Article

A 15-Item modification of the PSP rating scale to improve clinical meaningfulness and statistical performance

Tien Dam [1], Lili Yang [1], Colin Gillis[1], Yumeng Li[1], John O'Gorman [1] ✉,
Barbara Kolb[1], Adam L. Boxer [2], Lawrence Golbe[3] &
Samantha Budd Haeberlein [1] On behalf of the AL-108-231, PROSPERA, and
4RTNI investigators*

Progressive supranuclear palsy (PSP) is a rare neurodegenerative disorder characterized by physical, cognitive, and behavioral impairments. The PSP Rating Scale (PSPRS) is a widely used and validated, clinical scale to monitor disease progression. Here we show the modification of PSPRS to improve clinical meaningfulness and sensitivity. A conceptual framework was used to select meaningful items. Datasets were analyzed to improve the psychometric properties, understand factor structure, and develop a scoring algorithm. Fifteen items of the PSPRS were selected based on whether the items reflect concepts specific to the disease and those responsive to change over 12 months. Some items were rescored to reflect meaningful changes and to improve discrimination and reliability. The rescored 15-item PSPRS was more interpretable and sensitive to disease progression over 12 months. The approach to develop the 15-item PSPRS provides an example of how to modify an existing clinical outcome assessment for use as a primary endpoint in a clinical trial of a progressive neurological disorder.

Progressive supranuclear palsy (PSP) is a rare neurodegenerative disorder, with the most common phenotype, PSP–Richardson syndrome, characterized by progressive impairments in gait/postural instability, ocular motor, swallowing, speech, sleep, cognition, and behavior typically starting in the early 60 s and leading to death, on average, 7 years after symptom onset[1–3]. It is a rare and sporadic disease, occurring in adults with a prevalence of approximately 6 cases per 100,000 individuals[4,5]. Clinical symptoms appear to correlate with the distribution and density of 4 microtubule-binding domain repeat (4R) tau pathology in neurons and glia in the basal ganglia, diencephalon, brainstem, and cerebellum with restricted involvement of the neocortex[6].

There is currently no treatment available to slow the progression of this fatal disease. Tau targeting therapies are in development for indications such as Alzheimer's disease and PSP[7]. To measure the impact of therapies designed to slow disease progression, clinician- and patient-reported measures are often used as clinical outcome assessments (COAs) in clinical trials.

Since the sixth authorization of the Prescription Drug User Fee Act VI under Title I of the US Food and Drug Administration (FDA) Reauthorization Act of 2017, the FDA has developed a series of 4 methodological patient-focused drug development (PFDD) guidance documents to enhance incorporation of the patient's voice in medical product development and regulatory decision-making[8]. In PFDD guidance 1 and 2, the FDA provides methods to collect patient experience data and approaches to identify what is most important to patients as it relates to burden of disease and the burden of treatment. In PFDD guidance 3, the FDA provides a roadmap for patient-focused outcome

[1]Biogen Inc, Cambridge, MA, USA. [2]Memory and Aging Center, Department of Neurology, University of California, San Francisco, CA, USA. [3]Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, USA. *A list of authors and their affiliations appears at the end of the paper.
✉ e-mail: john.ogorman@biogen.com

measurement in clinical trials that includes suggestions to understand the disease, conceptualize the treatment benefit, and select or optimize a COA. A well-defined COA should specify (1) the concepts of interest that are relevant to an individual's experience or clinical, biological, physical, or functional state and (2) the context of use, including how the COA will be used, the definition of the target population, comparator groups, and timing and implementation of the assessments[8]. A challenge, however, is that well-developed and fit-for-purpose COAs do not exist for many disease/conditions. Furthermore, development of new COAs can be time and resource intensive. Thus, modification of an existing COA may be a preferred alternative.

In recent PSP clinical trials, the score from the PSP Rating Scale (PSPRS), a standardized clinician-rated instrument developed to monitor progression for routine patient care[9], has been used as the primary endpoint[10–16]. Change on the total PSPRS score correlates with changes on other assessments that measure various aspects of disease severity, including brain volume, cognition, and activities of daily living[17]. A linear change of approximately 11 points per year is generally observed in patients with PSP recruited in clinical trials[10,16,18]. The PSPRS score has demonstrated good inter-rater reliability and predicts gait ability and survival in patients with PSP[19]. While the PSPRS is a good scale for clinicians to measure progression, some of its clinimetric properties (i.e., cohesiveness of the scale, overweighting, and responsiveness to change) may limit its use as a primary endpoint in a clinical trial. Efforts to address some of these challenges have recently been published[20–22].

We report on the development process and evidence generated to support the modification of an existing COA, following the PFDD principles and using data from completed natural history studies and clinical trials. The objective of this work was to improve the psychometric properties and performance of the PSPRS score for use as a primary endpoint in the PASSPORT study (NCT03068468), a phase 2 clinical trial to evaluate the efficacy, safety, and tolerability of gosuranemab, a monoclonal antibody targeting extracellular tau, in participants diagnosed with PSP[11]. This work was completed while the PASSPORT study was ongoing, and data collection on the PSPRS scores remained blinded. The scoring algorithm for the 15-item PSPRS was included in the statistical analysis plan that was finalized prior to unblinding. This work to analyze and refine the structure of the PSPRS was based on advice and iterative engagements with the FDA Division of Neurology throughout the gosuranemab development program. Development of an improved COA that captures what matters to individuals with PSP, is reliable, and is sensitive to change is critical to aid in the diagnosis, assessment of disease progression, and evaluation of a clinically meaningful effect of an intervention.

Results

PSP Conceptual Framework

A PubMed literature search was conducted to identify articles exploring experiences of PSP from the patient perspective using a combination of the following search terms: ([progressive supranuclear palsy] OR Steele Richardson Olszewski) AND ([[(qualitative) OR focus group] OR interviews). The PubMed search identified 49 articles, which were screened by title/abstract; 5 underwent full-text review and 2 were included in the review[23,24]. Findings from the reviewed literature were used to develop an initial conceptual framework composed of 3 overall concepts: (1) motor symptoms and functions, (2) non-motor symptoms and functions, and (3) the impact of PSP on patients' lives.

