

# A short scale for the assessment of motor impairments and disabilities in Parkinson's disease: the SPES/SCOPA

J Marinus, M Visser, A M Stiggelbout, J Martin Rabey, P Martínez-Martín, U Bonuccelli, P H Kraus, J J van Hilten

*J Neurol Neurosurg Psychiatry* 2004;75:388–395

**Objectives:** To evaluate the reliability and validity of the Short Parkinson's Evaluation Scale (SPES)/Scales for Outcomes in Parkinson's disease (SCOPA)—a short scale developed to assess motor function in patients with Parkinson's disease (PD).

**Methods:** Eighty five patients with PD were assessed with the SPES/SCOPA, Unified Parkinson's Disease Rating Scale (UPDRS), Hoehn and Yahr (H&Y) scale, and Schwab and England (S&E) scale. Thirty four patients were examined twice by two different assessors who were blinded to each other's scores and test executions. Additionally, six items of the motor section of the SPES/SCOPA were assessed in nine patients and recorded on videotape to evaluate inter-rater and intra-rater reliability.

**Results:** The reproducibility of the sum scores in the clinical assessments was high for all subscales of the SPES/SCOPA. Inter-rater reliability coefficients for individual items ranged from 0.27–0.83 in the motor impairment section, from 0.58–0.82 in the activities of daily living section, and from 0.65–0.92 in the motor complications section. Inter-rater reliability of the motor items in the video assessments ranged from 0.70–0.87 and intra-rater reliability ranged from 0.81–0.95. The correlation between related subscales of the SPES/SCOPA and UPDRS were all higher than 0.85, and both scales revealed similar correlations with other measures of disease severity. The mean time to complete the scales differed significantly ( $p < 0.001$ ) and measured 8.1 (SD 1.9) minutes for the SPES/SCOPA and 15.6 (SD 3.6) minutes for the UPDRS.

**Conclusion:** The SPES/SCOPA is a short, reliable, and valid scale that can adequately be used in both research and clinical practice.

See end of article for authors' affiliations

Correspondence to:  
J J van Hilten, Department of Neurology (K5Q 94), Leiden University Medical Center, P.O. Box 9600, NL-2300 RC Leiden, The Netherlands;  
j.j.van\_hilten@lumc.nl

Received 25 April 2003  
In revised form 6 June 2003  
Accepted 9 June 2003

Over the past 15 years the Unified Parkinson's Disease Rating Scale (UPDRS) has become a standard tool in the clinical evaluation of patients with Parkinson's disease (PD). The UPDRS is the most frequently used scale in PD trials<sup>1</sup> and has acceptable inter-rater and intra-rater reliability for most items.<sup>2–3</sup> The construct validity with other scales is adequate but the content validity has been questioned, especially with respect to its conceptual clearness and balance between items that represent symptoms responsive to dopaminergic treatment and those more resistant to this intervention.<sup>2</sup> Other critiques include the length of the scale<sup>4</sup> and the redundancy of items.<sup>5</sup> The mean time to complete the scale is about 17 minutes for experienced users,<sup>4</sup> which makes it less suitable for clinical application. Van Hilten *et al*<sup>5</sup> demonstrated that the UPDRS can be shortened by removing redundant items from the motor section and conceptually unclear items from the activities of daily living (ADL) section without negative consequences for reliability or validity. A shorter scale with similar clinimetric properties may have advantages for patients, clinicians, and researchers.

As a result of the aforementioned considerations, the Short Parkinson's Evaluation Scale (SPES) was developed.<sup>6</sup> This scale is short, conceptually clear, and displays good reliability and validity.<sup>2–6</sup> The instrument is considered easy to use by its evaluators<sup>6</sup> but has only been used in a few studies.<sup>7–10</sup>

Careful inspection of the SPES, however, indicates that the consistency in the framing of response options may be improved. The item "swallowing" represents an impairment and should be moved from the disability (ADL) section to the motor impairment (MI) section in order to be consistent with current methodological concepts of scale construction.<sup>11–12</sup> Additionally, some clinimetric aspects of the SPES have not

been addressed to date. These aspects involve intra-rater and inter-rater reliability between two assessors who perform the clinical assessments separately. Hence, we first modified the SPES according to the aforementioned considerations and subsequently evaluated this scale the SPES/SCOPA. The development of this scale is part of a larger research project on Scales for Outcomes in Parkinson's disease (SCOPA)<sup>13</sup> in which short, practical, and clinimetric sound scales for all relevant domains in PD are selected or developed.

