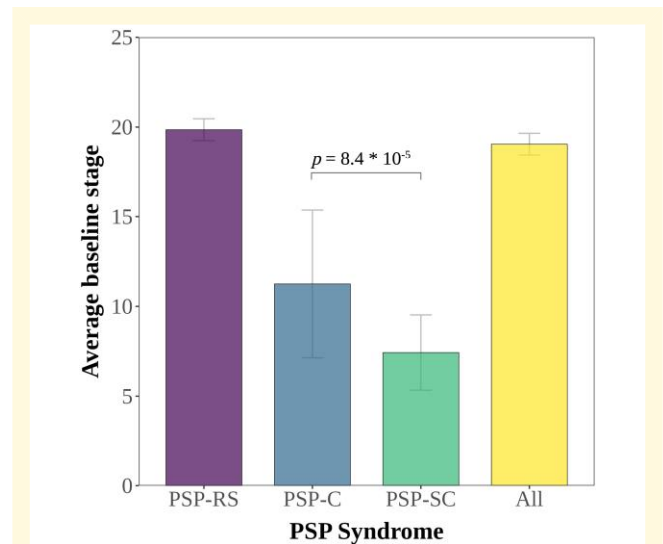
## Subtype demographics and clinical characteristics

Table 2 gives an overview of demographics, clinical diagnosis, and test scores by subtype. With an average subtype probability assignment of 0.94 compared to 0.85 for the ‘cortical’ subtype ( $t = -6.5$ ,  $P < 0.005$ ), 74% of the subtypable scans were assigned to the ‘subcortical’ subtype. Those in the ‘subcortical’ subtype were both slightly younger at symptom onset [63.8 (SD  $\pm$  6.9) years versus 65.5.0 (SD  $\pm$  8.3)] and at time of baseline scan [68.1 (SD  $\pm$  6.3) years versus 70.0 (SD  $\pm$  8.1)], though this did not reach statistical significance in either case. Pairwise comparisons of clinical scores demonstrated that PSP rating scale scores [37.0 (SD  $\pm$  13.6) versus 30.3 (SD  $\pm$  16.9),  $t = -3.7$ ,  $P < 0.005$ ] and UPDRS [30.0 (SD  $\pm$  16.2) versus 22.8 (SD  $\pm$  21.0)] were higher (i.e. more severe motor predominant disease burden) in the ‘subcortical’ subtype. Average MMSE was similar between subtypes [‘subcortical’ subtype 25.5 (SD  $\pm$  3.6) versus ‘cortical’ subtype 24.9 (SD  $\pm$  4.2),  $t = -1.3$ ,  $P < 0.20$ ].

The average stage for subtypable individuals within each subtype was similar [19.0 (SD  $\pm$  10.5) for ‘subcortical’ versus 18.3 (SD  $\pm$  9.1) for ‘cortical’,  $\beta = 8 \times 10^{-6}$ ,  $P = 0.85$ ] (Supplementary Table 7). However, PSP–subcortical cases (82%) assigned to the ‘subcortical’ SuStaIn subtype were on average at a lower stage [7.4 (SD  $\pm$  5.8)] compared to PSP-RS cases assigned to either the ‘subcortical’ [19.9 (SD  $\pm$  10.2),  $P < 0.003$ ] (Fig. 2) or the ‘cortical’ subtype [18.9 (SD  $\pm$  8.7),  $P < 0.003$ ]. We then tested whether the PSP–subcortical and PSP-RS cases assigned to the ‘subcortical’ subtype showed differences in the rate of progression (defined as change in subtype per year). PSP–subcortical cases in the ‘subcortical’ subtype progressed on average 0.66 stages per year, compared to 1.86 stages per year for the PSP-RS cases ( $t = 2.49$ , 95% CI 0.1–2.4,  $P = 0.046$ ). One PSP–subcortical case progressed from stage 12 (no cortical involvement) to stage 14 (insula and posterior frontal lobe abnormal), whilst two cases had more extensive cortical involvement at baseline (stage 16 and stage 26, respectively, at baseline, and stage 16 and stage 27, respectively, at follow-up).

## Association between stage, subtype, and disease severity

Table 3 shows the relationship between clinical test scores and the SuStaIn subtype and stage across all subtypable



**Figure 2** Average stage of PSP clinical syndromes assigned to the ‘subcortical’ SuStaIn subtype. Mean SuStaIn stage for PSP syndrome, with associated standard error bars. ANOVA: PSP syndrome = 12.1,  $P = 8 \times 10^{-6}$ ; SuStaIn subtype =  $1 \times 10^{-4}$ ,  $P = 0.97$ . Tukey *post hoc* honest significance test: in ‘subcortical’ SuStaIn subtype PSP-RS versus PSP-SC estimate =  $-12.4$ ,  $P = 8.4 \times 10^{-5}$ . No other *post hoc* comparisons significant.

cases, accounting for age at first scan and sex. The performance on the total PSP rating scale score (and history, bulbar, ocular motor, and gait midline subscores) was worse in the ‘Subcortical’ subtype and related to the stage, suggesting that these scores decline with disease progression in both subtypes, but the overall scores are worse in the ‘subcortical’ subtype. The UPDRS score was worse in the ‘subcortical’ subtype but unrelated to the SuStaIn stage, suggesting that test performance shows a stronger decline with disease progression in the ‘subcortical’ subtype. Worsening SEADL, PSP rating scale limb motor, and mentation subscores were associated with increasing SuStaIn stage, suggesting that these scores decline with disease progression in both subtypes. All these associations survived Bonferroni correction for multiple comparisons. MMSE showed no difference between subtypes and was not associated with the SuStaIn stage.