Patients, carers, and physician interviews were conducted between January and March 2018 by a third-party market research vendor to (1) explore the PSP journey, including first symptoms to diagnosis, timeline of progression, and (2) patient, caregiver, and physician preferences. For patient preferences, participants were asked specifically: Which symptoms are most bothersome? With a disease-modifying therapy, which symptoms should receive priority?

The interview sample included 46 patients and/or their carers and 78 PSP-treating physicians, such as movement disorder specialists, general neurologists, and primary care physicians from the US, Germany, France, Italy, UK, Spain, and Japan. Interviewed participants were not PASSPORT trial participants and were not selected based on pre-specified criteria such as age, sex, disease severity, or PSP subtype; participant information was blinded.

A subset of English-speaking interviews (19 patients/caregivers and 16 clinicians from the US and UK) were transcribed, and the transcripts were evaluated to elicit core disease-related concepts, concepts relevant and important to patients, and the potential impacts of these concepts and to understand which aspects of patient experience with the disease should be targeted by a medical product for a meaningful treatment benefit. Line-by-line coding of the transcripts resulted in a rich pool of 564 unique codes and produced a comprehensive set of concepts important to individuals living with PSP. The interviews confirmed and expanded on the 3 overall concepts identified through the literature review. New item-level concepts within the motor symptoms and functions included akinesia, dystonia, stiffness, ocular apraxia, and posture and facial expression issues. Within the non-motor symptoms and functions overall concept, additional item-level concepts included time loss episodes and issues with reasoning, problem-solving, and visuospatial processing. Neuropsychiatric symptoms such as personality and behavioral problems, as well as sleep problems, abnormal dreams, and fatigue, were added to the conceptual framework. Concepts important to patients and/or caregivers are shown in Fig. 1.

During the interviews, patients indicated that slowing the progression of deterioration in speech, swallowing, vision, and cognitive and behavioral issues was important. Patients, caregivers, and clinicians prioritized vision, swallowing, and mobility as important targets for treatment.

Steps to Modify the PSPRS

The 28 PSPRS items were mapped and qualitatively compared with the content of the conceptual framework. Overall, RMT analyses demonstrated that the 28-item PSPRS total score appropriately targeted the patient population and the range of PSP disability. However, the RMT analysis identified that some items may be ambiguous and non-cohesive with the overall scale. Although the PSP conceptual framework identified 3 overall concepts, we focused on PSP motor symptoms and functional issues as the concept of interest. To improve the ability of the PSPRS to capture motor symptoms and functions, we considered several modifications, including evaluating the content and method of scoring for each item (ie, collapsing response options or changing the scoring algorithm). Because the phase 2 PASSPORT study was enrolling and data collection with the PSPRS was ongoing, adding or modifying instructions/training materials, modifying the recall period, adding items, or modifying the wording of items or response options was not possible.

Step 1: Item selection

Each PSPRS item was evaluated based on several considerations, including whether patients identified these items as clinically meaningful to daily activities, whether the measured concept was considered core to disease and not a downstream consequence of the disease or other comorbidities, and responsiveness to change over a 12-month clinical trial duration. Fifteen of the 28 items met these criteria and were selected for retention in the modified PSPRS. In addition, these selected items were the most responsive to change over 12 months, and the inclusion of these items maximized the overall ability of the scale to measure disease progression in the natural history datasets.

The remaining 13 items were excluded from the modified PSPRS for different reasons, including (1) items were least related to daily

Fig. 1 | Concepts Important to Patients With PSP Based on Literature Review and Patient/Caregiver and Clinician Interviews. Color coding: Domains are reflected by color coding which matches the corresponding overarching domains. Sub-domains are reflected by two types of lighter shading of same color as associated domains. Sub-domains are sorted alphabetically within each domain. Patient only ‾, Clinician only ᵓ, Article only ˣ, ADL = Activities of daily living, IADL = Instrumental activities of daily living. Definitions: Activities of daily living (ADL) relates to routine activities including eating, bathing, dressing, toileting, transferring, and continence. Instrumental activities of daily living (IADL) relates to independent living and include preparing meals, managing money, shopping for groceries, performing housework, using a telephone, and doing laundry.

Table 1 | Excluded PSPRS Items and Rationale

Excluded PSPRS items	Rationale for exclusion
Sleep difficulty Urinary incontinence	Not primary manifestations of PSP; may be downstream consequences or confounded by comorbidities
Tremor Limb rigidity Irritability. Disorientation Emotional incontinence	Contributed little to disease progression or were least responsive to change over time
Withdrawal Bradyphrenia	Cognitive or non-motor symptoms
Finger tapping Toe tapping	Based on semi-quantitative neurological examination-based assessments that are not clinically meaningful to patients
Apraxia of hand movement	High potential for variability in clinician ratings within and across patients
Grasping	The change in patient ability to function associated with a change in score from absent to present is not assessed by the PSPRS

PSPRS Progressive Supranuclear Palsy Rating Scale.

activities or in some cases may have represented a downstream impact of the disease or may have been confounded by comorbidities (eg, sleep difficulty and urinary incontinence)[25,26]; (2) items, although considered primary or a core concept related to the disease, contributed little to disease progression or disability (eg, tremor, limb rigidity, irritability, disorientation, and emotional incontinence); or (3) items related to cognition or non-motor symptoms (eg, bradyphrenia and withdrawal) were excluded as other COAs could measure the concepts with greater specificity and sensitivity[27]. Additionally, a few items were removed as their response options did not reflect clear and meaningful differences in patients' conditions. Although finger tapping and toe tapping measure bradykinesia, the response options were based on semi-quantitative neurological examination-based

assessments that may not directly capture meaningful functional impact compared with other items that were included (ie, item 4 ability to hold a knife and fork). The excluded items and rationale for exclusion are summarized in Table 1.