The objective of this study was to evaluate the reliability (intra-rater and inter-rater reliability, internal consistency) and construct validity (correlation with related scales, "known groups" comparisons) of the SPES/SCOPA.

## METHODS

### Development of the SPES/SCOPA

The SPES/SCOPA consists of three sections: MI, ADL, and motor complications. There are four response options ranging from 0 (normal) to 3 (severe). In comparison with the original SPES, some modifications to the three sections were made based on findings in the literature and empirical testing of some of the items. The mental section was removed altogether because we felt that these important functions could not be assessed in a reliable and valid way by a few

**Abbreviations:** ADL, activities of daily living; ANOVA, analysis of variance; H&Y, Hoehn and Yahr; ICC, intra-class correlation coefficient; MI, motor impairment; PD, Parkinson's disease; S&E, Schwab and England; SCOPA, Scales for Outcomes in Parkinson's disease; SPES, Short Parkinson's Evaluation Scale; UPDRS, Unified Parkinson's Disease Rating Scale

single questions; as a part of the SCOPA project, we tested and developed separate instruments for these functions.<sup>13</sup>

### Motor impairments

**Tremor.** Pooled data of several studies<sup>8 14–17</sup> (total 1361 patients) revealed that in less than 4% of the patients the tremor score of the legs was higher than that of the arms, and that in only 2% of the patients a tremor in the legs was present whereas it was absent in the arms. For reasons of efficiency we therefore decided to evaluate tremor in the upper extremities only. Additionally, in the response options we linked the amplitude of the tremor to a displacement in centimetres to improve the quantification of “small”, “moderate”, and “severe”. **Bradykinesia.** “Finger tapping” was replaced by “rapid alternating movements” on the basis of the results of a separate study in which we compared five different tests for bradykinesia.<sup>18</sup> This study revealed that this test had the highest intra-rater and inter-rater reliability and the highest correlation with measures of disease severity. Both tremor and bradykinesia were assessed during 20 seconds. A time window was not indicated in the original SPES. **Rigidity.** The phrase “detectable only on activation of the contralateral arm” was removed from the response options because a pilot study showed that the muscle tone in the ipsilateral arm also increased in many healthy individuals if the contralateral arm was raised. Rigidity was now evaluated by “perceived difficulty in trying to reach the end positions in elbow or wrist”, which proved to be a useful criterion in a pilot study we held in 17 patients with PD. **Postural stability.** This item was modified on the basis of a study in which we compared six tests for postural stability.<sup>19</sup> The word “retropulsion” was removed from the response options, and the scoring was now determined by the number of steps patients took to restore balance and by whether a patient would fall or not. **Arising from chair, gait, and speech.** These items only underwent minimal changes compared with the original SPES and do not need detailed discussion. Two other impairments were added and evaluated historically—namely, swallowing and freezing during “on”. We considered it more important to assign items to the appropriate section than to assign items to sections on the basis of the way they are elicited (examination *v* history). “Swallowing” was therefore moved from the ADL section to the MI section. “Freezing during on” was added because it was considered a useful progression marker of PD that is less responsive to dopaminergic interventions. The maximum score in the MI section is 42.

### Activities of daily living

The response options were framed as uniformly as possible. Responses reflected no difficulty (normal), some difficulty (no assistance needed), considerable difficulty (possibly needing assistance), and unable (or needing complete or almost complete assistance). As stated before, “swallowing” (previously addressed under “eating”) was removed. “Turning and getting out of bed” was extended to “changing positions” to include all important transfers of daily life. The maximum score in this section is 21.

### Motor complications

The original section on complications of treatment was modified and now evaluated both the presence and severity of dyskinesias and motor fluctuations. Freezing and dystonia were removed because these phenomena cannot solely be attributed to dopaminergic treatment and, hence, may not reflect complications of treatment. Additionally, in patients treated with levodopa, dystonia may emerge while patients are “on” or “off” or during the switch of one state to the other and coincide with other dyskinesias. In view of this

complexity and the fact that “off” and dyskinesias are already assessed, we decided not to evaluate dystonia separately. Items were framed to address the impairment level. The maximum score is 6 for the dyskinesia section and 6 for the motor fluctuations section.

### Patients

Eighty five consecutive, non-demented patients who visited the outpatient clinic of the Department of Neurology of the Leiden University Medical Center, and who fulfilled the United Kingdom Parkinson's Disease Society Brain Bank criteria for idiopathic PD,<sup>20</sup> were included in the clinical assessments. Patients were excluded if they also had other diseases of the central nervous system or were not able to understand Dutch. Nine patients that met the same eligibility criteria were included in video assessments. The study was approved by the medical ethics committee of the Leiden University Medical Center.