## Longitudinal consistency of subtypes

Given that we used cross-sectional MRI data to infer distinct longitudinal trajectories, evaluating how well longitudinal data fits the SuStaIn model is a key aspect of validation. We tested this in two ways: firstly, by assessing whether subtype assignments were longitudinally stable and, secondly, by testing whether individuals progressed to later stages at follow-up. A total of 355 follow-up scans (355/367) were subtypable at follow-up from 289 PSP cases (224 had one follow-up scan, 64 had two, and one individual had three). Mean follow-up time was 0.91 years with a SD of 0.38 years.

Overall, the SuStaIn subtype assignments showed good stability at follow-up (Table 4), with 95% (347 out of 367

**Table 3 Comparison of adjusted clinical scores between subtypes**

	SuStaln subtype		SuStaln stage		Subtype with worse score	Change with SuStaln stage
	t-value	P-value	t-value	P-value		
PSP rating scale score						
Total	-4.12	$5 \times 10^{-5b}$	5.21	$3 \times 10^{-7b}$	Subcortical subtype	Worsens
History	-3.98	$8 \times 10^{-5b}$	4.21	$3 \times 10^{-5b}$	Subcortical subtype	Worsens
Mentation	-0.86	0.39	3.74	$2 \times 10^{-4b}$		Worsens
Bulbar	-2.17	0.03 <sup>a</sup>	3.42	$7 \times 10^{-4b}$	Subcortical subtype	Worsens
Ocular motor	-4.56	$7 \times 10^{-6b}$	3.90	$1 \times 10^{-4b}$	Subcortical subtype	Worsens
Limb motor	-0.40	0.69	3.25	0.001 <sup>b</sup>		Worsens
Gait and midline	-4.02	$7 \times 10^{-5b}$	3.12	0.002 <sup>b</sup>	Subcortical subtype	Worsens
SEADL	1.03	0.30	-5.57	$5 \times 10^{-8b}$		Worsens
UPDRS	-2.67	0.009 <sup>b</sup>	1.70	0.08	Subcortical subtype	
MMSE	-1.03	0.31	-1.42	0.16		

Linear model of clinical score ~ subtype + stage + age + sex. SEADL, Schwab and England Activities of Daily Living; UPDRS, Unified Parkinson's Disease Rating Scale; MMSE, Mini-Mental State Examination. <sup>a</sup>Statistically significant at  $P < 0.05$ , uncorrected for multiple comparisons. <sup>b</sup>Statistically significant at  $P < 0.05$ , corrected for multiple comparisons (10 items,  $P < 0.0125$ ).

**Table 4 Longitudinal consistency of subtype assignments**

Classification at previous visit	Classification at follow-up visit		
	Normal appearing <sup>a</sup>	Subcortical subtype	Cortical subtype
Normal appearing <sup>a</sup>	9 (69%)	3 (23%) <sup>b</sup>	1 (8%) <sup>b</sup>
Subcortical subtype	3 (1%)	265 (97%) <sup>b</sup>	5 (2%)
Cortical subtype	0 (0%)	3 (4%)	78 (96%) <sup>b</sup>

<sup>a</sup>Normal appearing = not subtypable. Note that this only includes 13 individuals that were not subtypable at baseline and had a follow-up scan. <sup>b</sup>An observation is longitudinally consistent if individuals remain in the same group or progress from the normal-appearing group to the subcortical or cortical subtype at follow-up visit. Entries indicate the number of visits  $n$ , with the % of the total individuals in classification at previous visit in classification at follow-up in brackets. Longitudinally consistent observations highlighted in bold.

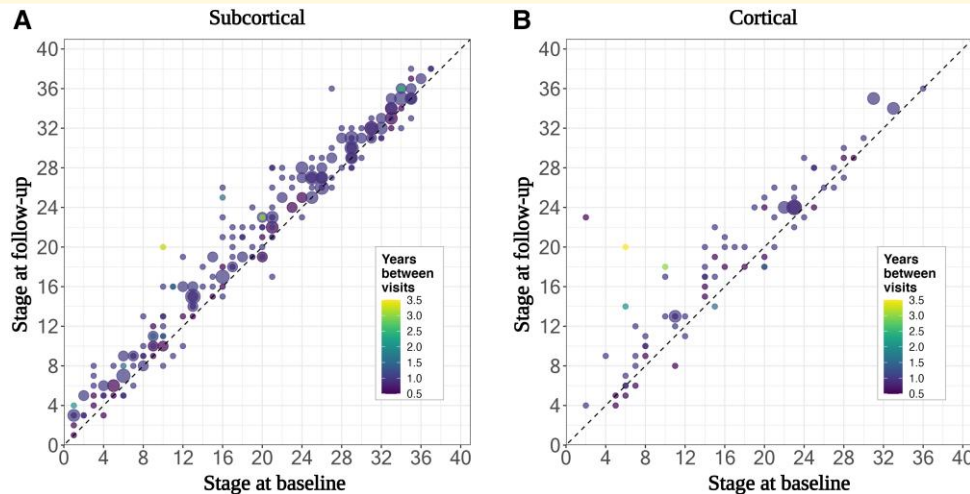
visits) either remaining in the same subtype or progressing from the normal-appearing (not subtypable) group to the 'subcortical' or 'cortical' subtypes. Ninety-seven per cent (265/273) of PSP scans assigned to the 'subcortical' subtype at previous scan remained in that subtype at follow-up; 2% (five scans) switched to the 'cortical' subtype at follow-up scan, whilst 1% (three scans) reverted to 'normal' (not subtypable). Of those scans assigned to the 'cortical' subtype, 96% (78/81) showed stable subtype assignment, whilst 4% (three scans) switched to the 'subcortical' subtype.

We next tested how the SuStaln stage progressed over time (Fig. 3) by comparing the assigned stage at follow-up to the baseline stage. As expected, the majority of individuals (90%) either progressed in stage (75%, 318/355) or stayed at the same stage (15%, 53/355), i.e. are on or above the line  $y = x$ . For those individuals assigned to the 'subcortical' subtype (Fig. 3A), 92% stayed at the same stage or progressed (17% and 75%, respectively); for those diagnosed with PSP-RS, 92% stayed at the same stage or progressed, for PSP-cortical 100%, and for PSP-subcortical 100%. In the 'cortical' subtype (Fig. 3B), 83% stayed at the same stage

or progressed (9% and 74%, respectively) in the 'cortical' subtype; the breakdown for clinical phenotypes assigned to this subtype was 83.9% for PSP-RS, PSP-subcortical 100%, and PSP-cortical 77%.