Step 2: Item Response Options

The response options for each item were reviewed to assess whether the response options were non-overlapping and differences among adjacent response categories reflected true differences in motor function. The RMT analysis demonstrated that multiple items had overlapping response options, which could lead to inconsistencies in response patterns and add potential bias and variability to the data. Of the 15 items selected for the modified PSPRS, only the response

Table 2 | 15-Item PSPRS With Collapsing and Re-Scoring

PSPRS item no.	PSPRS item name	Method for collapsing and re-scoring response options
3	Dysphagia for solids	No collapsing
4	Using knife and fork, buttoning clothes, washing hands and face	No collapsing
5	Falls	Collapsing per Rasch analyses
12	Dysarthria	Collapsing based on patient relevant functional differences
13	Dysphagia for liquids	Collapsing per Rasch analyses
14	Voluntary upward saccades	Collapsing per Rasch analyses
15	Voluntary downward saccades	Collapsing per Rasch analyses
16	Voluntary left and right saccades	Collapsing per Rasch analyses
17	Eyelid dysfunction	Collapsing based on patient relevant functional differences
19	Limb dystonia	Collapsing per Rasch analyses
24	Neck rigidity	Collapsing per Rasch analyses
25	Arising from chair	Collapsing based on patient relevant functional differences
26	Gait	Collapsing based on patient-relevant functional differences
27	Postural instability	Collapsing based on patient-relevant functional differences
28	Sitting down	Collapsing based on patient-relevant functional differences

PSPRS Progressive Supranuclear Palsy Rating Scale.

options for item 3 (dysphagia for solids) and item 4 (using knife and fork, buttoning clothes, washing hands and face) were non-overlapping and meaningfully different. The response options of the remaining 13 items did not all represent clear and meaningful differences in a patient's condition. Hence, response options required collapsing (or combining) into discrete, functionally meaningful categories to improve interpretability to patients and increase reproducibility. To guide collapsing of response options, 2 methods were considered: (1) re-scoring to reflect clear and functional meaningful differences in a patient's condition or (2) method based on Rasch analysis results (Table 2 and Supplemental Table 1).

However, collapsing response options resulted in some items being measured using fewer intervals than was the case in the original PSPRS. Items measured with different intervals would result in differential weighting, with collapsed items contributing less weight to the total score. Thus, we re-scored response options to provide each item with a 4-point maximum as follows: a score of 0 was assigned for no impairment or slight impairment without an impact on function; a score of 4 was assigned for maximum impairment or total loss of function; and a score of 1, 2, or 3 was assigned for intermediate levels of impairment. Items measured with a 3-point interval score of 0, 2, or 4 were assigned; items measured with a 4-point interval score of 0, 1, 2, or 4 were assigned. The modified response options are presented in Table 2.

Step 3: Factor Structure and Total Scoring Algorithm
To understand whether multiple items could be combined to generate a total score, CFA was conducted. Among individuals experiencing PSP, clinical rationale, presumed underlying functional neuroanatomical etiopathology, and CFAs, we hypothesized 3 distinct underlying disease constructs for the 15-item PSPRS before conducting the CFA: (1) gait and limb function, (2) ocular motor, and (3) bulbar. The correlation between these 3 constructs ranged from 0.36 to 0.56, with the highest correlation between gait/limb function and bulbar and the lowest correlation between gait/limb function and ocular motor. Model-fitting indices of the CFA, comparative fit index=0.964, TFI = 0.956, RMSEA = 0.046, and SRMR = 0.051, supported a good model fit with the 3-factor structure. The factor-loading estimates from the CFA are summarized in Supplement Table 2. In addition, a CFA assuming 1-factor structure was conducted; model fit indices did not support a 1-factor structure.

Finally, analyses with different weighting approaches in factor score and total score calculation were conducted and compared to

evaluate the scoring algorithm. The 3 approaches were as follows: (a) using factor-loading estimates from the CFA; before the addition of items in the same factor, the factor loading of each item was multiplied by the scale score of each item; (b) using factor-loading estimates from the CFA, for which the factor loadings were rounded to 1, 0.5. or 0.25 (factor loadings ≥0.75 are rounded to 1, factor loadings >0.25 and <0.75 are rounded to 0.5, and factor loadings ≤0.25 are rounded to 0.25); and (c) not using the factor-loading estimates from the CFA (an unweighted approach). Each item in a factor was given a weight of 1. Similar results (mean SD ratio for the change from baseline score at 1 year) were observed between the weighted and unweighted approaches. Based on these results, an unweighted approach was used for the total score calculation, except for the 3 highly correlated bulbar items. Given the high correlation between the voluntary saccade items (eg, upward, downward, and left and right), the average of these 3 items was used to calculate the total score. The 15-item PSPRS total score algorithm is presented in Table 3. Each domain score is calculated as the sum of item score multiplied by the weight, and the total score is the sum of the 3 domain scores.

Performance of the 15-Item PSPRS Compared With the PSPRS
Comparative performance of the PSPRS and 15-item PSPRS was validated once the PASSPORT study was completed and unblinded. Longitudinal data analyses of change from baseline to week 52 in PSPRS and 15-item PSPRS using MMRM resulted in a mean-to-SD ratio of 1.15 for the PSPRS and 1.22 for the 15-item PSPRS at week 52 for the placebo group (Table 4). For a 2-arm study with 1:1 randomization and 20% dropout rate at week 52, the sample size needed per group for 80% power to detect a treatment effect of 25% slowing vs placebo for change from baseline at week 52 is 240 for the PSPRS and 213 for the 15-item PSPRS, an 11% improvement.

Discussion
We sought to create a modified version of the PSPRS that improved both clinical meaningfulness and sensitivity to detect disease progression and potentially responsiveness to treatment. This article outlines the process and analyses to refine the structure of the PSPRS. The analyses presented demonstrate that the re-scored 15-item PSPRS performed better and successfully improved the psychometric properties of the 28-item PSPRS. The 15-item PSPRS is fit for use as a primary endpoint in a 12-month clinical trial since it (1) measures what matters to patients, (2) has a clear well-defined concept of interest and context

of use, (3) uses a data-driven approach to support that the optimized scale reliably measures motor signs and symptoms of PSP-Richardson syndrome, and (4) has enhanced sensitivity that could reduce the required sample size of a future clinical trial.