### Assessment procedures

The patients were evaluated with the SPES/SCOPA (appendix), UPDRS parts II–IV,<sup>21</sup> the Hoehn and Yahr (H&Y) scale,<sup>22</sup> and the Schwab and England (S&E) scale.<sup>23</sup> The latter three scales were included to evaluate the construct validity of the SPES/SCOPA. One global question evaluated overall ADL functioning and patients were asked to indicate on a seven-point scale (ranging from very good to very bad) how well they were able to carry out various daily activities in the past month. Patients also indicated whether they were “on” or “off” at the time of assessment. Additional information gathered included medication, disease duration, and comorbidity.

Reproducibility of the SPES/SCOPA was assessed in three different ways: inter-rater reliability of *clinical assessments*, inter-rater reliability of *video assessments*, and intra-rater reliability of *video assessments*. In the clinical assessments, 34 patients were assessed twice by the investigators (JM, MV) who assessed patients separately in immediate succession and were blind to each other's scores and test executions. Reproducibility was calculated in patients with a stable response to medication during this period—that is, who had no “on–off” transitions. In the video assessments, inter-rater reliability was assessed by an international panel of movement disorders specialists (from Italy, Israel, Spain, Germany, and the Netherlands) who rated nine videotaped patients with a stable response to medication during the time the recordings were made. This panel rated the videotaped patients twice with an interval of 7–14 days. In the video recordings, six items from the MI section were presented. The other four items of this section were not appropriate for video scoring—that is, rigidity, speech, and the two historic items. Rigidity could not be evaluated from video for obvious reasons, and the other three items were not included because patients and raters spoke different languages.

### Statistical analysis

Reproducibility was assessed with the intra-class correlation coefficient (ICC; one-way random effects model) if two ratings were compared (inter-rater reliability in clinical assessments and intra-rater reliability in video assessments), and with Kendall's coefficient of concordance *W* if it concerned agreement between more than two raters (inter-rater reliability in video assessments). The ICC is equivalent to the weighted  $\kappa$  if quadratic weights are used.<sup>24</sup> We therefore used the “strength-of-agreement” classification as proposed by Landis and Koch,<sup>25</sup> who classified strength of agreement as slight (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), or almost perfect (0.81–1.00). Internal consistency of subscales was evaluated with

Cronbach's  $\alpha$ . Construct validity was evaluated by determining the correlation between scales, using Pearson's  $r$  for the correlation of SPES/SCOPA with UPDRS and S&E, and the Spearman's test  $r_s$  for the correlation with H&Y and "global functioning". "Known groups" validity was assessed by comparing scores of patients with different disease severity—that is, by H&Y—using analysis of variance (ANOVA) and ordinal regression.

## RESULTS

The characteristics of patients that participated in the clinical assessment are presented in table 1. The five men and four women recorded on video had a mean age of 62.4 years and a mean disease duration of 12.4 years. Three of these patients were in H&Y stage 2, three in H&Y 3, and three in H&Y 4 (data not shown).

### Practicality

The mean time necessary to complete the scales was 8.1 (SD 1.9) minutes for the SPES/SCOPA and 15.6 (SD 3.6) minutes for the UPDRS. This difference was significant ( $p < 0.001$ ).

### Reliability of clinical assessments

One patient changed from "on" to "off" between both clinical assessments and was removed from the analysis; hence, inter-rater reliability was assessed in 33 patients. The inter-rater reliability coefficients for the motor sections of the clinical assessments were all at least "moderate" according to the Landis & Koch criteria, except for two items in the SPES/SCOPA ("postural tremor right hand", "rigidity right arm") and eight items in the UPDRS (table 2). The latter eight items also included the two items that had "fair" reliability in the SPES/SCOPA. The mean reliability coefficient, calculated over the items that were shared by both scales, was 0.56 for both. The mean ICC calculated over all the items of the motor sections was 0.58 for the SPES/SCOPA and 0.50 for the UPDRS. Two items in the UPDRS (rest tremor left and right leg) could not reliably be calculated as a result of insufficient dispersion; however, percentage agreement for these items was high (table 2). Results for the ADL section are presented in table 3. All agreements were at least "substantial" except "changing positions" in the SPES/SCOPA. The mean reliability coefficient calculated over the shared items of the ADL sections was 0.69 for the SPES/SCOPA and 0.71 for the UPDRS. The mean ICC over all items of the ADL section of

the UPDRS was 0.76. The motor complication sections of both scales (table 4) shared only one item in both the dyskinesia and the motor fluctuation section. The mean ICC for the items in the dyskinesia section was 0.83 for the SPES/SCOPA and 0.75 for the UPDRS. The mean ICC for items in the motor fluctuation sections was 0.67 for the SPES/SCOPA and 0.60 for the UPDRS. The ICCs of the sumscores of the UPDRS were generally higher than those of the SPES/SCOPA (table 5).