## Discussion

The clinical heterogeneity of PSP is increasingly recognized<sup>7,8</sup> and although post-mortem studies<sup>3,9,10</sup> suggest that this heterogeneity is related to differences in the severity and neuro-anatomical distribution of pathology, there is an urgent need to better delineate this variability in vivo. To this end, we applied SuStaln to a large PSP MRI data set, encompassing the spectrum of PSP clinical syndromes, and empirically identified two subtypes characterized by distinct temporal patterns of atrophy. We referred to these two subtypes as 'subcortical' and 'cortical' based on the earliest regions to show abnormality. Clinical scores of disease severity worsened with increasing stage and the 'subcortical' subtype was associated with more severe disease compared to the 'cortical' subtype, as measured by the PSP rating scale. It is worth noting, however, that the PSP rating scale was originally designed to assess disease severity in PSP-RS and as such is heavily weighted towards measuring motor predominant disease burden. It is therefore unlikely to be a good measure of overall disease burden in the cortical PSP syndromes,<sup>42</sup> and the scores in these cases relative to the PSP-RS/PSP-subcortical variants need to be interpreted with caution. As expected, the 'cortical' subtype was enriched for patients clinically diagnosed with PSP-cortical variants (PSP-CBS, PSP-F, and PSP-SL). The 'subcortical' subtype was enriched for patients clinically diagnosed with both PSP-RS and PSP-subcortical variants (PSP-P, PSP-PGF), though on average, the PSP-subcortical variants were at a lower subtype stage with a slower rate of progression compared to the PSP-RS cases. The distinct patterns of MRI atrophy in these subtypes provides unique insights into disease mechanisms across the



**Figure 3 Stage progression at follow-up visits by the SuStaln subtype.** Scatter plots of (A) the ‘subcortical’ subtype and (B) the ‘cortical’ subtype showing the predicted stage at baseline (x-axis) versus the predicted stage at follow-up scan (y-axis) for those PSP cases with a follow-up scan ( $n = 355$ ). The area of the circle is weighted by the number of scans at each point, and the colour of the circle represents the time (years) between visits.

disease course and supports the stratification of patients by subtype in clinical trials.

The ‘subcortical’ and ‘cortical’ SuStaln subtypes share some common early features (Fig. 1A and Supplementary Fig. 3), though they are distinguished by much earlier cortical involvement in the latter. Both subtypes demonstrate early involvement of the midbrain and ventral diencephalon (which includes the subthalamic nucleus and substantia nigra), in keeping with previous post-mortem studies,<sup>3,9</sup> and several neuropathological studies of incidental or early-stage PSP cases.<sup>43-50</sup> In contrast to the ‘subcortical’ subtype, the ‘cortical’ subtype has concomitant atrophy in the insula, frontal lobes, thalamus, and basal ganglia in these early stages. The subsequent cortical atrophy in the ‘cortical’ subtype then progresses to the parietal, occipital, and cingulate and finally the temporal lobe. At equivalent stages, the subcortical atrophy progresses to the medulla, SCP, dentate, and globus pallidus, followed by the amygdala, and finally the caudate and cerebellar cortex. In contrast, in the ‘subcortical’ subtype, atrophy has already progressed through the whole brainstem, SCP, and dentate nucleus, before it reaches the insula and posterior frontal lobe (supplementary motor cortex). Once it reaches the cortex, the sequence of cortical atrophy is broadly similar to the ‘cortical’ subtype.

The sequence of atrophy in the ‘subcortical’ subtype broadly mirrors the sequential spread of tau pathology proposed by Kovacs *et al.* in their post-mortem PSP-RS staging system and is consistent with other *in vivo* disease progression models of PSP-RS.<sup>25,51</sup> Whilst Scotton *et al.*<sup>25</sup> used an event-based modelling approach on PSP-RS cases to identify this sequence, Saito *et al.*<sup>51</sup> applied SuStaln to a small cohort of PSP-RS and CBS cases. They found that the optimal model consisted of two subtypes, with one subtype associated with CBS and the other to PSP-RS. Given the small sample size of the Saito study and the absence of any vPSP cases in the

sample, they were unable to extract finer grained information on PSP heterogeneity.<sup>25,51</sup>

An important difference in our study was that in addition to the majority of PSP-RS cases (81%) being assigned to the ‘subcortical’ subtype, the same was also true for the PSP-subcortical (PSP-P and PSP-PGF) cases (82%). The implication is that atrophy in PSP-subcortical variants progresses along the same trajectory as PSP-RS cases. Previous work shows that PSP-P and PSP-PGF cases develop similar clinical phenomenology to PSP-RS cases in the later stages of the disease course,<sup>52-54</sup> albeit at a slower rate<sup>17,54</sup> resulting in longer survival times.<sup>20</sup> *Post hoc* analysis of the ‘subcortical’ SuStaln subtype supports this, with the PSP-SC cases at an earlier stage in the sequence than PSP-RS cases (Fig. 2) (7.4 versus 19.9,  $t = 12.4$ ,  $P = 8.4 \times 10^{-5}$ ), with a slower stage progression per year (0.66 versus 1.86,  $t = 2.49$ , 95% CI 0.1–2.4,  $P = 0.046$ ).

