RESEARCH ARTICLE

The Progressive Supranuclear Palsy Clinical Deficits Scale

Ines Piot, MD,[1,2,3] Kerstin Schweyer, MD,[1,2] Gesine Respondek, MD,[1,2,8] Maria Stamelou, MD, PhD,[4,5] DescribePSP study group, ProPSP study group, MDS-endorsed PSP study group, Philipp Sckopke, Dr.,[6] Thomas Schenk, Dr.,[6] Christopher G. Goetz, MD,[7] Glenn T. Stebbins, PhD,[7] and Günter U. Höglinger, MD[1,2,8*]

[1]*German Center for Neurodegenerative Diseases (DZNE), Munich, Germany*
[2]*Department of Neurology, Technical University of Munich, Munich, Germany*
[3]*Department of Neurology, University of Basel, Basel, Switzerland*
[4]*Parkinson's disease and Movement Disorders Department, HYGEIA Hospital and First Department of Neurology, Aiginiteion Hospital, National and Kapodistrian University of Athens, Athens, Greece*
[5]*Neurology Clinic, Philipps University, Marburg, Germany*
[6]*Department of Psychology, Ludwig-Maximilians-University, Munich, Germany*
[7]*Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois, USA*
[8]*Department of Neurology, Hannover Medical School, Hannover, Germany*

ABSTRACT: Background: There is currently no undisputed, validated, clinically meaningful measure for deficits in the broad spectrum of PSP phenotypes.
Objective: To develop a scale to monitor clinical deficits in patients with PSP across its broad phenotypes.
Methods: The Progressive Supranuclear Palsy Clinical Deficits Scale was conceptualized to cover seven clinical domains (**A**kinesia-rigidity, **B**radyphrenia, **C**ommunication, **D**ysphagia, **E**ye movements, **F**inger dexterity, and **G**ait & balance), each scored from 0 to 3 (no, mild, moderate, or severe deficits). User guidelines were developed to standardize its application. Progressive Supranuclear Palsy Clinical Deficits Scale scores were collected in patients fulfilling the MDS-PSP diagnostic criteria in two independent, multicenter, observational studies, both cross-sectionally (exploratory DescribePSP cohort; confirmatory ProPSP cohort) and longitudinally (12-months' follow-up, both cohorts).
Results: Cognitive pretesting demonstrated easy scale utility. In total, 164 patients were scored (70.4 ± 7.6 years; 62% males, 35% variant phenotypes). Mean Progressive Supranuclear Palsy Clinical Deficits Scale completion time was 4 minutes. The Progressive Supranuclear Palsy Clinical Deficits Scale total score correlated with existing scales (e.g., Progressive Supranuclear Palsy Rating Scale: $R = 0.88$; $P < 0.001$). Individual Progressive Supranuclear Palsy Clinical Deficits Scale items correlated well with similar constructs in existing scales. Internal consistency (Cronbach's alpha: 0.75), inter-rater reliability (0.96), and test-retest stability (0.99) were acceptable. The PSP-CDS showed significant 12-month change (baseline, 8.6 ± 3.6; follow-up: 10.8 ± 3.6; annualized difference: 3.4 ± 3.4; $n = 49$; $P < 0.0001$). Sample sizes required per arm for a two-arm, 1-year follow-up therapeutic trial to detect 50% change in Progressive Supranuclear Palsy Clinical Deficits Scale progression was estimated to be 65 (two-sided, two-sample *t* test).
Conclusion: The Progressive Supranuclear Palsy Clinical Deficits Scale is a rapidly completed, clinimetrically sound scale for clinical care and research involving PSP.

Key Words: progressive supranuclear palsy; clinical rating scales; outcome measures; power calculation

***Correspondence to:** Prof. Dr. Günter U. Höglinger, Department of Neurology, Hannover Medical School, Carl-Neuberg-Straße 1, D-30625 Hannover, Germany; E-mail: guenter.hoeglinger@dzne.de.

Ms. Piot and Dr. Schweyer made an equal contribution.

Funding agencies: The project was supported by the German Center for Neurodegenerative Diseases e.V. (DZNE) and the German Parkinson Society (DPG).

Relevant conflicts of interest/financial disclosures: Nothing to report.

Full financial disclosures and author roles may be found in the online version of this article.

Received: 19 August 2019; **Revised:** 4 December 2019; **Accepted:** 6 December 2019

Published online 00 Month 2020 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.27964

PSP is a four-repeat tauopathy presenting with symptoms including ocular motor dysfunction, postural instability, akinesia/rigidity, frontal cognitive/behavioral dysfunction, speech/language dysfunction, and bulbar symptoms (dysarthria/dysphagia).[1,2] The most frequent clinical manifestation of PSP is Richardson's syndrome (PSP-RS), but a broad spectrum of variant PSP manifestations (vPSP) has been acknowledged and operationalized in the International Parkinson and Movement Disorder Society (MDS) diagnostic criteria.[2] There are currently no effective symptomatic or disease-modifying therapies available, but tau-directed clinical trials are ongoing and in preparation.[1,3]

Rating scales are fundamental in clinical practice and research to assess disease severity cross-sectionally and progression or treatment response longitudinally.[4] Scales as outcome measures for clinical trials have been increasingly criticized, given that traditional scales often prove to be inadequate, producing, at best, circumstantial evidence suggesting therapeutic efficacy of investigational drugs.[4] Both the European Medicines Agency (EMA)[5] and the U.S. Food and Drug Administration (FDA)[6-8] have issued guidelines for scales in clinical trials. Key points of the EMA recommendations on outcome assessments are external validation, reliability, sensitivity to change, and the focus on "functional and global domains with greater emphasis on activities of daily living (ADL)," capturing "clinically meaningful benefits."[5] The FDA advocates the importance of "patient-focused clinical trial endpoints," assessing how a patient "feels, functions and survives."[8] Quality-of-life scales as primary outcome measures are controversial, because they are prone to "response shifts" attributable to social, psychological, or emotional factors, which may be confused with truly neurological treatment effects.[5] Patient-reported outcome measures are being recommended,[6] but may be of limited value in patients with cognitive and frontal lobe dysfunction impairing insight into existing deficits.[2,9] Thus, both EMA and FDA recommend scales focusing on patient-focused deficits in the clinical domains affected by the disease under investigation.

Currently, several generic scales are being used to monitor the status and change of PSP patients' health:

- The UPDRS Part III (MDS-UPDRS III; 14 items)[10] has been developed to assess motor deficits in Parkinson's disease. It is also recommended for this purpose in PSP.[11,12]
- The Frontal Assessment Battery (FAB; 6 items)[13] evaluates frontal lobar behavioral and cognitive functions. In PSP, the FAB identifies cognitive deficits,[14,15] but is not sensitive to change.[16,17]
- The Montreal Cognitive Assessment (MoCA; 11 items)[18] is also used to identify cognitive deficits,[19] but longitudinal data have not yet been reported in PSP.
- The Schwab and England Activities of Daily Living Scale (SEADL; 1 item)[20] briefly evaluates functional dependence in ADL, showing good sensitivity to change in PSP.[16]
- The Clinical Global Impression of Illness Severity Scale (CGI-S; 1 item)[21] assesses global illness levels relative to the raters' general experience; it has been applied in PSP and also shows sensitivity to change.[16,22]

There are only two established disease-specific scales for PSP:

- The PSP-Quality of Life Scale (PSP-QoL; 46 items)[23] is a disease-specific, patient-rated quality of life, but longitudinal data have not been published so far.
- The PSP Rating Scale (PSPRS; 28 items)[24] is a physician-rated measure of disease severity. It remains the only prospectively validated, disease-specific scale for PSP, yielding reproducible annual progression rates with reasonable effect sizes in multicenter settings.[16,25,26] Still, concerns have been voiced: The disease concept for its construction was PSP-RS. Features of vPSP do occur in the PSPRS, but contribute to a lesser extent to the total score than PSP-RS hallmark features. In some items, response categories may not be objectively distinguishable. Score changes may not proportionally imply clinically meaningful changes, because some items contribute higher scores than others, because some functional domains are represented by more items than others, and because some items may be functionally more relevant than others. There are no inter-rater reliability or test-retest data available so far. Finally, the PSPRS is a time-consuming specialists' tool and is not very handy for clinical routine. The PSP Staging System (PSPSS), which comes with the PSPRS, classifies disease severity based on different grades of postural instability and gait only.