### Reliability of video assessments

The inter-rater reliability coefficients for the items in the video assessments were 0.70 or higher and therefore all at least "substantial" (table 6). Intra-rater reliability coefficients were higher, with all items above 0.80 ("almost perfect").

### Internal consistency

The internal consistencies of the SPES/SCOPA scales were higher than those of the UPDRS except for the MI scale (table 5). "Sensory symptoms" in the UPDRS-ADL had a negative corrected item-total correlation ( $-0.02$ ). Other items with corrected correlations below 0.20 involved rest tremor of the right hand (0.15) and swallowing (0.18) in the SPES/SCOPA; rest tremor of head, right hand, left hand, and right leg (0.03, 0.11, 0.16, and 0.18, respectively) in the UPDRS motor section; and tremor (0.05) in the UPDRS-ADL section. The corrected item-total correlation of the tremor of the left hand was considerably higher in both scales (0.40 in the SPES/SCOPA and 0.31 in the UPDRS).

### Validity

Correlations between related sections of the SPES/SCOPA and the UPDRS were 0.88 for MI, 0.86 for ADL, 0.86 for dyskinesias, and 0.95 for motor fluctuations. The correlations of these sections with the H&Y and S&E scale were all similar: all differences in correlations in a range between 0.02 and 0.10. The correlation between these sections and disease duration also bore strong resemblance, with coefficients of 0.38 (SPES/SCOPA) and 0.23 (UPDRS) for the motor sections, and 0.29 (SPES/SCOPA) and 0.36 (UPDRS) for the ADL sections. The correlation with global ADL functioning was 0.49 for the SPES/SCOPA ADL and 0.48 for the UPDRS ADL.

Mean scores of patients grouped by their H&Y stages (table 7) indicated significant differences between groups for both the motor and ADL sections of both scales (ANOVA; all  $p$  values  $< 0.001$ ). Post-hoc  $t$  tests showed no significant differences between patients in H&Y 2 and H&Y 3 in both scales, but differences between stages 2 and 4, and between stages 3 and 4, were significant. A significant trend was present in both sections of the scales, with higher scores for patients with more advanced PD.

## DISCUSSION

We evaluated several aspects of the clinimetric performance of the SPES/SCOPA. Inter-rater reliability and internal consistency obtained by clinical assessment among 85 patients were similar for the SPES/SCOPA and the UPDRS. The reproducibility of the sumscores of the SPES/SCOPA was high. Two items in the motor section displayed less than moderate agreement—that is, both "postural tremor" and "rigidity" of the right arm. The same items on the same side also performed only fairly in the UPDRS, and previous studies on the UPDRS have also found lower scores for these items.<sup>4 26 27</sup> Inter-rater reliability of the SPES/SCOPA MI section assessed from video displayed higher scores than the clinical assessments, with all values above 0.70. Intra-rater reliability is even higher, with all reliability coefficients over 0.80. Items in the ADL section were only evaluated in the

**Table 1** Patient characteristics

Number of patients	85
Mean (SD) age (y)	62.4 (10.4)
Number (%) males	52 (61.2)
Mean (SD) disease duration (y)	10.5 (5.8)
H&Y stage	
1	0
2	25
3	42
4	18
5	0
Mean (SD) SPES/SCOPA-MI	13.8 (5.1)
Mean (SD) SPES/SCOPA-ADL	8.5 (3.2)
Mean (SD) UPDRS-motor evaluation	32.6 (10.9)
Mean (SD) UPDRS-ADL	15.9 (6.0)
Mean (SD) S&E	75.7 (12.4)
Number assessed while "on"/"off"/missing	78/5/2
Number (%) on levodopa	72 (84.7)
Mean (SD) levodopa dose (mg) in users	535 (407)
Number (%) on dopamine agonists	63 (74.1)

ADL, activities of daily living; H&Y, Hoehn and Yahr; MI, motor impairment; S&E, Schwab and England.