This work may potentially have broad applicability to drug development programs in progressive neurological disorders and/or other clinical areas. Valid, reliable, and sensitive COAs do not exist for many diseases. This report provides an example of the approach/process to evaluate, modify, and generate evidence to support the modification of an existing COA, following the FDA's guidance documents for PFDD[8]. Patient-focused drug development is critical to understanding the impact of disease on patients and what they value most in terms of alleviation. This may assist regulators in conducting risk-benefit assessments and drug developers to identify areas of unmet need and the development of outcome measures that support meaningful clinical benefit in medical product labeling.

Per PFDD guidance, patient input should be used in developing instruments to collect data in clinical studies and to identify clinical outcomes that are meaningful to patients. The literature review and qualitative analyses to gather patient perspectives illustrate the diversity and complexity of individuals living with and caring for PSP. The systematic approach using interviews with patients and multiple stakeholders supported the concepts measured by the 15-item PSPRS; patient interviews captured the patient's voice, while caregiver and clinician interviews confirmed aspects of the patient experience that were important from their perspective. Findings from these analyses were inductively categorized in a conceptual framework comprising 3 overall concepts: proximal motor symptoms and functional issues,

non-motor symptoms and functional issues, and the impact of PSP on patients' lives. The conceptual framework was used to understand the disease and conceptualize clinical benefits and risks. The qualitative interviews elicited key unmet needs (gait, vision, cognition) considered by patients and caregivers to be core disease-related symptoms and identified that slowing deterioration of these symptoms are key priorities to inform clinical benefit of a potential therapy.

Although the PSPRS total score captures the spectrum of symptoms and impairments experienced by patients with PSP-Richardson syndrome, some of the items are not meaningful to daily activities or do not contribute to change. Rasch analysis identified that some items may have issues with ambiguity and can impact the cohesiveness of the overall COA. We focused on developing a scale that would capture the motor symptoms and functional issues experienced with PSP. Poor mobility, slowness of movement, and gait difficulty are the most commonly reported symptoms at disease onset[28]. The earliest symptom domains to demonstrate impairment in PSP is generally ocular motor, followed by gait, and then limb motor; cognitive decline is generally a later manifestation of PSP[28]. Previous clinimetric analysis of motor domain of the PSPRS suggested that removal of limb dystonia, tremor, and dysphagia would improve internal consistency[29]. The confirmatory factor analyses supported a 3-factor structure with 3 key motor domains of (1) gait and limb function, (2) ocular motor, and (3) bulbar. Each of the 15 items mapped to one of these 3 motor domains. By focusing on the motor aspects of PSP-Richardson syndrome, the 15-item PSPRS is more fit for use as the selected items are more cohesive and better define the concepts that are clinically relevant and important to patients.

While cognitive impairment is a core feature of PSP and is directly associated with tau burden[30], cognitive items assessed by PSPRS (ie, disorientation, emotional incontinence, withdrawal, irritability, and bradyphrenia) are limited; exclusion of these 5 items improved the sensitivity of the modified PSPRS. To better measure cognition in PSP, multi-domain cognitive scales, such as the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)[31,32] or another cognitive composite scale, may be more specific and sensitive to cognitive changes than the PSPRS. Based on this, we developed a separate COA to more thoroughly capture cognitive symptoms experienced in PSP. This cognitive composite, based on RBANS (only the picture naming is excluded), letter number sequencing test, and phonemic fluency test, is more sensitive than the PSPRS and RBANS as it accounts for ceiling/floor effects and the impact of PSP motor impairments on individual cognitive tests. This PSP cognitive composite was used as a secondary endpoint in the PASSPORT study[27].

By omitting some items that may be confounded by comorbidities or contribute relatively little to disability over a 12-month period given their low prevalence[9] and re-scoring response options into distinct, non-overlapping, and functionally relevant options, our results suggest that the 15-item PSPRS had improved overall performance. Re-scoring response options such that all items were measured on a similar 5-point interval (score range of 0 to 4) addressed issues with over-weighting and ensured that each item equally contributes to the total score. The 15-item PSPRS was able to detect change over a 12-

Table 3 | 15-Item PSPRS total score algorithm

Motor domains	Items	Item names	Weight for total score
Gait and limb function	04	Using knife and fork, buttoning clothes, washing hands and face	1
	05	Falls	1
	19	Limb dystonia	1
	24	Neck rigidity	1
	25	Arising from chair	1
	26	Gait	1
	27	Postural instability	1
	28	Sitting down	1
Ocular motor	14	Voluntary upward saccades	1/3
	15	Voluntary downward saccades	1/3
	16	Voluntary left and right saccades	1/3
	17	Eyelid dysfunction	1
Bulbar	03	Dysphagia for solids	1
	12	Dysarthria	1
	13	Dysphagia for liquids	1

PSPRS Progressive Supranuclear Palsy Rating Scale.

Table 4 | Comparison of 28- and 15-Item PSPRS based on data from the PASSPORT study at week 52

Outcome	Adjusted mean difference (SD) from baseline at week 52, placebo	Mean to SD ratio	Effect size[a]	No. evaluable per arm	Sample size per arm after incorporating 20% dropout[b]
28-Item PSPRS	10.59 (9.19)	1.15	0.287	192	240
15-Item PSPRS	7.57 (6.19)	1.22	0.305	170	213

Adjusted mean (SD) differences are from mixed-model repeated-measures analyses.
[a]Effect size = (treatment effect of 25% slowing of mean placebo change × MSDR).
[b]Sample sizes required for a 2-arm, 52-week duration to detect a treatment effect of 25% slowing vs placebo with 80% power and 1:1 randomization.
MSDR mean-to-SD ratio, PSPRS Progressive supranuclear palsy rating scale.

month period. Variability over a 12-month period was decreased, likely a result of removing items that were contributing to noise, and patient or caregiver improved ability to understand and reliably select response options that reflect the disease over time.