Power calculations on longitudinal data demonstrated the PSPRS to be the most reliable single progression measure, requiring the least number of patients (49/arm) to detect a 50% change in disease progression within 1 year, followed by the SEADL (70/arm) and CGI-S (81/arm).[16]

Acknowledging the emerging need of a scale assessing meaningful deficits in the clinical domains relevant for PSP, being equally applicable to PSP-RS and vPSP-phenotypes, yielding predictable annual progression rates, and being usable in clinical care and multicenter research settings, the MDS-endorsed PSP study group set out to develop the physician-reported PSP–Clinical Deficits Scale (PSP-CDS).

Materials and Methods

Scale Development

Members of the MDS-endorsed PSP study group, with input from members of the MDS Rating Scales

Program (C.G.G., G.T.S.), conceptualized the PSP-CDS. To identify the clinical domains most pertinent to PSP, we reanalyzed the systematic literature review conducted for the purpose of generating the MDS-PSP diagnostic criteria.[27] In brief, literature was searched on PubMed, Cochrane, Medline, and PSYCInfo databases for entries from 1996 until 2015, using search terms for PSP and for the question "Which signs, symptoms, or syndromes are present in neuropathologically defined PSP?"[27] Research articles, systematic reviews, and meta-analyses in English language, using postmortem analysis or the National Institute of Neurological Disorders and Stroke and Society for Progressive Supranuclear Palsy (NINDS-SPSP) criteria as a diagnostic standard, were selected to identify characteristic PSP-associated clinical deficits with relevance to ADL. The members of the study group reviewed and annotated the literature and the principal component analyses. Based on their written summary report, a subgroup (I.P., K.S., G.R., M.S., and G.U.H.) developed the scale by agreeing on the clinical domains to be included and the response categories in an iterative Delphi-like approach achieving final consensus. Domains with presumed impact on instrumental ADL (preparing meals, shopping, telecommunication, and managing finances) and self-care ADL (hygiene, dressing, self-feeding, using toilets, managing medications, and not being bedridden) were prioritized. For all domains, four response categories were generated, defining deficit degrees (0 = no deficit; 1 = mild deficits considered by the scale developers not to affect ADL; 2 = moderate deficits considered to necessitate partial external support to maintain ADL; 3 = severe deficits considered to incapacitate ADLs requiring continuous external support). The scale was developed and validated in the English language.

Cognitive Pretesting

Cognitive pretesting of the PSP-CDS was conducted with the aim to identify aspects that required improvement in comprehensibility and applicability, using a standardized questionnaire based on two techniques[28-30]:

- Think-aloud: To reveal possible sources for difficulties in decision making, raters reported their (dis) agreement with individual items/response categories while applying the PSP-CDS.
- Verbal probing: After choosing a response for an item, raters were asked to paraphrase their answer, to report the reason for selecting specific responses, and assess their ease of decision (1 = very easy; 2 = easy; 3 = difficult; 4 = very difficult).

Patients and Assessments

Cross-sectional and longitudinal data were collected within two German multicenter, observational cohort studies of PSP patients. The DescribePSP study, run by the German Center for Neurodegenerative Diseases (DZNE), served to generate an exploratory data set. The ProPSP study, run by the German Parkinson and Movement Disorders Society (DPG), generated an independent confirmatory sample. Independent patients, regardless of PSP phenotype, were consecutively recruited into both cohorts since 2017 at multiple centers (14 ProPSP, 8 DescribePSP) with expertise in movement disorders, geographically distributed across the country. Patients were assessed in the German language. Scales requiring active participation of the patient (e.g., MoCA) were provided in the German language. Physician-reported outcomes were used in the English language. Inclusion and exclusion criteria for patients were the MDS criteria for clinical diagnosis of PSP.[2] Information required for PSP-CDS rating was collected by individually trained neurologists through a semistructured interview, with or without help by a reliable caregiver, and a short structured clinical examination, at baseline and 12-months' follow-up examination. For evaluation of inter-rater reliability, a subset of patients was assessed by two independent raters. For test-retest stability, patients were assessed by the same rater twice (within 1–14 days' follow-up). To avoid a bias by disease progression or diurnal fluctuations, both examinations were performed, at most, 14 days apart and at the same time of the day. Additionally, the following data were collected at baseline: demographic information, PSPRS, PSPSS, PSP-QoL, MDS-UPDRS III, SEADL, MoCA, and CGI-S. Data from DescribePSP and ProPSP were collected by two independent web-based data capture systems. Both cohort studies and PSP-CDS development were approved by ethical committees at all participating centers.

Statistical Analysis

All statistics were conducted using GraphPad Prism software (version 8.0.1. for MacOS; GraphPad Software Inc., San Diego, CA). Descriptive data are reported as mean, standard deviation (SD), and range.

Frequency distribution of baseline scores and disease duration served to assess floor and ceiling effects, considered as present if >15% of patients achieved worst or best scale score.[53] The nonparametric Spearman's rank coefficient was calculated for correlation analysis. To relate PSP-CDS progression to clinically relevant disease, we operationalized milestones based on PSPRS items (3, 5, and 26) and MoCA scores and analyzed percentages of patients having reached such milestones by PSP-CDS tertiles (operationalizing mild, moderate, and severe disease stages). Item-total correlations and Cronbach's alpha (CA) were used as a measure for internal consistency. Inter-rater reliability and test-retest

stability were calculated using intraclass correlations. Power calculations were performed on patients with 12-month follow-up data, calculating the annualized change versus baseline, standardized effect sizes, and estimated sample sizes needed to detect 30% or 50% changes in annualized progression (80% power; two-sided, two-sample *t* test and Mann-Whitney U test).

Short Scale Versions

The full PSP-CDS addresses deficits in seven clinical domains with three degrees of severity each. The full version thus may also be referred to as the PSP-CDS.^{7x3} After several feedback rounds, we decided to develop two shorter versions of the scale, one omitting the first response category (mild deficits) and one without the item "eye movements": Given that the response categories 1 (= mild deficit) of the PSP-CDS^{7x3} purposefully identify clinical deficits without functional relevance for patients, we devised an abbreviated PSP-CDS^{7x2} scale for use in research settings, where such items are of no interest (merging categories 0 and 1 of PSP-CDS^{7x3}; Supporting Information S8). While we believe in the strong impact of the item "Eye movements" on ADL in PSP (affecting the ability to look down to the plate while eating, to toes while descending stairs, and to obstacles while walking), the item's functional relevance is sometimes controversially debated. Therefore, we constructed a PSP-CDS^{6x2} (omitting "Eye movements" from PSP-CDS^{7x2}; Supporting Information S9). We performed basis statistics on both shorter scales to confirm their usability as well.