Traditional cross-sectional imaging studies show that the PSP-subcortical variants usually have less severe atrophy in the midbrain, medulla, and SCP relative to PSP-RS cases.<sup>2,33,55-59</sup> Although this may seem at odds with SuStaln assigning PSP-RS and PSP-subcortical to the same trajectory, it can be reconciled by the finding that at the average stage for PSP-subcortical variants (7.4), the  $z$ -score has only reached 2 sigma (two standard deviations from controls) for the midbrain and 1 sigma for the medulla and SCP, versus 3, 2, and 2 sigma, respectively, by stage 20 (the average stage for the PSP-RS cases). This suggests that on average, PSP-SC cases have less atrophy in those regions at the time of baseline scan compared to PSP-RS cases, which would be consistent with the cross-sectional imaging findings that do not account for disease stage heterogeneity. It is currently unclear as to why the PSP-subcortical variants show a less aggressive disease, though it may be at least partly due to protective genetic variants.<sup>60</sup>

There was also a strong assignment of PSP–cortical cases (PSP-F, PSP-SL, and PSP-CBS) to the ‘cortical’ subtype, which drove a lower average PSP rating scale and UPDRS score compared to the ‘subcortical’ subtype. As previously discussed, the ‘cortical’ subtype had early cortical atrophy predominantly affecting the insula and frontal lobes alongside the subcortical involvement, which is in keeping with cross-sectional MRI studies of PSP–cortical variants.<sup>33,55,61,62</sup> Although there are few longitudinal imaging studies of these variants due to their rarity, a recent retrospective cohort of PSP-SL demonstrated that the majority of these cases developed symptoms typical of PSP-RS as their disease progressed.<sup>62</sup> Our data supports this finding given that the end-stage atrophy pattern of the ‘cortical’ SuStaIn subtype is very similar to the ‘subcortical’ atrophy pattern. Further work is required to better understand what is driving the difference between the ‘subcortical’ and ‘cortical’ subtype atrophy patterns, especially early in the disease trajectory. The early cortical involvement in the ‘cortical’ subtype may, at least in part, be due to higher astroglial or oligodendroglial pathology relative to neuronal tau pathology early in the disease course,<sup>3</sup> though this is an area of ongoing research.

## Limitations

This study has a number of limitations and highlights opportunities for future work.

Although we built a large international cohort of PSP cases with baseline and follow-up imaging, the sample size for vPSP cases was still small (52 PSP–cortical and 17 PSP–subcortical variants). Whilst the SuStaIn algorithm uses a cross-validation framework, ideally, one should have a separate training and test set to validate results. Given the low numbers for the vPSP syndromes in particular, we opted not to do this with our data set so as not to further underpower to model to find patterns associated with these syndromes. Future work will be required to confirm the validity of the two-subtype model on external PSP data sets. We may find that there are further distinct subtypes as vPSP numbers increase, though for this data set, the two-subtype model was optimal. Related to this, we decided to group the PSP–cortical (PSP-F, PSP-S/L, and PSP-CBS) and PSP–subcortical (PSP-P and PSP-PGF) syndromes together in the *post hoc* analyses. This is an established approach in the PSP research setting,<sup>20,33,60,63</sup> though it will be interesting to look at the finer-grained SuStaIn subtype:clinical syndrome associations when we have larger sample sizes for the vPSP syndromes.

We collected MRI scans from a range of international centres across a number of different scanners. It is well known that using data from different scanners can introduce bias into downstream analyses, through variations in imaging quality, signal homogeneity, and image contrast.<sup>64</sup> In addition to stringent visual QC of raw images and post-segmentation scans, we also regressed out both scanner manufacturer and field strength when generating adjusted regional volumes to try and account for this. In multi-centre clinical trials of a rare disease such as PSP, the use of multiple

scanner types is the rule rather than the exception. We believe that this study’s inclusion of cases across multiple scanner types (albeit adjusted for in the analysis) supports stronger generalizability of the findings to the wider clinical setting.

Another limitation was the use of the Lawton *et al.*<sup>34</sup> method to convert MoCA scores to MMSE. This method has only been validated in Parkinson’s disease, though it has been used previously in a PSP study.<sup>65</sup> Given that the MMSE is likely to be less sensitive than the MoCA in PSP,<sup>66</sup> the cognitive impairment estimated in this study is likely to be on the conservative side.

SuStaIn fits data based on the assumption that there are a distinct set of trajectories, though it is possible that there is a spectrum of disease progression patterns within the data.<sup>1</sup> In this situation, the identified SuStaIn subtype trajectory could have been created by appending unrelated disease trajectories into one subtype.<sup>26</sup> When assessing the two-subtype model, we checked the average subtype probability by stage by subtype (Supplementary Fig. 5B), based on the hypothesis that if a separate sequence is appended to another, we would expect the average probability assignment to drop below 50% at some point in the sequence. This was not the case in our model supporting that these are indeed two distinct trajectories. In addition, we found that the majority of individuals remained in the same subtype and progressed to later stages in that subtype which supports the model validity. The high association between PSP–subcortical syndromes and the ‘subcortical’ subtype at baseline could also be due to this caveat, i.e. they don’t necessarily progress through the stages of the subtype to cortical involvement. However, when assessing the stage at follow-up, one of the PSP-SC cases progressed from a baseline stage where there is only subcortical involvement (<stage 13) to a follow-up stage where the cortex started to be involved (>stage 13). Two other cases already had cortical involvement at baseline, progressing to more extensive cortical involvement at follow-up (Supplementary Table 7). Taken together, these findings give confidence that in our model, appending of different disease trajectories into one is unlikely to be the case.

Another consideration is that SuStaIn has no explicit time-scale and is only able to extract information on the relative position that the individual is in the sequence within a given subtype. A recently developed generative model called the temporal event-based model addresses this issue, by using longitudinal information to extract transition times between events.<sup>67</sup> Work to integrate this framework into the SuStaIn algorithm is ongoing. It will be interesting in future work to see whether a temporal SuStaIn model identifies a third subtype, given the finding in this study that the majority of PSP-SC cases follow the same trajectory as PSP-RS cases though with a slower stage progression rate.