The sensitivity to change on the 15-item PSPRS was higher than that of the PSPRS, and a smaller sample size is required for the 15-item PSPRS. For a 2-arm study with 80% power and a treatment effect of 25%, the sample size per arm would be reduced from 240 to 213, approximately an 11% reduction in sample size by using the 15-item PSPRS. This enhanced sensitivity indicates that these modifications could improve the power of the scale to detect a potential benefit from an anti-tau therapy.

The 15-item PSPRS has shown itself to be superior for use in clinical trials of patients with PSP-Richardson syndrome. However, we recognize that the original PSPRS was designed primarily as a clinical tool to monitor progression in clinical practice, and many of the 13 items removed from the 28-item PSPRS are of potential utility to clinicians in monitoring patients' disabilities as part of routine clinical care. Salient examples are the items on sleep, urinary incontinence, limb rigidity, irritability, disorientation, emotional incontinence, withdrawal, apraxia, and involuntary manual grasping. Thus, the 28-item PSPRS remains an important tool in clinical practice and could continue to be used in clinical trials to enable both the 15-item and 28-item scores to be accessible.

There are some limitations to this study. Because the PASSPORT trial was already in progress at the time of this work, items could not be revised, new items could not be added, response options could not be rewritten or changed, and response options could only be combined/collapsed. Evidence to support the modification of the PSPRS used external datasets. Data from the PASSPORT trial were not used. These analyses were conducted on data from previously enrolled clinical trials (AL-108-231, PROSPERA), a longitudinal natural history study (4RTNI), or longitudinal clinical care observations (Rutgers dataset). As such, results were not biased by looking at PASSPORT data for signs of a response to gosuranemab. These available natural history studies were not diverse in terms of geographic regions and/or underrepresented population; thus, future analyses should be conducted in a more diverse study population.

Additionally, data from these studies were primarily from patients with PSP with 1 clinical PSP syndrome, Richardson syndrome, which was consistent with the population enrolled in the PASSPORT study and contemporary clinical studies because this population can be reliably diagnosed with some certainty. Thus, the 15-item PSPRS and the PSP cognitive composite scale developed for PSP-Richardson syndrome may not be as sensitive in other PSP phenotypes. Recent PSP diagnostic criteria[33,34] were designed with the aim of broadening the definition of PSP to include more people at earlier disease stages in clinical trials. Two recent publications[35,36] have demonstrated the performance of the PSPRS and a modified PSPRS among individuals with 4R tauopathies. First, analyses in the PROSPECT natural history study suggest that the PSPRS and modified PSPRS[20] perform well and are quite sensitive to disease progression in multiple PSP phenotypes, including PSP-Richardson syndrome, PSP-cortical group, and PSP-subcortical group[35]. To advance therapeutic trials in PSP, development of new COAs remains a critical need. Additional work to develop new COAs for PSP that capture a broad spectrum of motor and non-motor symptoms across all PSP phenotypes, particularly for mild severity or early disease, is needed for use in future clinical trials in these PSP subtypes. The recent development of the Cortical Basal Ganglia Functional Scale[37] may be used in combination with the PSPRS to assess motor and non-motor experiences in daily living.

In summary, the 15-item PSPRS is a comprehensive COA that captures clinically meaningful concepts important to patients with PSP-Richardson syndrome and their caregivers. The items and response options included in the instrument are functionally relevant and interpretable to patients, and the scale demonstrated improved sensitivity relative to the original 28-item PSPRS. The use of only 3 motor domains provides the 15-item PSPRS improved interpretability, reliability, and sensitivity. As such, the 15-item PSPRS instrument is fit for use to measure motor signs and symptoms in patients diagnosed with PSP-Richardson syndrome and may be considered for use as a primary endpoint in future clinical trials for PSP-Richardson syndrome.

Methods

The research in this manuscript complies with all relevant ethical regulations. For the PASSPORT study, the study protocol was approved by Advarra's institutional review board (https://www.advarra.com/irb-services/sponsors-cros/) for 6 US sites: Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ; University of Florida Center For Movement Disorders and Neurorestoration, Gainesville, FL; Banner Sun Health Research Institute, Sun City, AZ; University of South Florida–Morsani College of Medicine, Tampa, FL; QUEST Research Institute, Farmington Hills, MI; and St Joseph's Hospital & Medical Center/Barrow Neurology Clinics, Phoenix, AZ. For all other sites, the institutional review board or ethics committee at the institutions listed for the PASSPORT study group investigators approved the study protocol.

Participants and study design

Participant-level PSPRS data from 5 observational and clinical trials were obtained, and data from the first 4 were pooled, including (1) Rutgers clinical data (unpublished), (2) PROSPERA clinical trial data[15], (3) davunetide clinical trial data[10], (4) 4-Repeat Tauopathy Neuroimaging Initiative (4RTNI) natural history study data[38], and (5) PASSPORT clinical trial data[11]. The PASSPORT clinical trial data were used as a separate validation cohort after the study was completed and results were unblinded. Study and recruitment materials were approved by institutional review boards or ethics committees at each site. Written informed consent was obtained from all participants before they underwent any study evaluations. Clinical trials were performed in accordance with the principles outlined in the Declaration of Helsinki and Good Clinical Practice guidelines. The Rutgers study includes clinical practice data from Lawrence Golbe, MD.

PROSPERA (NCT01187888) was a phase 2, multinational, double-blind, randomized, placebo-controlled trial that examined the efficacy and safety of rasagiline 1 mg or placebo (randomized 1:1) administered orally once daily for 52 weeks in 44 individuals with PSP who fulfilled the National Institute of Neurological Disorders and Stroke/Society for PSP (NINDS-PSP) criteria[15].