Results

Scale Conceptualization

The literature review identified 103 relevant studies, 37 of which were considered particularly applicable to the current project. For two independent clinicopathological series of autopsy-confirmed PSP patients, a principal components analysis of clinical features, extracted from the patients' charts, had been performed.[31,32] Both analyses identified supranuclear gaze palsy and related ocular motor dysfunctions, postural instability leading to gait and balance problems, akinesia-rigidity, and related parkinsonian features as key factors.[31,32] One of these studies additionally identified cognitive dysfunction as a key element.[31] More specifically, frontal cognitive/behavioral[33-35] and speech/language[36-38]

Functional Domain	0 = No Deficit	1 = Mild Deficit	2 = Moderate Deficit	3 = Severe Deficit	Score
Akinesia-rigidity	No akinesia or rigidity	Slow movements, but full range possible	Reduced range in active movements	Reduced range in passive movements	
Bradyphrenia	No bradyphrenia	Equivocal or mild, but not interfering with activities of daily living	Interfering moderately with activities of daily living	Interfering severely with activities of daily living	
Communication	No communicative dysfunction	Mild communicative dysfunction, but easily understood	Moderate communicative dysfunction, partly not understood	Severe communicative dysfunction, cannot be understood	
Dysphagia	No dysphagia	Mild dysphagia, but no dietary adaptations required	Moderate dysphagia, dietary adaptations required	Severe dysphagia , oral nutrition impossible	
Eye movements	No ocular motor dysfunction	Slow vertical saccades	Vertical supranuclear gaze palsy	Vertical & horizontal supranuclear gaze palsy	
Finger dexterity	No impairment in finger dexterity	Somewhat slow, but no help required when using knife and fork, buttoning clothes, washing hands and face.	Extremely slow or occasional help required	Considerable help or total assistance needed	
Gait & balance	No postural instability	Postural instability, but unassisted gait possible	Gait possible with walking aid	Gait impossible	
				Total Score	

FIG. 1. PSP-CDS Scale scoring table. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1. Demographic and clinical data of the study population

PSP Phenotypes		Exploratory Analysis (DescribePSP) All (n = 88)	Exploratory Analysis (DescribePSP) PSP-RS (n = 55)	Exploratory Analysis (DescribePSP) Variants (n = 33)	Confirmatory Analysis (ProPSP) All (n = 76)	Confirmatory Analysis (ProPSP) PSP-RS (n = 52)	Confirmatory Analysis (ProPSP) Variants (n = 24)	Joint Analysis All (n = 164)	Joint Analysis PSP-RS (n = 107)	Joint Analysis Variants (n = 57)
Sex (m/f), N (%)		60/28 (68/32)	35/20 (64/36)	25/8 (76/24)	40/34 (54/46)	31/20 (61/39)	9/14 (39/61)	100/62 (62/38)	66/40 (62/38)	34/22 (61/39)
Age at symptom onset (y)		67.2 ± 7.8 (43–82)	68.5 ± 6.3 (52–80)	65.1 ± 9.5 (43–82)	66.2 ± 7.7 (50–81)	66.1 ± 6.8 (50–81)	66.4 ± 9.5 (50–81)	66.8 ± 7.7 (43–82)	67.4 ± 6.6 (50–81)	65.7 ± 9.4 (43–82)
Age at diagnosis (y)		69.2 ± 7.7 (45–82)	70.2 ± 6.6 (56–82)	67.4 ± 9.2 (45–82)	68.9 ± 8.0 (50–82)	68.6 ± 7.3 (52–82)	69.6 ± 9.6 (50–82)	69.1 ± 7.8 (45–82)	69.5 ± 6.9 (52–82)	68.3 ± 9.3 (45–82)
Age at examination (y)		70.0 ± 7.8 (44–83)	71.1 ± 6.3 (56–82)	68.5 ± 9.4 (44–83)	70.6 ± 7.5 (53–85)	70.7 ± 7.0 (54–85)	70.3 ± 8.8 (53–81)	70.4 ± 7.6 (44–85)	70.9 ± 6.6 (54–85)	69.3 ± 9.1 (44–83)
Disease duration (y)		3.0 ± 2.1 (0–10)	2.9 ± 2.0 (0–8)	3.3 ± 2.1 (0–10)	4.0 ± 2.9 (0–14)	4.1 ± 2.5 (0–11)	3.9 ± 3.7 (0–14)	3.5 ± 2.5 (0–14)	3.4 ± 2.3 (0–11)	3.6 ± 2.9 (0–14)
SEADL		60.5 ± 22.7 (10–90)	62.4 ± 21.2 (10–90)	52.3 ± 24.9 (10–90)	47.6 ± 23.2 (10–90)	40.4 ± 21.3 (10–80)	63.8 ± 20.2 (20–90)	54.5 ± 23.7 (10–90)	51.7 ± 23.8 (10–90)	60.0 ± 23.1 (10–90)
UPDRS III		38.6 ± 18 (6–105)	36.2 ± 15.4 (9–69)	42.6 ± 21.3 (6–105)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
CGI-S		4.2 ± 1.0 (2–7)	4.1 ± 0.9 (2–7)	4.4 ± 1.2 (2–6)	4.7 ± 1.0 (2–7)	4.9 ± 1.0 (3–7)	4.3 ± 1.1 (2–6)	4.4 ± 1.1 (2–7)	4.5 ± 1.0 (2–7)	4.3 ± 1.1 (2–6)
MoCA		21.0 ± 6.2 (3–30)	21.3 ± 5.4 (6–29)	20.5 ± 7.4 (3–30)	19.1 ± 7.9 (0–29)	19.1 ± 7.6 (0–28)	19.1 ± 8.8 (0–29)	20.1 ± 7.1 (0–30)	20.2 ± 6.7 (0–29)	19.8 ± 8.0 (0–30)
PSP-QoL		34.8 ± 17.3 (3.3–65.9)	37.0 ± 17.0 (3.3–65.9)	31.3 ± 17.6 (9.4–60.8)	40.3 ± 19.5 (3.8–98.9)	42.4 ± 20.5 (3.8–98.9)	35.7 ± 16.7 (9.4–67.4)	37.9 ± 18.7 (3.3–98.9)	40.2 ± 19.2 (3.3–98.9)	33.7 ± 17.1 (9.4–67.4)
PSPSS		2.9 ± 1.1 (1–5)	2.9 ± 1.1 (1–5)	2.9 ± 1.2 (1–5)	3.1 ± 1.0 (1–5)	3.4 ± 0.9 (1–5)	2.5 ± 1.0 (1–4)	3.0 ± 1.1 (1–5)	3.2 ± 1.0 (1–5)	2.7 ± 1.1 (1–5)
PSPRS		33.4 ± 12.4 (11–71)	32.2 ± 11.4 (15–71)	33.8 ± 14.1 (11–65)	37.6 ± 15.5 (9–75)	43.5 ± 14.3 (20–75)	24.8 ± 9.1 (9–45)	35.4 ± 14.0 (9–75)	38.3 ± 13.8 (15–75)	30.0 ± 12.9 (9–65)
PSP-CDS	Total	8.9 ± 3.2 (3–17)	9.0 ± 3.1 (4–17)	8.9 ± 3.5 (3–16)	10.3 ± 4.3 (2–21)	11.8 ± 4.0 (4–21)	7.0 ± 2.6 (2–14)	9.6 ± 3.8 (2–21)	10.4 ± 3.8 (4–21)	8.1 ± 3.3 (2–16)
	A	1.4 ± 0.7 (0–3)	1.3 ± 0.7 (0–3)	1.4 ± 0.8 (0–3)	1.5 ± 0.7 (0–3)	1.6 ± 0.8 (0–3)	1.3 ± 0.5 (0–2)	1.4 ± 0.7 (0–3)	1.5 ± 0.8 (0–3)	1.4 ± 0.7 (0–3)
	B	1.2 ± 0.8 (0–3)	1.1 ± 0.8 (0–3)	1.3 ± 0.9 (0–3)	1.3 ± 0.9 (0–3)	1.5 ± 0.9 (0–3)	0.8 ± 0.8 (0–3)	1.2 ± 0.9 (0–3)	1.3 ± 0.9 (0–3)	1.1 ± 0.9 (0–3)
	C	1.3 ± 0.6 (0–3)	1.2 ± 0.7 (0–3)	1.3 ± 0.6 (0–3)	1.4 ± 0.9 (0–3)	1.5 ± 0.8 (0–3)	1.2 ± 1.0 (0–3)	1.3 ± 0.8 (0–3)	1.3 ± 0.7 (0–3)	1.2 ± 0.8 (0–3)
	D	0.8 ± 0.7 (0–3)	0.7 ± 0.7 (0–3)	0.8 ± 0.8 (0–2)	0.9 ± 0.8 (0–3)	1.2 ± 0.8 (0–3)	0.5 ± 0.6 (0–2)	0.8 ± 0.8 (0–3)	0.9 ± 0.8 (0–3)	0.7 ± 0.7 (0–2)
	E	1.7 ± 0.9 (0–3)	1.9 ± 0.7 (0–3)	1.3 ± 1.0 (0–3)	1.9 ± 1.0 (0–3)	2.3 ± 0.7 (1–3)	1.2 ± 1.2 (0–3)	1.8 ± 0.9 (0–3)	2.1 ± 0.7 (0–3)	1.3 ± 1.0 (0–3)
	F	1.5 ± 1.0 (0–3)	1.5 ± 1.0 (0–3)	1.6 ± 1.0 (0–3)	1.7 ± 1.1 (0–3)	2.0 ± 1.0 (0–3)	1.1 ± 1.0 (0–3)	1.6 ± 1.0 (0–3)	1.7 ± 1.0 (0–3)	1.4 ± 1.0 (0–3)
	G	1.2 ± 0.9 (0–3)	1.3 ± 0.8 (0–3)	1.2 ± 1.0 (0–3)	1.5 ± 0.8 (0–3)	1.8 ± 0.7 (0–3)	1.0 ± 0.7 (0–2)	1.4 ± 0.9 (0–3)	1.5 ± 0.8 (0–3)	1.1 ± 0.9 (0–3)