## Conclusion

The SuStain model provides data-driven evidence for the existence of two spatiotemporal subtypes of atrophy in

clinically diagnosed PSP, giving insights into the relationship between PSP pathology and clinical syndrome. These image-based subtypes are differentially enriched for PSP clinical syndromes and show different clinical characteristics. The results suggest that the PSP-RS and PSP-subcortical syndromes share a similar trajectory of atrophy, though the latter tends to be at an early stage at diagnosis and progresses at a slower rate. Being able to accurately subtype and stage PSP patients at baseline has important implications for screening patients on entry into clinical trials, as well as for tracking disease progression. Future work should focus on validating these results in larger data sets with a higher number of vPSP syndromes that ideally have autopsy-confirmed PSP pathology, extracting information on time to transition between subtype stages, and assessing the clinical relevance of these imaging subtypes in real-world settings.

## Supplementary material

‘Supplementary material is available at *Brain Communications* online’

## Acknowledgements

We thank the research participants and their families for their contribution to the study. Part of the data used in the preparation of this manuscript were obtained from the PROgressive Supranuclear Palsy CorTico-Basal Syndrome Multiple System Atrophy Longitudinal Study (PROSPECT), a UK-wide longitudinal study of patients with atypical parkinsonian syndromes (Queen Square Research Ethics Committee 14/LO/1575). Part of the data used in the preparation of this manuscript were obtained from the 4-Repeat Neuroimaging Initiative (4RTNI) database and the Frontotemporal Lobar Degeneration Neuroimaging Initiative (FTLDNI) (<http://4rtni-ftldni.ini.usc.edu/>). 4RTNI was launched in early 2011 and is funded through the National Institute of Aging and The Tau Research Consortium. The primary goal of 4RTNI is to identify neuroimaging and biomarker indicators for disease progression in the 4-repeat tauopathy neurodegenerative diseases, progressive supranuclear palsy (PSP), and cortico-basal degeneration (CBD). FTLDNI is also founded through the National Institute of Aging and started in 2010. The primary goals of FTLDNI are to identify neuroimaging modalities and methods of analysis for tracking frontotemporal lobar degeneration (FTLD) and to assess the value of imaging versus other biomarkers in diagnostic roles. The Principal Investigator of 4RTNI is Dr. Adam Boxer, MD, PhD, at the University of California, San Francisco. The data is the result of collaborative efforts at four sites in North America. For more information on 4RTNI, please visit: <https://memory.ucsf.edu/research-trials/research/4rtni-2>. The Principal Investigator of NIFD is Dr Howard Rosen,

MD, at the University of California, San Francisco. The data is the result of collaborative efforts at three sites in North America. For up-to-date information on participation and protocol, please visit <http://memory.ucsf.edu/research/studies/nifd>.

## Funding

W.J.S. is supported by a Wellcome Trust Clinical PhD fellowship (220582/Z/20/Z). C.S. is supported by the UK Research and Innovation Medical Research Council (MR/S03546X/1). M.B. is supported by a fellowship award from the Alzheimer’s Society, UK (AS-JF-19a-004-517), and the UK Dementia Research Institute. D.M.C. is supported by the UK Dementia Research Institute, as well as Alzheimer’s Research UK (ARUK-PG2017-1946), and the University College London/University College London Hospitals, National Institute for Health and Care Research Biomedical Research Centre. H.H. is supported by the National Institutes of Health (R01AG038791, U19AG063911). A.L.Y. is supported by a Skills Development Fellowship from the Medical Research Council (MR/T027800/1). N.P.O. is a UK Research and Innovation Future Leaders Fellow (MR/S03546X/1). L.V.V. is supported by the National Institutes of Health (R01AG038791, K23AG073514) and the Alzheimer’s Association. D.C.A. is supported by the Engineering and Physical Sciences Research Council (EP/M020533/1), Medical Research Council (MR/T046422/1), and Wellcome Trust (UNS113739). J.B.R. is supported by the Wellcome Trust (220258), National Institute for Health and Care Research Cambridge Biomedical Research Centre (BRC-1215-20014), PSP Association, Evelyn Trust, and Medical Research Council (SUAG051 R101400). H.R.M. is supported by Parkinson’s UK, Cure Parkinson’s Trust, PSP Association, CBD Solutions, Drake Foundation, Medical Research Council, and the Michael J Fox Foundation. A.L.B. is supported by the National Institutes of Health (U19AG063911, R01AG038791, R01AG073482, and U24AG057437), the Rainwater Charitable Foundation, the Bluefield Project to Cure FTD, and the Alzheimer’s Association and the Association for Frontotemporal Degeneration. J.D.R. is supported by the Miriam Marks Brain Research UK Senior Fellowship and has received funding from a Medical Research Council Clinician Scientist Fellowship (MR/M008525/1) and the National Institute for Health and Care Research Rare Disease Translational Research Collaboration (BRC149/NS/MH). P.A.W. is supported by a Medical Research Council Skills Development Fellowship (MR/T027770/1).

The Dementia Research Centre is supported by Alzheimer’s Research UK, Alzheimer’s Society, Brain Research UK, and The Wolfson Foundation. This work was supported by the National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre, the Leonard Wolfson

Experimental Neurology Centre (LWENC) Clinical Research Facility, and the UK Dementia Research Institute, which receives its funding from UK DRI Ltd., funded by the UK Medical Research Council, Alzheimer's Society, and Alzheimer's Research UK. The PROSPECT study is funded by the PSP Association and CBD Solutions. The 4-Repeat Tauopathy Neuroimaging Initiative (4RTNI) and FTLDNI are funded by the National Institutes of Health Grant (R01 AG038791) and through generous contributions from the Tau Research Consortium. Both are coordinated through the University of California, San Francisco, Memory and Aging Center. 4RTNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

## Competing interests

The authors report no competing interests.