The davunetide study (NCT01110720) was a phase 2/3, multinational, double-blind, randomized, placebo-controlled trial that examined the efficacy and safety of davunetide 30 mg or placebo (randomized 1:1) administered intranasally twice daily for 52 weeks in 313 individuals with possible or probable PSP[10]. Enrolled individuals met the modified NINDS-PSP criteria defined as at least a 12-month history of postural instability or falls occurring during the first 3 years that symptoms were present, decreased downward saccade velocity or supranuclear ophthalmoplegia, and an akinetic-rigid syndrome with prominent axial rigidity.

The 4RTNI studies (NCT01804452 and NCT02966145) were observational studies to examine the use and reliability of volumetric magnetic resonance imaging in measuring disease progression in 4R tauopathies. Fifty-five individuals with PSP who met the modified NINDS-PSP criteria were enrolled and followed for 1 year[38].

PASSPORT was a phase 2, multinational, double-blind, randomized, placebo-controlled trial that examined the efficacy and safety of gosuranemab 2000 mg or placebo (randomized 2:1) in 486 individuals with possible or probable PSP[11]. Participants were recruited at 90 outpatient specialized movement disorders clinic study sites across 13 countries (sites listed as affiliations for the PASSPORT study group

investigators). Recruitment/screening visits began on April 24, 2017, and the final data collection for the primary outcome occurred on September 6, 2019. The primary efficacy outcome was change from baseline in the PSPRS score at week 52 in participants treated with gosuranemab relative to participants treated with placebo. The primary safety outcomes were frequency of death, serious adverse events, adverse events leading to discontinuation, and grade 3/4 laboratory abnormalities graded by numerical criteria from the Common Terminology Criteria for Adverse Events version 4.0.3. Secondary efficacy endpoints included change from baseline to week 52 on the Movement Disorder Society Unified Parkinson's Disease Rating Scale part II (motor experiences of daily living), Clinical Global Impression (CGI)–Change score at week 52, change from baseline to week 52 on the PSP cognitive composite battery and PSP–Quality of Life scores, change from baseline to week 48 in the Schwab and England Activities of Daily Living scale score, change from baseline to week 52 in CGI-Severity (CGI-S) score, change from baseline to week 48 in a phonemic fluency test, and absolute change from baseline to week 52 in volumes of the total lateral ventricles, whole brain, midbrain, and pons on magnetic resonance imaging scans. Exploratory endpoints included gosuranemab concentrations in blood and cerebrospinal fluid and unbound N-terminal tau concentrations in cerebrospinal fluid.

Sex and/or gender were not considered in the design for this endpoint development. No analyses of sex and/or gender were performed for this manuscript. Sex and/or gender are reported elsewhere. Sex was based on participants self-reported information in the PASSPORT study[10,11,15,38].

Progressive supranuclear palsy rating scale

The PSPRS[9] is a clinician-reported outcome measure comprising 28 items organized into 6 categories. The first category relies on a patient interview about daily activities and symptoms over the previous 30 days. The other 5 categories (mentation, bulbar, ocular motor, limb motor, and gait/midline examination) rely on neurological examination. The PSPRS total score sums all 28 items and ranges from 0 (normal) to 100 points, with higher scores indicating greater severity.

Statistics and reproducibility

The sample sizes for the analyses of the Rutgers clinical data, PROSPERA clinical trial, davunetide clinical trial, and 4RTNI study included data from evaluable patients, who were defined as ambulatory with complete data at baseline and at least one post-baseline visit through year 2. Complete data was defined as all 28 items in the PSPRS recorded at a visit. Ambulatory patients were defined as patients with a baseline PSPRS item 26 gait response of 0, 1, or 2. In the Rutgers dataset, which followed patients in routine clinical care, many patients did not have complete data at a point 12 months after the initial evaluation (eg, patients had data at month 15 or 20 but not at year 1). No statistical method was used to predetermine the sample size. No data were excluded from the analysis. Data from the PROSPERA and davunetide clinical trials are based on randomized experiments and the investigators were blinded to treatment assignment.

Mixed-methods psychometric research was used to improve the performance and interpretability of the PSPRS. Qualitative methods identified concepts important to the lives of patients with PSP. Concepts important to patients with PSP were identified by (1) reviewing published literature and (2) analyzing interview transcripts of patients with PSP, their caregivers, and clinicians who care for patients with PSP. Standard analytical techniques of conceptual framework development were used to categorize data in higher-order overarching categories referred to as concepts, subdomains, and domains[39]. A saturation analysis was conducted at the subdomain level to determine the point at which no new relevant information emerges from additional qualitative data[40]. These data were then used to create a conceptual framework to guide the identification of PSPRS items that were meaningful to patients.

Quantitative methods evaluated the performance and interpretability of the PSPRS. Rasch Measurement Theory (RMT) analyses[41] assessed measurement properties of the PSPRS items, including evaluation of how well the PSPRS fit the intended target population, evaluation of sample measurements, whether the items worked together to define a single measurement construct, item independence, responsiveness to change, and whether the response options worked as intended or required re-scoring. An analysis of item-level responsiveness to change over 12 months was used to identify sensitive items to retain in the modified PSPRS. For these analyses, data from 402 ambulatory participants with complete data through year 2 were analyzed (244 participants in the pooled PROSPERA, davunetide, and 4RTNI dataset and 158 participants in the Rutgers dataset).

To understand the underlying structure of the selected items and develop a scoring algorithm, confirmatory factor analysis (CFA) was conducted. A 3-factor structure was hypothesized for the 15-item PSPRS and was tested using function CFA in R package lavaan. For CFA analyses, data from 620 participants with baseline data were analyzed. Validity of the factor structure was assessed using the following model-fit statistics and criteria: comparative fit index ≥0.95, Tucker-Lewis index (TFI) ≥ 0.95, standardized root-mean squared residual (SRMR) ≤ 0.08, and root-mean-squared error of approximation (RMSEA) ≤ 0.08[42]. Comparative fit and TFI values closer to 1 indicate better fit (values range from 0 to 1); SRMR and RMSEA values between 0.05 and 0.08 indicate acceptable fit, with values < 0.05 indicating good fit. Multiple weighted and unweighted scoring algorithms were compared to determine the best total score algorithm.