Data are given as mean ± SD (range), unless indicated otherwise. N is the total number of patients.
y, years; m/f, male/female; n, number of patients; variants, variant PSP phenotypes; n.a., not available (UPDRS III was only performed in the DescribePSP study); PSP-^{7x3}CDS, Progressive Supranuclear Palsy 7x3 Clinical Deficits Scale; A, Akinesia rigidity; B, Bradyphrenia; C, Communication; D, Dysphagia; E, Eye movements; F, Finger dexterity; G, Gait and balance.

TABLE 2. Cross-sectional analysis: correlation of PSP-CDS with other established scales

	Analysis	SEADL	CGI-S	MoCA	PSP-QoL	PSPSS	PSPRS
All phenotypes	Exploratory (DescribePSP)	–0.69***	0.60***	–0.35**	0.47**	0.65***	0.85***
		n = 87	n = 87	n = 76	n = 55	n = 84	n = 87
	Confirmatory (ProPSP)	–0.73***	0.58***	–0.40**	0.50***	0.53***	0.91***
		n = 75	n = 76	n = 76	n = 74	n = 76	n = 76
	Joint analysis	–0.71***	0.60***	–0.37***	0.50***	0.59***	0.88***
		n = 162	n = 163	n = 152	n = 129	n = 160	n = 163
PSP-RS	Exploratory (DescribePSP)	–0.71***	0.51***	*–0.26*	0.36*	0.53***	0.77***
		n = 54	n = 54	n = 48	n = 34	n = 52	n = 54
	Confirmatory (ProPSP)	–0.71***	0.63***	–0.56***	0.53***	0.58***	0.93***
		n = 51	n = 52	n = 52	n = 50	n = 52	n = 52
	Joint analysis	–0.74***	0.64***	–0.41***	0.47***	0.59***	0.86***
		n = 105	n = 106	n = 100	n = 84	n = 104	n = 106
Variant PSP phenotypes	Exploratory (DescribePSP)	–0.69***	0.74***	–0.45**	0.59*	0.81***	0.94***
		n = 33	n = 33	n = 28	n = 21	n = 32	n = 33
	Confirmatory (ProPSP)	–0.63**	*0.40*	*–0.27*	0.52*	*–0.01*	0.78***
		n = 24	n = 24	n = 24	n = 24	n = 24	n = 24
	Joint analysis	–0.66***	0.59***	–0.36*	0.51**	0.52***	0.89***
		n = 57	n = 57	n = 52	n = 45	n = 56	n = 57

Spearman r correlation coefficients. N is the number of analyzed pairs.
All correlations were statistically significant, except for italic values (not significant): *P* values: * < 0.05; ** < 0.005; *** < 0.0001.

disorders and corticobasal syndrome[39] have been consistently reported as cognitive deficits in PSP. Also, bulbar dysfunction leading to dysarthria and dysphagia were frequently reported in PSP.[40] Because of their particular relevance for PSP, these aspects have been included as core or supportive features into the MDS-PSP diagnostic criteria.[2] Thus, the PSP-CDS was constructed using the following seven domains: Akinesia-rigidity, Bradyphrenia, Communication, Dysphagia, Eye movements, Finger dexterity, and Gait & balance (Fig. 1).

Cognitive Pretesting

Twenty-six questionnaires were completed at four different study centers. Completion time of the PSP-CDS was 4.0 ± 1.1 versus 13.6 ± 3.5 minutes for the PSPRS ($P < 0.001$). Raters reported difficulties in choosing a score mainly when a patient's history report differed markedly from the clinical exam (e.g., in gait and dysphagia items). In general, items were considered to be easy to evaluate (mean item rating: 1.7 ± 0.2; range, 1 [very easy] to 4 [very difficult]). Given that no item was considered particularly difficult, we did not change any item, but rather developed a User Instruction (Supporting Information S1).

A second round of cognitive pretesting of the PSP-CDS provided to eight independent raters along with the User Instruction confirmed the scale to be easily applicable (item ratings: Akinesia-rigidity 1.4 ± 0.5, Bradyphrenia 1.9 ± 0.6, Communication 1.3 ± 0.5, Dysphagia 1.0 ± 0.0, Eye movements 1.5 ± 0.8, Finger dexterity 1.3 ± 0.5, and Gait & balance 1.1 ± 0.4).

Cross-Sectional Analyses: Sample Description

Of the 190 patients recruited, 26 were excluded because of missing PSP-CDS or PSPRS data, leaving 164 patients for cross-sectional analysis (88 from DescribePSP, 76 from ProPSP). Their demographic and clinical data are shown in Table 1.

A total of 107 PSP-RS and 57 vPSP phenotypes were included. A detailed description of their diagnostic certainties and predominance types by the MDS-PSP criteria is shown in Supporting Information S2.

Figure 2A to 2C shows the frequency distribution of disease duration, PSPRS scores, and PSP-CDS scores.

Supporting Information S3 shows the individual PSP-CDS items' scores in the entire cross-sectional cohort. These analyses demonstrated no floor or ceiling effects, considered present if >15% of patients achieved worst or best scores.

Cross-Sectional Analyses: Reliability

Item-to-total correlations for the PSP-CDS were above the common threshold of 0.4 for all phenotypes in the entire study population (Supporting Information S4).