## Data availability

Source data are not publicly available but non-commercial academic researcher requests may be made to the chief investigators of the seven source studies, subject to data access agreements and conditions that preserve participant anonymity. The underlying SuStaIn model code is publicly available at <https://github.com/ucl-pond/pySuStaIn>.<sup>68</sup>

## Appendix

### 4RTNI Consortium

**Bradley F. Boeve**—Department of Neurology, Mayo Clinic, Rochester, MN 55905, USA.

**Brad C. Dickerson**—Departments of Neurology and Psychiatry, Frontotemporal Disorders Unit and Alzheimer's Disease Research Center, Boston Massachusetts USA.

**Carmela M. Tartaglia**—Tanz Centre for Research in Neurodegenerative Diseases University of Toronto Toronto Canada.

**Irene Litvan**—Department of Neurosciences, University of California San Diego, La Jolla, California, USA.

**Murray Grossman**—Department of Neurology, University of Pennsylvania, Philadelphia, USA.

**Alex Pantelyat**—Department of Neurology. School of Medicine, Johns Hopkins University, Baltimore, MD, USA.

**Edward D. Huey**—Department of Psychiatry and Neurology, Columbia University, New York, New York, USA.

**David J. Irwin**—Penn Center for Neurodegenerative Disease Research, University of Pennsylvania School of Medicine, Philadelphia, PA, USA.

**Anne Fagan**—Department of Neurology, Washington University School of Medicine, St Louis, MO, USA.

**Suzanne L. Baker**—Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

**Arthur W. Toga**—Laboratory of Neuro Imaging, Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States.

## PROSPECT Consortium

**Alyssa A. Costantini**, MSc—Department of Clinical and Movement Neurosciences, UCL (University College London) Queen Square Institute of Neurology, London, United Kingdom. [a.costantini@ucl.ac.uk](mailto:a.costantini@ucl.ac.uk).

**Henry Houlden**, FRCP, PhD—Department of Clinical and Movement Neurosciences, UCL (University College London) Queen Square Institute of Neurology, London, United Kingdom; Movement Disorders Centre, UCL Queen Square Institute of Neurology, London, United Kingdom; Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London, United Kingdom. [h.houlden@ucl.ac.uk](mailto:h.houlden@ucl.ac.uk).

**Christopher Kobylecki**, FRCP, PhD—Department of Neurology, Manchester Academic Health Science Centre, Salford Royal NHS (National Health Service) Foundation Trust, University of Manchester, Manchester, United Kingdom. [Christopher.Kobylecki@srft.nhs.uk](mailto:Christopher.Kobylecki@srft.nhs.uk).

**Michele T. M. Hu**, FRCP, PhD—Division of Neurology, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom [michele.hu@ndcn.ox.ac.uk](mailto:michele.hu@ndcn.ox.ac.uk).

**Nigel Leigh**, FRCP, PhD—Department of Neuroscience, Brighton and Sussex Medical School, Brighton, United Kingdom. [P.Leigh@bsms.ac.uk](mailto:P.Leigh@bsms.ac.uk).

## References

1. Young AL, Marinescu R V, Oxtoby NP, *et al.* Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nat Commun.* 2018;9:4273.
2. Stamelou M, Respondek G, Giagkou N, Whitwell JL, Kovacs GG, Höglinger GU. Evolving concepts in progressive supranuclear palsy and other 4-repeat tauopathies. *Nat Rev Neurol.* 2021;17:601-620.
3. Kovacs GG, Lukic MJ, Irwin DJ, *et al.* Distribution patterns of tau pathology in progressive supranuclear palsy. *Acta Neuropathol.* 2020;140:99-119.
4. Respondek G, Stamelou M, Kurz C, *et al.* The phenotypic spectrum of progressive supranuclear palsy: A retrospective multicenter study of 100 definite cases. *Mov Disord.* 2014;29:1758-1766.
5. Höglinger GU, Litvan I, Mendonca N, *et al.* Safety and efficacy of tilavonemab in progressive supranuclear palsy: A phase 2, randomised, placebo-controlled trial. *Lancet Neurol.* 2021;20:182-192.
6. Dam T, Boxer AL, Golbe LI, *et al.* Safety and efficacy of anti-tau monoclonal antibody gosuranemab in progressive supranuclear palsy: A phase 2, randomized, placebo-controlled trial. *Nat Med.* 2021;27:1451-1457.
7. Franzmeier N, Höglinger GU. Inferring the sequence of brain volume changes in progressive supranuclear palsy using MRI. *Brain Commun.* 2022;4:fcac113.