To compare the sensitivity of PSPRS with that of the 15-item PSPRS, we analyzed changes from baseline to week 52 in the placebo groups from the PASSPORT study using a mixed-model repeated measures (MMRM) approach with fixed effects of treatment group, time, treatment group-by-time interaction, baseline value, baseline-value-by-time interaction, region (US or non-US), and baseline Color Trails 2 test (≤ 170 or >170 s; a measure of frontal cognitive and visual function). The mean-to-SD ratio was calculated using (absolute value of adjusted mean change for the placebo group)/(SD from the variance-covariance matrix from the MMRM model).

SAS version 9.4 was used for data preparation, descriptive statistics, and statistical modeling. RUMM2030 was used for Rasch analyses. R was used for CFA. Reproducibility of the experimental findings was done via analytical replication. Analyses were produced by a primary statistical programmer or statistician. These analyses were reproduced by an independent statistical programmer or another statistician. All attempts at analytical replication were successful.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

For the PASSPORT study, the Biogen Clinical Trial Transparency and Data Sharing Policy is available at https://clinicalresearch.biogen.com/. Individual participant data collected during the trial that support the research proposal may be shared after anonymization and upon approval of the research proposal. Biogen commits to share patient-level data, study-level data, clinical study reports, and protocols with qualified scientific researchers who provide a methodologically sound proposal. Data requests are initially reviewed by Vivli and Biogen for completeness and other parameters (relating to scope and meeting Biogen's policies) and are then reviewed by an independent review panel. Deidentified data and documents will be shared under agreements that further protect against participant reidentification. To request access to data, please visit https://vivli.org/. The 4RTNI dataset

can be freely downloaded from the Laboratory of Neuroimaging Imaging and Data Archive at https://ida.loni.usc.edu/login.jsp. The davunetide and PROSPERA datasets may be obtained through a request to Dr Adam Boxer (adam.boxer@ucsf.edu) and will require a data use agreement with UCSF. For the Rutgers dataset, please visit https://rarediseases.org/rdca-dap/.