CA for the PSP-CDS in the joint cross-sectional cohort was 0.75 (Supporting Information S5).

Inter-rater reliability and test-retest stability showed excellent correlations (Supporting Information S6A,B).

Cross-Sectional Analysis: Validity

Individual PSP-CDS items' scores correlated well with scores from other scales designed to measure similar conceptual constructs, analyzing all predominance

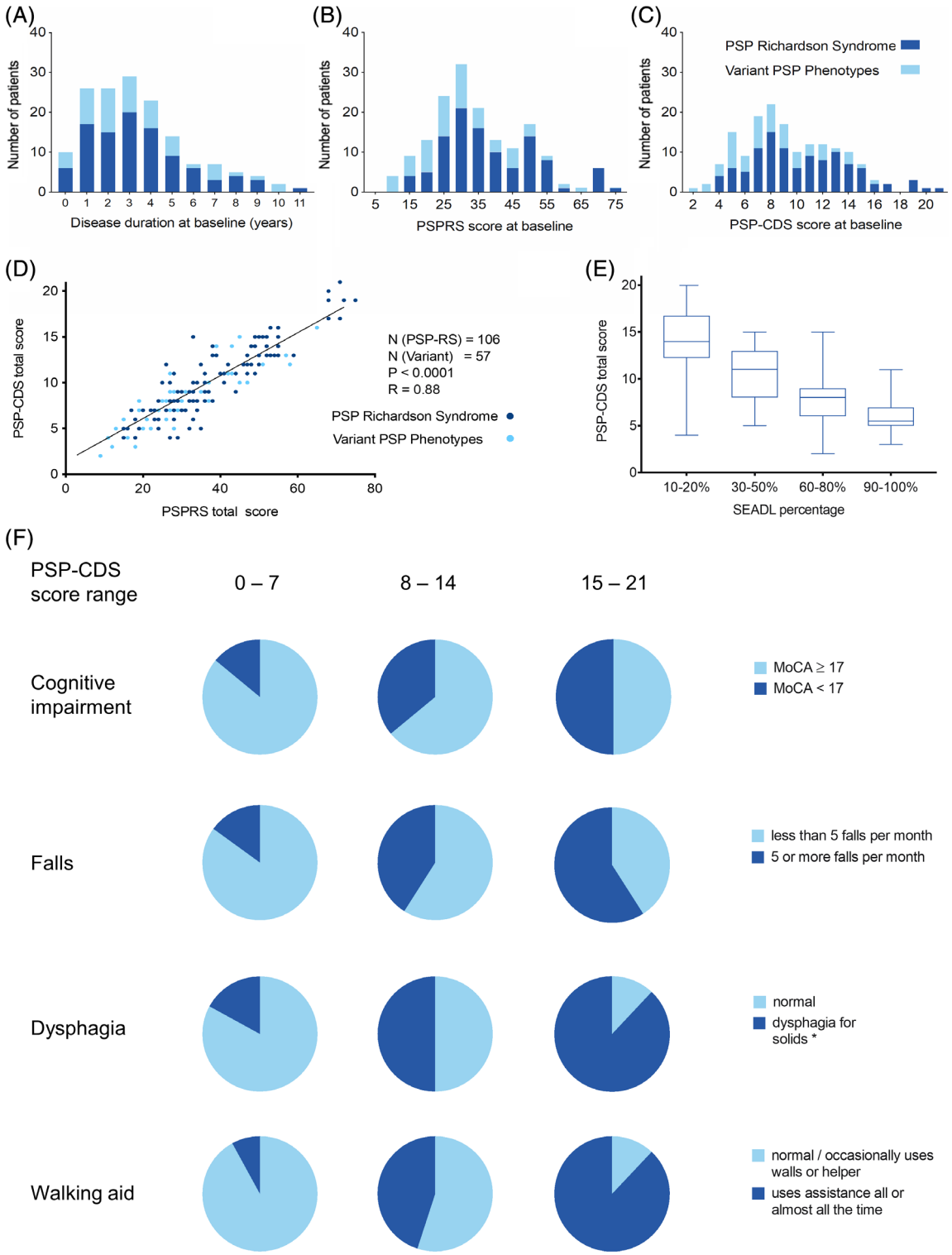

FIG. 2. PSP-CDS performance. (A–C) Bar graphs of disease duration (A) and frequency distribution of PSP-CDS (B) and PSPRS scores (C). (D) Linear regression of PSP-CDS total score against PSPRS total score. (E) Box and whiskers plot shows distribution of total PSP-CDS scores by ascending SEADL quartiles. Boxes represent interquartile ranges; the horizontal line in the middle shows the median value. Whiskers show minimum and maximum values. (F) Pie charts depicting disease milestones reached according to PSP-CDS score progression (score divided into three equally sized subgroups [tertiles], corresponding to mild, moderate, and severe disease stage). Milestones were conceptualized based on PSPRS items N.3, N.5, and N.26 and MoCA total score. *Ranging from cutting up tough foods to requiring artificial feeding. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 3. Longitudinal analysis: sensitivity to change

			Values at		Annualized Difference			30% Change	50% Change
		N	Baseline	Follow-up	Mean ± SD (% of baseline)	*P* Value	Standardized Effect Size	Sample Size	Sample Size
Sex m/f, N (%)		49	31 / 18 (63/37)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Age at examination (y)		49	68.7 ± 7.6 (52–85)	69.8 ± 7.5 (53–86)	n.a.	n.a.	n.a.	n.a.	n.a.
Disease duration at examination (y)		47	3.4 ± 2.1 (0–9)	4.4 ± 2.1 (1–9)	n.a.	n.a.	n.a.	n.a.	n.a.
PSP-CDS	Total	49	8.6 ± 3.6 (3–19)	10.8 ± 3.6 (5–20)	3.4 ± 3.4 (39.5% ± 39.5%)	<0.0001	1.00	177 (205)	65 (75)
	A	49	1.2 ± 0.7 (0–3)	1.6 ± 0.8 (0–3)	0.8 ± 1.3 (66.7% ± 108.3%)	<0.0001	0.65	410 (474)	148 (172)
	B	49	1.1 ± 0.8 (0–3)	1.5 ± 0.7 (0–3)	0.5 ± 1.2 (45.5% ± 109.1%)	0.0039	0.45	855 (989)	309 (357)
	C	49	1.2 ± 0.7 (0–3)	1.5 ± 0.7 (0–3)	0.4 ± 0.8 (33.3% ± 66.7%)	0.0232	0.45	854 (989)	309 (357)
	D	49	0.7 ± 0.7 (0–2)	0.9 ± 0.6 (0–2)	0.3 ± 1.0 (42.9% ± 142.9%)	0.0712	0.28	2,228 (2578)	803 (929)
	E	49	1.8 ± 0.9 (0–3)	2.2 ± 0.7 (0–3)	0.5 ± 1.0 (27.8% ± 55.6%)	0.0080	0.49	726 (840)	262 (303)
	F	49	1.5 ± 0.9 (0–3)	1.8 ± 0.9 (0–3)	0.4 ± 1.1 (26.7% ± 73.3%)	0.0972	0.31	1,813 (2099)	654 (757)
	G	49	1.1 ± 0.7 (0–3)	1.5 ± 0.8 (0–3)	0.6 ± 1.0 (54.5% ± 90.9%)	0.0037	0.55	585 (677)	212 (245)
PSPRS		49	31.9 ± 13.4 (11–75)	39.2 ± 14.7 (12–80)	10.8 ± 9.4 (33.9% ± 29.5%)	0.0006	1.15	133 (154)	49 (56)
PSP-QoL		32	30.0 ± 15.2 (3.8–75.0)	39.1 ± 19.5 (6.7–85.2)	13.1 ± 27.0 (43.7% ± 90.0%)	0.0184	0.49	742 (859)	268 (310)
PSPSS		46	2.5 ± 1.0 (1–4)	3.0 ± 0.9 (1–5)	0.5 ± 1.0 (20.0% ± 40.0%)	0.0121	0.54	595 (689)	215 (249)
UPDRS III		20	27.8 ± 13.7 (6–51)	41.4 ± 16.9 (16–79)	14.8 ± 18.6 (53.3% ± 66.9%)	0.0063	0.79	278 (322)	101 (117)
SEADL		49	61.8 ± 24.5 (10–90)	51.2 ± 24.0 (10–90)	−16.7 ± 22.5 (27.0% ± 36.4%)	0.0023	−0.74	319 (370)	116 (134)
MoCA		39	22.3 ± 5.1 (10–28)	21.0 ± 5.8 (5–30)	−1.7 ± 5.6 (7.6% ± 25.1%)	0.1774	−0.31	1,853 (2144)	668 (773)