8. Höglinger GU, Respondek G, Stamelou M, *et al.* Clinical diagnosis of progressive supranuclear palsy: The movement disorder society criteria. *Mov Disord.* 2017;32:853-864.
9. Williams DR, Holton JL, Strand C, *et al.* Pathological tau burden and distribution distinguishes progressive supranuclear palsy-parkinsonism from Richardson's syndrome. *Brain.* 2007;130:1566-1576.
10. Williams DR, De Silva R, Paviour DC, *et al.* Characteristics of two distinct clinical phenotypes in pathologically proven progressive supranuclear palsy: Richardson's syndrome and PSP-parkinsonism. *Brain.* 2005;128:1247-1258.
11. Ali F, Martin PR, Botha H, *et al.* Sensitivity and specificity of diagnostic criteria for progressive supranuclear palsy. *Mov Disord.* 2019;34:1144-1153.
12. Respondek G, Grimm M, Piot I, *et al.* Validation of the Movement Disorder Society criteria for the diagnosis of 4-repeat tauopathies. *Mov Disord.* 2020;35:171-176.
13. Grimm MJ, Respondek G, Stamelou M, *et al.* How to apply the movement disorder society criteria for diagnosis of progressive supranuclear palsy. *Mov Disord.* 2019;34:1228-1232.
14. Höllerhage M, Kliez M, Höglinger GU. Disease modification in parkinsonism: Obstacles and ways forward. *J Neural Transm.* 2022;129:1133-1153.
15. Shoeibi A, Litvan I, Juncos JL, *et al.* Are the International Parkinson Disease and Movement Disorder Society progressive supranuclear palsy (IPMDS-PSP) diagnostic criteria accurate enough to differentiate common PSP phenotypes? *Park Relat Disord.* 2019;69:34-39.
16. Frank A, Peikert K, Linn J, Brandt MD, Hermann A. MDS criteria for the diagnosis of progressive supranuclear palsy overemphasize Richardson syndrome. *Ann Clin Transl Neurol.* 2020;7:1702-1707.
17. Shoeibi A, Litvan I, Tolosa E, Ser Td, Lee E. Progression of two progressive supranuclear palsy phenotypes with comparable initial disability. *Park Relat Disord.* 2019;66:87-93.
18. Jellinger KA. Different tau pathology pattern in two clinical phenotypes of progressive supranuclear palsy. *Neurodegener Dis.* 2008;5:339-346.
19. Nath U, Ben-Shlomo Y, Thomson RG, Lees AJ, Burn DJ. Clinical features and natural history of progressive supranuclear palsy: A clinical cohort study. *Neurology.* 2003;60:910-916.
20. Guasp M, Molina-Porcel L, Painous C, *et al.* Association of PSP phenotypes with survival: A brain-bank study. *Parkinsonism Relat Disord.* 2021;84:77-81.
21. O'Sullivan SS, Massey LA, Williams DR, *et al.* Clinical outcomes of progressive supranuclear palsy and multiple system atrophy. *Brain.* 2008;131(Pt 5):1362-1372.
22. Dell'Aquila C, Zoccollella S, Cardinali V, *et al.* Predictors of survival in a series of clinically diagnosed progressive supranuclear palsy patients. *Parkinsonism Relat Disord.* 2013;19:980-985.
23. Couto B, Martinez-Valbuena I, Lee S, *et al.* Protracted course-progressive supranuclear palsy (PC-PSP). *Eur J Neurol.* 2022;29:1-12.
24. Stamelou M, Giagkou N, Höglinger GU. One decade ago, one decade ahead in progressive supranuclear palsy. *Mov Disord.* 2019;34:1284-1293.
25. Scotton WJ, Bocchetta M, Todd E, *et al.* A data-driven model of brain volume changes in progressive supranuclear palsy. *Brain Commun.* 2022;4:fac098.
26. Vogel JW, Young AL, Oxtoby NP, *et al.* Four distinct trajectories of tau deposition identified in Alzheimer's disease. *Nat Med.* 2021;27:871-881.
27. Eshaghi A, Young AL, Wijeratne PA, *et al.* Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. *Nat Commun.* 2021;12:1-12.
28. Young AL, Bocchetta M, Russell LL, *et al.* Characterizing the clinical features and atrophy patterns of MAPT-related frontotemporal dementia with disease progression modeling. *Neurology* 2021;97:e941-e952.
29. Dutt S, Binney RJ, Heuer HW, *et al.* Progression of brain atrophy in PSP and CBS over 6 months and 1 year. *Neurology.* 2016;87:2016-2025.
30. Zhang Y, Walter R, Ng P, *et al.* Progression of microstructural degeneration in progressive supranuclear palsy and corticobasal syndrome: A longitudinal diffusion tensor imaging study. *PLoS One.* 2016;11:e0157218.
31. Boxer AL, Lang AE, Grossman M, *et al.* Davunetide in patients with progressive supranuclear palsy: A randomised, double-blind, placebo-controlled phase 2/3 trial. *Lancet Neurol.* 2014;13:676-685.
32. VandeVrede L, Dale ML, Fields S, *et al.* Open-label phase 1 futility studies of salsalate and young plasma in progressive supranuclear palsy. *Mov Disord Clin Pract.* 2020;7:440-447.
33. Jabbari G, Holland N, Chelban V, *et al.* Diagnosis across the spectrum of progressive supranuclear palsy and corticobasal syndrome. *JAMA Neurol.* 2020;77:377-387.
34. Lawton M, Kasten M, May MT, *et al.* Validation of conversion between mini-mental state examination and Montreal Cognitive Assessment. *Mov Disord.* 2016;31:593-596.
35. Cardoso MJ, Modat M, Wolz R, *et al.* Geodesic information flows: Spatially-variant graphs and their application to segmentation and fusion. *IEEE Trans Med Imaging.* 2015;34:1976-1988.
36. Johnson EB, Gregory S, Johnson HJ, *et al.* Recommendations for the use of automated gray matter segmentation tools: Evidence from Huntington's disease. *Front Neurol.* 2017;8:519.
37. Perlaki G, Horvath R, Nagy SA, *et al.* Comparison of accuracy between FSL's FIRST and FreeSurfer for caudate nucleus and putamen segmentation. *Sci Rep.* 2017;7:1-9.
38. Diedrichsen J, Balsters JH, Flavell J, Cussans E, Ramnani N. A probabilistic MR atlas of the human cerebellum. *Neuroimage.* 2009;46:39-46.
39. Iglesias JE, Van Leemput K, Bhatt P, *et al.* Bayesian Segmentation of brainstem structures in MRI. *Neuroimage.* 2015;113:184-195.
40. Collij LE, Salvadó G, Wotschel V, *et al.* Spatial-temporal patterns of amyloid- $\beta$  accumulation: A subtype and stage inference model analysis. *Neurology.* 2022;98:e1692-e1703.
41. Marinescu R V, Eshaghi A, Alexander DC, Golland P. Brainpainter: A software for the visualisation of brain structures, biomarkers and associated pathological processes. *Multimodal Brain Image Anal Math Found Comput Anat.* 2019;11846:112-120.
42. Piot I, Schweyer K, Respondek G, *et al.* The progressive supranuclear palsy clinical deficits scale. *Mov Disord.* 2020;35:650-661.
43. Dugger BN, Hentz JG, Adler CH, *et al.* Clinicopathological outcomes of prospectively followed normal elderly brain bank volunteers. *J Neuropathol Exp Neurol.* 2014;73:244-252.
44. Evidente VGH, Adler CH, Sabbagh MN, *et al.* Neuropathological findings of PSP in the elderly without clinical PSP: Possible incidental PSP? *Parkinsonism Relat Disord.* 2011;17:365-371.
45. Halliday GM, Macdonald V, Henderson JM. A comparison of degeneration in motor thalamus and cortex between progressive supranuclear palsy and Parkinson's disease. *Brain.* 2005;128:2272-2280.
46. Nogami A, Yamazaki M, Saito Y, *et al.* Early stage of progressive supranuclear palsy: A neuropathological study of 324 consecutive autopsy cases. *J Nippon Med Sch.* 2015;82:266-273.
47. Sakai K, Yamada M. Early-stage progressive supranuclear palsy with degenerative lesions confined to the subthalamic nucleus and substantia nigra. *Neuropathology.* 2011;31:77-81.
48. Yokota O, Tsuchiya K, Arai T, *et al.* Clinicopathological characterization of Pick's disease versus frontotemporal lobar degeneration with ubiquitin/TDP-43-positive inclusions. *Acta Neuropathol.* 2009;117:429-444.
49. Yoshida K, Hata Y, Kinoshita K, Takashima S, Tanaka K, Nishida N. Incipient progressive supranuclear palsy is more common than expected and may comprise clinicopathological subtypes: A forensic autopsy series. *Acta Neuropathol.* 2017;133:809-823.