References

1. Boxer, A. L. et al. Advances in progressive supranuclear palsy: new diagnostic criteria, biomarkers, and therapeutic approaches. *Lancet Neurol.* **16**, 552–563 (2017).
2. Nath, U., Ben-Shlomo, Y., Thomson, R. G., Lees, A. J. & Burn, D. J. Clinical features and natural history of progressive supranuclear palsy: a clinical cohort study. *Neurology* **60**, 910–916 (2003).
3. Williams, D. R. et al. Characteristics of two distinct clinical phenotypes in pathologically proven progressive supranuclear palsy: Richardson's syndrome and PSP-parkinsonism. *Brain* **128**, 1247–1258 (2005).
4. Nath, U. et al. The prevalence of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome) in the UK. *Brain* **124**, 1438–1449 (2001).
5. Schrag, A., Ben-Shlomo, Y. & Quinn, N. P. Prevalence of progressive supranuclear palsy and multiple system atrophy: a cross-sectional study. *Lancet* **354**, 1771–1775 (1999).
6. Dickson, D. W., Ahmed, Z., Algom, A. A., Tsuboi, Y. & Josephs, K. A. Neuropathology of variants of progressive supranuclear palsy. *Curr. Opin. Neurol.* **23**, 394–400 (2010).
7. VandeVrede, L., Boxer, A. L. & Polydoro, M. Targeting tau: Clinical trials and novel therapeutic approaches. *Neurosci. Lett.* **731**, 134919 (2020).
8. *FDA Patient-Focused Drug Development Guidance Series for Enhancing the Incorporation of the Patient's Voice in Medical Product Development and Regulatory Decision Making*, <https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical> (2024).
9. Golbe, L. I. & Ohman-Strickland, P. A. A clinical rating scale for progressive supranuclear palsy. *Brain* **130**, 1552–1565 (2007).
10. Boxer, A. L. et al. Davunetide in patients with progressive supranuclear palsy: a randomised, double-blind, placebo-controlled phase 2/3 trial. *Lancet Neurol.* **13**, 676–685 (2014).
11. Dam, T. et al. Safety and efficacy of anti-tau monoclonal antibody gosuranemab in progressive supranuclear palsy: a phase 2, randomized, placebo-controlled trial. *Nat. Med* **27**, 1451–1457 (2021).
12. Höglinger, G. U. et al. Safety and efficacy of tilavonemab in progressive supranuclear palsy: a phase 2, randomised, placebo-controlled trial. *Lancet Neurol.* **20**, 182–192 (2021).
13. Apetauerova, D. et al. CoQ10 in progressive supranuclear palsy: A randomized, placebo-controlled, double-blind trial. *Neurol. Neuroimmunol. Neuroinflamm* **3**, e266 (2016).
14. Leclair-Visonneau, L. et al. Randomized placebo-controlled trial of sodium valproate in progressive supranuclear palsy. *Clin. Neurol. Neurosurg.* **146**, 35–39 (2016).
15. Nuebling, G. et al. PROSPERA: a randomized, controlled trial evaluating rasagiline in progressive supranuclear palsy. *J. Neurol.* **263**, 1565–1574 (2016).
16. Tolosa, E. et al. A phase 2 trial of the GSK-3 inhibitor tideglusib in progressive supranuclear palsy. *Mov. Disord.* **29**, 470–478 (2014).
17. Bang, J. et al. Predicting disease progression in progressive supranuclear palsy in multicenter clinical trials. *Parkinsonism Relat. Disord.* **28**, 41–48 (2016).
18. Stamelou, M. et al. Power calculations and placebo effect for future clinical trials in progressive supranuclear palsy. *Mov. Disord.* **31**, 742–747 (2016).
19. Golbe, L. I., Ohman-Strickland, P., Beisser, E. B. & Elghoul, F. T. A convenient prognostic tool and staging system for progressive supranuclear palsy. *Mov. Disord. Clin. Pr.* **7**, 664–671 (2020).
20. Grötsch, M. T. et al. A modified progressive supranuclear palsy rating scale. *Mov. Disord.* **36**, 1203–1215 (2021).
21. Piot, I. et al. The progressive supranuclear palsy clinical deficits scale. *Mov. Disord.* **35**, 650–661 (2020).
22. Wills, A. M. et al. A modified progressive supranuclear palsy rating scale for virtual assessments. *Mov. Disord.* **37**, 1265–1271 (2022).
23. Schrag, A. et al. Health-related quality of life in patients with progressive supranuclear palsy. *Mov. Disord.* **18**, 1464–1469 (2003).
24. Wiblin, L., Durcan, R., Lee, M. & Brittain, K. The importance of connection to others in QoL in MSA and PSP. *Parkinsons Dis.* **2017**, 5283259 (2017).
25. Walsh, C. M. et al. Rest-activity rhythm disruption in progressive supranuclear palsy. *Sleep. Med* **22**, 50–56 (2016).
26. Walsh, C. M. et al. Sleepless night and day, the plight of progressive supranuclear palsy. *Sleep* **40** https://doi.org/10.1093/sleep/zsx154 (2017).
27. Jaeger, J. et al. Development of a cognitive composite for measuring change in progressive supranuclear palsy. *Parkinsonism Relat. Disord.* **92**, 94–100 (2021).
28. Brittain, C. et al. Severity dependent distribution of impairments in PSP and CBS: Interactive visualizations. *Parkinsonism Relat. Disord.* **60**, 138–145 (2019).
29. Hall, D. A. et al. Clinimetric analysis of the motor section of the progressive supranuclear palsy rating scale: Reliability and factor analysis. *Mov. Disord. Clin. Pr.* **3**, 65–67 (2016).
30. Koga, S. et al. Distribution and characteristics of transactive response DNA binding protein 43 kDa pathology in progressive supranuclear palsy. *Mov. Disord.* **32**, 246–255 (2017).
31. Brown, R. G. et al. Cognitive impairment in patients with multiple system atrophy and progressive supranuclear palsy. *Brain* **133**, 2382–2393 (2010).
32. Duff, K., McDermott, D., Luong, D., Randolph, C. & Boxer, A. L. Cognitive deficits in progressive supranuclear palsy on the repeatable battery for the assessment of neuropsychological status. *J. Clin. Exp. Neuropsychol.* **41**, 469–475 (2019).
33. Höglinger, G. U. et al. Clinical diagnosis of progressive supranuclear palsy: The movement disorder society criteria. *Mov. Disord.* **32**, 853–864 (2017).
34. Jabbari, E. et al. Diagnosis across the spectrum of progressive supranuclear palsy and corticobasal syndrome. *JAMA Neurol.* **77**, 377–387 (2020).
35. Street, D. et al. Progression of atypical parkinsonian syndromes: PROSPECT-M-UK study implications for clinical trials. *Brain* **146**, 3232–3242 (2023).
36. VandeVrede, L. et al. Evaluation of plasma phosphorylated tau217 for differentiation between alzheimer disease and frontotemporal lobar degeneration subtypes among patients with corticobasal syndrome. *JAMA Neurol.* **80**, 495–505 (2023).
37. Lang, A. E. et al. The Cortical basal ganglia functional scale (CBFS): Development and preliminary validation. *Parkinsonism Relat. Disord.* **79**, 121–126 (2020).
38. Dutt, S. et al. Progression of brain atrophy in PSP and CBS over 6 months and 1 year. *Neurology* **87**, 2016–2025 (2016).
39. Thomas, D. R. A general inductive approach for analyzing qualitative evaluation data. *Am. J. Eval.* **27**, 237–246 (2006).
40. Morse, J. M. The significance of saturation. *Qual. Health Res* **5**, 147–149 (1995).
41. Andrich, D. *Rasch Models for Measurement*. (Sage Publications, New York City, 1988)
42. Brown, T. A. *Confirmatory factor analysis for applied research*. (The Guilford Press, New York City, 2006).

Author contributions
T.D., L.Y., and J.O. had full access to all data and take responsibility for the integrity of the data and the accuracy of the data analysis. *Acquisition, analysis, or interpretation of data*: T.D., L.Y., C.G., Y.L., J.O., B.K., A.B., L.G., S.B.H and AL-108-231, PROSPERA, and 4RTNI investigators. *Drafting of the manuscript*: T.D and J.O. *Statistical analysis*: L.Y., Y.L., and J.O.

Competing interests
Tien Dam, Lili Yang, Colin Gillis, Yumeng Li, Barbara Kolb, and Samantha Budd Haeberlein are former employees of and may hold stock/stock options in Biogen. John O'Gorman is an employee of and may hold stock in Biogen. Adam Boxer receives research support from the National Institutes of Health, Tau Research Consortium, Association for Frontotemporal Degeneration, Bluefield Project to Cure Frontotemporal Dementia, Alzheimer's Drug Discovery Foundation, Alzheimer's Association, Biogen, Eisai, and Regeneron and has served as a consultant for AGTC, Alector, Arkuda, Arvinas, AviadoBio, AZTherapeutics, Boehringer Ingelheim, GSK, Humana, Life Edit Therapeutics, Oligomerix, Roche, True-Binding, and Wave. Lawrence Golbe has received research support from Biogen, travel support from CurePSP, and Consultant fees from AI Therapeutics, Amylyx, Apellis, Aprinoia, Ferrer, Mitochon, Roche, and UCB. These analyses were funded by Biogen (Cambridge, MA, USA). The 4 Repeat Tauopathy Imaging Initiative is funded by National Institutes of Health grant R01AG038791.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-55442-0.

Correspondence and requests for materials should be addressed to John O'Gorman.

Peer review information *Nature Communications* thanks Timothy Rittman and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

the AL-108-231, PROSPERA, and 4RTNI investigators

Adam L. Boxer ©[2] & Lawrence Golbe[3]