Sample sizes required for a two-arm, 1-year follow-up therapeutic trial to detect 30% or 50% change.
Data statistics at baseline (BL) and follow-up (FU) as well as annualized rate of change, defined as follow-up scale score minus baseline score divided by time in years, and power calculations. Estimated sample sizes needed to detect a 30% and 50% rate of change based on 80% power, two-sided, two-sample *t* test, were calculated. Approximations of the sample size for the Mann-Whitney U test are given in parentheses. *P* values were calculated with *t* tests. Data are given as mean ± SD (range), unless indicated otherwise. N is the total number of patients.
n.a., not applicable; A, Akinesia-rigidty; B, Bradyphrenia; C, Communication; D, Dysphagia; E, Eye movements; F, Finger dexterity; G, Gait and balance.

types in the entire study population (Supporting Information S7).

Correlations of the PSP-CDS total score with established clinical scales commonly used to monitor disease severity in PSP were good to excellent, both in the exploratory, confirmatory, and joint cohorts and when analyzing all predominance types, PSP-RS, or vPSP predominance types only (Table 2).

Specifically, the PSP-CDS total score was positively correlated with the PSPRS total score, for both PSP-RS and vPSP phenotypes, in the entire study cohort (Fig. 2D).

Decreasing SEADL quartiles, indicating decreasing ADLs, had increasing mean PSP-CDS total scores (Fig. 2E).

In turn, increasing PSP-CDS tertiles yielded higher percentages of patients reaching predefined, clinically meaningful disease milestones (Fig. 2F). These milestones were the MoCA threshold suggestive of dementia (<17),[18,41,42] five or more falls per month, dysphagia for solid foods, and regular use of a walking aid.

Longitudinal Analysis: Sensitivity to Change

Twelve-month follow-up data were obtained from 51 patients, 2 of whom had to be excluded because of incomplete data, leaving 49 patients for longitudinal

analyses (24 from DescribePSP, 25 from ProPSP). Their demographic and clinical data are shown in Table 3. Their diagnostic certainties and predominance types by the MDS-PSP criteria are shown in Supporting Information S2.

Annualized score changes and power calculations for the PSP-CDS and established scales are shown in Table 3. The PSP-CDS showed significant 1-year change (baseline, 8.6 ± 3.6; follow-up: 10.8 ± 3.6; annualized difference: 3.4 ± 3.4; $P < 0.001$). The standardized effect size for the PSP-CDS (1.0) was similar to the PSPRS (1.15). Sample sizes required per arm for a two-arm, 1-year follow-up therapeutic trial to detect 50%-change were estimated to be 65 (two-sided, two-sample *t* test).

Short Scale Versions

Statistical analysis of both shorter versions show very similar performance as the PSP-CDS^{7x3} (Supporting Information S10 and S11), with the exception of power calculations, where both shorter versions had lower sensitivity to change (Supporting Information S12): The sample sizes required per arm for a two-arm, 1-year follow-up therapeutic trial to detect 50% change in the score progression (two-sided, two-sample *t* test) were estimated to be 65 for PSP-CDS^{7x3}, 88 for PSP-CDS^{7x2}, and 160 for PSP-CDS^{6x2}.

Discussion

Based on a systematic analysis of the literature and expert consensus, the PSP-CDS was conceptualized to cover the broad clinical spectrum of PSP. Our aim was to create a simple, straightforward scale, applicable to all PSP phenotypes equally. One of our other most important aspects was time efficiency compared to other established PSP scales, while still trying to maintain excellent clinimetrics.

First, we assessed its psychometric properties and usability. The most frequently rated score was 1 for most items (Supporting Information S3), because our patients were in relatively early disease stages (Table 1). The only item with a peak score at 2 was "Eye movements," most likely because of its diagnostic hallmark nature for the disease. Floor and ceiling effects were not observed (Fig. 2A–C). Based on cognitive pretesting, we developed a PSP-CDS User Instruction (Supporting Information S2), allowing the standardized collection of information required for rating by a semistructured interview with the patient and/or a reliable caregiver and a structured clinical examination.

Criterion and construct validity were demonstrated by strong correlations of PSP-CDS items' scores and total scores with other patient- and physician-reported outcomes. Correlations of the PSP-CDS total score (Table 2) were strongest with the PSPRS (see also Fig. 2D). A significant correlation with the PSP-QoL suggests that the PSP-CDS also captures subjective and quality-of-life–oriented aspects of the disease. The correlation with the MoCA was weaker, maybe because cognitive deficits in PSP involve mainly frontal and executive dysfunction, apathy, and language disorders,[33,34,43-45] which are not the focus of the MoCA. The PSP-CDS items Bradyphrenia and Communication aim to cover those functionally relevant cognitive features. All correlations were similarly strong in the PSP-RS and vPSP subgroups, suggesting that the PSP-CDS—despite its shortness and unlike the PSPSS—may be well suited for the broader spectrum of PSP predominance types. Correlations of the PSP-CDS with other scales were strong for PSP patients with "probable" diagnostic certainty, whereas for "possible" and "suggestive of" diagnostic certainties they were significant only with PSPRS and SEADL (Supporting Information S13). We ascribe this to the low numbers of patients and high clinical heterogeneity in the "possible" and "suggestive of" patient groups and will follow this up in future work.

All findings were comparable between the exploratory and confirmatory cohort, as well as their joint analysis, demonstrating the reproducibility. Decreasing independence level, measured by ascending SEADL quartiles, matched well with increasing PSP-CDS scores (Fig. 2E), implying that the PSP-CDS measures features that lead to increased dependency. In this context, further investigations should be conducted to examine the correlation of the PSP-CDS with caregiver burden. The ability of the PSP-CDS to indicate functionally relevant restrictions in daily living was further demonstrated by relating the PSP-CDS total score to well-defined, functionally relevant disease milestones (Fig. 2F). As a milestone for cognitive decline, we used a predefined cutoff of the MoCA score,[18,41,42] which indicated increasing cognitive impairment with increasing PSP-CDS scores (Fig. 2F). Consistently, the correlation of PSP-CDS scores with the MoCA scores (Table 2) and of the PSP-CDS bradyphrenia item with the PSPRS mentation subscore (Supporting Information S7) were both significant.

In summary, these analyses show that the PSP-CDS captures clinical deficits reported by physicians and experienced by the patient equally well.