50. Kovacs GG, Milenkovic I, Wöhrer A, *et al.* Non-Alzheimer neurodegenerative pathologies and their combinations are more frequent than commonly believed in the elderly brain: A community-based autopsy series. *Acta Neuropathol.* 2013;126:365-384.
51. Saito Y, Kamagata K, Wijeratne PA, *et al.* Temporal progression patterns of brain atrophy in corticobasal syndrome and progressive supranuclear palsy revealed by Subtype and Stage Inference (SuStaIn). *Front Neurol.* 2022;13:814768.
52. Necpál J, Borsek M, Jeleňová B. “Parkinson’s disease” on the way to progressive supranuclear palsy: A review on PSP-parkinsonism. *Neurol Sci.* 2021;42:4927-4936.
53. Srulijes K, Mallien G, Bauer S, *et al.* In vivo comparison of Richardson’s syndrome and progressive supranuclear palsy-parkinsonism. *J Neural Transm.* 2011;118:1191-1197.
54. Jecmenica-Lukic M, Petrovic IN, Pekmezovic T, Kostic VS. Clinical outcomes of two main variants of progressive supranuclear palsy and multiple system atrophy: A prospective natural history study. *J Neurol.* 2014;261:1575-1583.
55. Whitwell JL, Tosakulwong N, Botha H, *et al.* Brain volume and flor-taucipir analysis of progressive supranuclear palsy clinical variants. *Neuroimage Clin.* 2020;25:102152.
56. Longoni G, Agosta F, Kostic VS, *et al.* MRI Measurements of brainstem structures in patients with Richardson’s syndrome, progressive supranuclear palsy-parkinsonism, and Parkinson’s disease. *Mov Disord.* 2011;26:247-255.
57. Agosta F, Kostic VS, Galantucci S, *et al.* The in vivo distribution of brain tissue loss in Richardson’s syndrome and PSP-parkinsonism: A VBM-DARTEL study. *Eur J Neurosci.* 2010;32:640-647.
58. Hong JY, Yun HJ, Sunwoo MK, *et al.* Comparison of regional brain atrophy and cognitive impairment between pure akinesia with gait freezing and Richardson’s syndrome. *Front Aging Neurosci.* 2015;7:180.
59. Nakahara K, Nakane S, Kitajima M, Masuda-Narita T, Matsuo H, Ando Y. Diagnostic accuracy of MRI parameters in pure akinesia with gait freezing. *J Neurol.* 2020;267:752-759.
60. Jabbari E, Woodside J, Tan MMX, *et al.* Variation at the TRIM11 locus modifies progressive supranuclear palsy phenotype. *Ann Neurol.* 2018;84:485-496.
61. Santos-Santos MA, Mandelli ML, Binney RJ, *et al.* Features of patients with nonfluent/agrammatic primary progressive aphasia with underlying corticobasal degeneration or progressive supranuclear palsy pathology. *JAMA Neurol.* 2016;73:733.
62. Whitwell JL, Stevens CA, Duffy JR, *et al.* An evaluation of the progressive supranuclear palsy speech/language variant. *Mov Disord Clin Pract.* 2019;6:452-461.
63. Jabbari E, Koga S, Valentino RR, *et al.* Genetic determinants of survival in progressive supranuclear palsy: A genome-wide association study. *Lancet Neurol.* 2021;20:107-116.
64. Höglinger GU, Schöpe J, Stamelou M, *et al.* Longitudinal magnetic resonance imaging in progressive supranuclear palsy: A new combined score for clinical trials. *Mov Disord.* 2017;32:842-852.
65. Brendel M, Barthel H, Eimeren TV, *et al.* Assessment of 18 F-PI-2620 as a biomarker in progressive supranuclear palsy. *JAMA Neurol.* 2020;77:1408-1419.
66. Fiorenzato E, Weis L, Falup-Pecurariu C, *et al.* Montreal Cognitive Assessment (MoCA) and Mini-Mental State Examination (MMSE) performance in progressive supranuclear palsy and multiple system atrophy. *J Neural Transm.* 2016;123:1435-1442.
67. Wijeratne PA, Alexander DC. Learning transition times in event sequences: The temporal event-based model of disease progression. Springer; 2021:583-595.
68. Aksman LM, Wijeratne PA, Oxtoby NP, *et al.* Pysustain: A Python implementation of the subtype and stage inference algorithm. *SoftwareX.* 2021;16:100811.