Internal consistency (CA; Supporting Information S5) was above the commonly requested threshold of 0.7 in the joint analysis. Considering that CA measures the mean correlation of the items among themselves, this value probably reflects the heterogeneity of the items that were chosen to cover deficits in the broad spectrum of PSP predominance types, as well as the concision of the PSP-CDS.[46]

Item-total correlations (Supporting Information S5) were above common thresholds of 0.4 to 0.5,[47,48]

demonstrating good homogeneity and cross-correlation of the scale despite its modest CA. In patients with vPSP phenotypes, who often display only limited symptoms during the early stages, item-total correlations were shown to be lower.

Inter-rater-reliability and test-retest-stability (Supporting Information S6) were excellent, but have to be interpreted with caution given our small sample and possible recall effects. Intraclass correlation coefficients indicated an excellent level of agreement,[49-51] suggesting the PSP-CDS to consist of well-defined items that are easy to understand and can be objectively applied.

Correlations of individual items' scores with similar constructs (Supporting Information S7) confirm the adequacy for measurement of the domains. Items were intended to cover a broad range of ADL-relevant deficits. For example, the communication item, designed to capture PSP-related communicative disabilities resulting from deficits in different neurological systems, correlates well with both the withdrawal and dysarthria items of the PSPRS.

To estimate the PSP-CDS' disease progression sensitivity, we conducted a longitudinal analysis over a 12-month period (Table 3). Annualized rates of change for the total PSP-CDS versus baseline were significant (+39.5%) and of similar magnitude as the PSPRS progression rate (+31.9%); however, this requires confirmation in independent and international research settings. The individual PSP-CDS items Akinesia-rigidity, Eye movements, and Gait & balance showed the strongest 1-year progression, whereas Communication, Dysphagia, Bradyphrenia, and Finger dexterity progressed less in our cohort. Given that our longitudinal cohort mainly consisted of PSP-RS in earlier disease stages, we might expect different progression rates of individual subitems in later disease stages and in cohorts with predominant vPSP patients.

Comparative power calculations of the scales collected in our longitudinal data set (Table 3) showed that the PSPRS requires the smallest sample size (49 patients per arm for a two-arm, 1-year follow-up therapeutic trial to detect a 1-year-50% change with a two-sided, two-sample *t* test, an estimate that is in line with previously published data).[16] The PSP-CDS was estimated to require slightly more patients (full PSP-CDS$^{7\times3}$: 65 patients; short PSP-CDS$^{7\times2}$: 88 patients; short PSP-CDS$^{6\times2}$: 160 patients). The UPDRS III (101 patients) and SEADL (116 patients) were in a similar range, but other scales required a notably higher number of patients. Thus, the PSP-CDS might be useful as a short and versatile tool to monitor disease status and progression. The relatively low number of vPSP patients in our longitudinal cohort does not allow one to draw final conclusions on the usefulness of the PSP-CDS for monitoring disease progression in vPSP subgroups, requiring further longitudinal analyses with a higher numbers of patients.

The PSP-CDS and the conducted analyzes have several limitations. First, because of the shortness of the PSP-CDS and the heterogeneity of symptoms captured therein, it is unavoidable that some psychometric features may be inferior to scales describing specific features in greater detail. By focusing on the core features from a broad spectrum of deficits in PSP, we risk neglecting deficits that might have functional impact on individual patients. On the other hand, the research group decided purposefully to include only clinical deficits considered to be most relevant to all PSP patients after performing a systematic literature analysis, to avoid distortions by including features which may be of minor clinical relevance to most PSP-patients or only present in a minority of patients.

A second limitation results from the predominant outpatient setting of our study, limiting the number of participating patients in late disease stages. This participation bias may have affected the clinimetric profile of the dysphagia item, which is expected to progress particularly in later stages.

A third limitation is the fact that the PSP-CDS is a physician-reported rather than a patient-reported scale. Again, this format was based on a deliberate decision to keep the scale as objective as possible in a disease which is well-known to limit the self-evaluation of clinical deficits attributable to frontal lobar cognitive dysfunction. On an exploratory basis, we have still tried to add individual PSP-QoL items to the PSP-CDS, which, however, compromised CA values. Also, it has to be taken into consideration that patient-reported quality-of-life scales often do not adequately reflect clinical progression.[52]

Finally, we cannot exclude some interdependency of the result of one scale from another in our data, given that the same examiners rated all scales for individual patients at the visits.

Moving toward the future, further validation in larger cohorts and longer follow-up periods, as well as more extensive psychometric evaluations including Rasch analysis are warranted. As a start, however, the PSP-CDS represents a reliable, valid, sensitive, and easily applicable scale for both research and clinical settings. Because of its well-defined clinical domains and distinct response categories, it is easy to memorize, while still broadly covering PSP phenotypes. Major advantages of the PSP-CDS are its time efficiency and lack of need for extensive training. With good sensitivity to change and the broad coverage of patient-relevant clinical domains, the PSP-CDS meets many objectives and requirements as a measure for clinical care, for observational studies and for therapeutic trials in PSP-RS as well as variant phenotypes.

Acknowledgments: This project was supported by the German Center for Neurodegenerative Diseases (DZNE) and the German Parkinson's Association (DPG).

References

1. Rösler TW, Tayaranian Marvian A, Brendel M, et al. Four-repeat tauopathies. Prog Neurobiol. 2019;180:101644.
2. Hoglinger GU, Respondek G, Stamelou M, et al. Clinical diagnosis of progressive supranuclear palsy: the Movement Disorder Society criteria. Mov Disord 2017;32:853–864.
3. Boxer AL, Yu JT, Golbe LI, Litvan I, Lang AE, Höglinger GU. Advances in progressive supranuclear palsy: new diagnostic criteria, biomarkers, and therapeutic approaches. The Lancet Neurology 2017;16:552–563.
4. Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurol 2007;6:1094–1105.
5. European Medicines Agency (EMA). Reflection paper on the regulatory guidance for the use of the health-related quality of life (HRQL) measures in the evaluation of medicinal products. Amsterdam, The Netherlands: EMA; 2016.
6. FDA. Patient reported outcome measures: use in medical product development to support labelling claims. Rockville, MD: Dockets Management, Food and Drug Administration; 2006.
7. Revicki DA, Regulatory I; Patient-Reported Outcomes Task Force for the International Society for Quality of Life Research. FDA draft guidance and health-outcomes research. Lancet 2007;369:540–542.
8. Campbell M. Measuring How Patients Feel and Function. Roadmap to Engaging CDER Public Workshop, FDA, 2017. https://www.fda.gov/media/104969/download.
9. Boxer AL, Yu JT, Golbe LI, Litvan I, Lang AE, Hoglinger GU. Advances in progressive supranuclear palsy: new diagnostic criteria, biomarkers, and therapeutic approaches. Lancet Neurol 2017;16:552–563.
10. Fahn S, Elton RL; UPDRS Program Members. Recent Developments in Parkinson's Disease. Florham Park, NJ: Macmillan Healthcare Information; 1987.
11. Kroonenberg PM, Oort FJ, Stebbins GT, Leurgans SE, Cubo E, Goetz CG. Motor function in Parkinson's disease and supranuclear palsy: simultaneous factor analysis of a clinical scale in several populations. BMC Med Res Methodol 2006;6:26.
12. Cubo E, Stebbins GT, Golbe LI, et al. Application of the Unified Parkinson's Disease Rating Scale in progressive supranuclear palsy: factor analysis of the motor scale. Mov Disord 2000;15:276–279.
13. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a Frontal Assessment Battery at bedside. Neurology 2000;55:1621–1626.
14. Hall DA, Forjaz MJ, Golbe LI, et al. Scales to assess clinical features of progressive supranuclear palsy: MDS Task Force report. Mov Disord Clin Pract 2015;2:127–134.
15. Borroni B, Alberici A, Agosti C, Cosseddu M, Padovani A. Pattern of behavioral disturbances in corticobasal degeneration syndrome and progressive supranuclear palsy. Int Psychogeriatr 2009;21:463–468.
16. Stamelou M, Schope J, Wagenpfeil S, et al. Power calculations and placebo effect for future clinical trials in progressive supranuclear palsy. Mov Disord 2016;31:742–747.
17. Ghosh BC, Carpenter RH, Rowe JB. A longitudinal study of motor, oculomotor and cognitive function in progressive supranuclear palsy. PLoS One 2013;8:e74486.
18. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005;53:695–699.
19. 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 (Vienna) 2016;123:1435–1442.
20. Schwab R, England A. Projection technique for evaluating surgery in Parkinson's disease In: Gillingham FJ, Donaldson IML, eds. Third Symposium on Parkinson's Disease. Edinburgh, UK: E&S Livingston; 1969:152–157.
21. Guy W, ed. ECDEU assessment manual for psychopharmacology. Rockville, MD: U.S. Dept. of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psychopharmacology Research Branch, Division of Extramural Research Programs; 1976.
22. Bang J, Lobach IV, Lang AE, et al. Predicting disease progression in progressive supranuclear palsy in multicenter clinical trials. Parkinsonism Relat Disord 2016;28:41–48.
23. Schrag A, Selai C, Quinn N, et al. Measuring quality of life in PSP: the PSP-QoL. Neurology 2006;67:39–44.
24. Golbe LI, Ohman-Strickland PA. A clinical rating scale for progressive supranuclear palsy. Brain 2007;130(Pt 6):1552–1565.
25. 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.
26. Tolosa E, Litvan I, Hoglinger GU, et al. A phase 2 trial of the GSK-3 inhibitor tideglusib in progressive supranuclear palsy. Mov Disord 2014;29:470–478.
27. Respondek G, Kurz C, Arzberger T, et al. Which ante mortem clinical features predict progressive supranuclear palsy pathology? Mov Disord 2017;32:995–1005.
28. Adair CE, Holland AC, Patterson ML, et al. Cognitive interviewing methods for questionnaire pre-testing in homeless persons with mental disorders. J Urban Health 2012;89:36–52.
29. Collins D. Pretesting survey instruments: an overview of cognitive methods. Qual Life Res 2003;12:229–238.
30. Tilley BC, LaPelle NR, Goetz CG, Stebbins GT; MDS-UPDRS Task Force. Using cognitive pretesting in scale development for Parkinson's disease: the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) example. J Parkinsons Dis 2014;4:395–404.
31. 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.
32. 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(Pt 6):1247–1258.
33. Donker Kaat L, Boon AJ, Kamphorst W, Ravid R, Duivenvoorden HJ, van Swieten JC. Frontal presentation in progressive supranuclear palsy. Neurology 2007;69:723–729.
34. Han HJ, Kim H, Park JH, et al. Behavioral changes as the earliest clinical manifestation of progressive supranuclear palsy. J Clin Neurol 2010;6:148–151.
35. Hassan A, Parisi JE, Josephs KA. Autopsy-proven progressive supranuclear palsy presenting as behavioral variant frontotemporal dementia. Neurocase 2012;18:478–488.
36. Boeve B, Dickson D, Duffy J, Bartleson J, Trenerry M, Petersen R. Progressive nonfluent aphasia and subsequent aphasic dementia associated with atypical progressive supranuclear palsy pathology. Eur Neurol 2003;49:72–78.
37. Ferrer I, Hernandez I, Boada M, et al. Primary progressive aphasia as the initial manifestation of corticobasal degeneration and unusual tauopathies. Acta Neuropathol 2003;106:419–435.
38. Josephs KA, Duffy JR. Apraxia of speech and nonfluent aphasia: a new clinical marker for corticobasal degeneration and progressive supranuclear palsy. Curr Opin Neurol 2008;21:688–692.
39. Ling H, O'sullivan SS, Holton JL, et al. Does corticobasal degeneration exist? A clinicopathological re-evaluation. Brain 2010;133(Pt 7):2045–2057.
40. Muller J, Wenning GK, Verny M, et al. Progression of dysarthria and dysphagia in postmortem-confirmed parkinsonian disorders. Arch Neurol 2001;58:259–264.
41. Freitas S, Simões MR, Alves L, Santana I. Montreal cognitive assessment: validation study for mild cognitive impairment and Alzheimer disease. Alzheimer Dis Assoc Disord 2013;27:37–43.

42. Damian AM, Jacobson SA, Hentz JG, et al. The Montreal Cognitive Assessment and the mini-mental state examination as screening instruments for cognitive impairment: item analyses and threshold scores. Dement Geriatr Cogn Disord 2011;31:126–131.
43. Brown RG, Lacomblez L, Landwehrmeyer BG, et al. Cognitive impairment in patients with multiple system atrophy and progressive supranuclear palsy. Brain 2010;133(Pt 8):2382–2393.
44. Rittman T, Coyle-Gilchrist IT, Rowe JB. Managing cognition in progressive supranuclear palsy. Neurodegener Dis Manag 2016;6:499–508.
45. Gerstenecker A, Mast B, Duff K, Ferman TJ, Litvan I, Group E-PS. Executive dysfunction is the primary cognitive impairment in progressive supranuclear palsy. Arch Clin Neuropsychol 2013;28:104–113.
46. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011;2:53–55.
47. Clark LA, Watson D. Constructing validity: new developments in creating objective measuring instruments. Psychol Assess 2019;31: 1412–1427.
48. Ladhari R. Developing e-service quality scales: a literature review. J Retailing Consumer Services 2010;17:464–477.
49. Cicchetti, DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items—applications to assessment of adaptive-behavior. Am J Ment Defic 1981;86:127–137.
50. Gamer M, Lemon J, Singh IFP. Irr: Various Coefficients of Interrater Reliability and Agreement. 2019. https://CRAN.R-project.org/package=irr. Access on 19th Nov. 2019.
51. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–428.
52. Higginson IJ, Carr AJ. Measuring quality of life: using quality of life measures in the clinical setting. BMJ 2001;322:1297–1300.
53. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.

APPENDIX 1

The DescribePSP study group

Thomas Gasser, Andreas Hermann, Günter Höglinger, Matthias Höllerhage, Okka Kimmich, Thomas Klockgether, Johannes Levin, Gerrit Machetanz, Antje Osterrath, Carla Palleis, Johannes Prudlo, and Annika Spottke.

APPENDIX 2

The ProPSP study group

Daniela Berg, Katrin Bürk, Joseph Claßen, Carsten Eggers, Andrea Greuel, Max-Joseph Grimm, Lennard Hermann, Vassilena Iankova, Klaus Jahn, Wolfgang Jost, Martin Klietz, Andrea Kühn, Franz Marxreiter, Steffen Paschen, Monika Poetter-Nerger, Marie-Therese Preisl, Lisa Prilop, Lars Tönges, Claudia Trenkwalder, Tobias Warnecke, Florian Wegner, and Jürgen Winkler.

APPENDIX 3

The MDS-endorsed PSP study group

Angelo Antonini, Kailash P. Bhatia, Adam L. Boxer, Carlo Colosimo, Yaroslau Compta, Jean-Christophe Corvol, Lawrence I. Golbe, Günter U. Höglinger, Anthony E. Lang, Irene Litvan, Huw R. Morris, Christer Nilsson, Alexander Pantelyat, Gesine Respondek, and Maria Stamelou.

Supporting Data

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