**Table 2** Inter-rater reliability (ICC) of motor impairment items in clinical assessment

Shared items	SPES/SCOPA	UPDRS	Non-shared items
Rest tremor R	0.63	0.53	
Rest tremor L	0.67	0.45	
Postural tremor R	0.34*	0.34*	
Postural tremor L	0.64	0.58	
Rapid alternating mov R	0.58	0.59	
Rapid alternating mov L	0.45	0.50	
Rigidity R	0.27*	0.38*	
Rigidity L	0.61	0.64	
Rise from chair	0.83	0.83	
Postural instability	0.60	0.59	
Gait	0.60	0.63	
Speech	0.55	0.68	
	0.57		Freezing in "on"
	0.79		Swallowing
		0.36*	Facial expression
		0.21*	Rest tremor head
		0.00†	Rest tremor R leg
		0.00†	Rest tremor L leg
		0.59	Rigidity head
		0.29*	Rigidity R leg
		0.44	Rigidity L leg
		0.46	Finger tap R
		0.39*	Finger tap L
		0.28*	Hand movements R
		0.67	Hand movements L
		0.53	Leg agility R
		0.61	Leg agility L
		0.66	Posture
		0.23*	Body bradykinesia

\*Agreement is fair (ICC = 0.21–0.40) – all other agreements are at least "moderate" (>0.40)<sup>25</sup>; †estimates unreliable (insufficient dispersion); % agreement 97.10 (R leg) and 91.20 (L leg). Eight items in UPDRS (U) and two items in SPES/SCOPA (S) with poor agreement (≤0.40). ICC, intra-class correlation coefficient; L, left; mov, movement; R, right; UPDRS, Unified Parkinson's Disease Rating Scale.

clinical assessment situation and all displayed at least "substantial agreement", with the exception of "changing positions". Items in the motor complication section all showed "substantial" reproducibility. The results from our study comply with previous findings in which the SPES and UPDRS were compared<sup>6</sup> and comparable items displayed similar reliability. The internal consistency of all SPES/SCOPA scales was above 0.70, which is considered the minimum for group comparisons.<sup>28</sup>

The correlation between related SPES/SCOPA and UPDRS sections is high, which is not too surprising given the shared components of both scales. Additionally, the similarity of the correlations between the SPES/SCOPA and UPDRS compared with measures of disease severity, such as H&Y, S&E, global ADL functioning, and disease duration, is striking. This

endorses the impression that the scales capture the same phenomena. Differences between patients grouped by their H&Y stages also display similar results for both scales. Although the SPES/SCOPA contains only half the number of items and has four response options instead of five, reliability and validity are apparently preserved.

Inter-rater reliability in our study was generally lower than that seen in other studies. This is not surprising given that most previous studies used either video recordings or a design where several raters assessed patients simultaneously, thereby excluding potential biases caused by changes in the patient's state and differences in test executions. Video assessments are useful because they provide information on reproducibility in standardised situations and thus present the opportunity to locate weaker items that may benefit from

**Table 3** Inter-rater reliability (ICC) of ADL items in clinical assessment

Shared items	SPES/SCOPA	UPDRS	Non-shared items
Speech	0.66	0.68	
Feeding	0.63	0.61	
Dressing	0.80	0.76	
Hygiene	0.70	0.71	
Changing positions	0.58*	0.73	
Walking	0.61	0.66	
Handwriting	0.82	0.85	
		0.91	Salivation
		0.90	Swallowing
		0.73	Falling (unrelated to freezing)
		0.77	Freezing when walking
		0.83	Tremor
		0.68	Sensory complaints

\*All agreements "substantial" (>0.60), except changing positions (S15) where agreement is "moderate". ADL, activities of daily living; ICC, intra-class correlation coefficient; UPDRS, Unified Parkinson's Disease Rating Scale.

**Table 4** Inter-rater reliability (ICC) of motor complication items in clinical assessment

Shared items	SPES/SCOPA	UPDRS	Non-shared items
<b>Dyskinesias</b>			
Presence dyskinesias	0.92	0.96	Severity dyskinesias
	0.74	0.94	How disabling dyskinesias?
		0.71	How painful dyskinesias?
		0.39*	Early morning dystonia?
<b>Motor fluctuations</b>			
Presence "off" periods	0.69	0.65	Severity "off" periods
	0.65	0.71	Off periods predictable?
		0.41†	Off periods unpredictable?
		0.62	Sudden offs?

All agreements "substantial", except for \*U35 ("fair") and †U37 ("moderate").  
ICC, intra-class correlation coefficient; UPDRS, Unified Parkinson's Disease Rating Scale.

**Table 5** Inter-rater reliability and internal consistency of SPES/SCOPA and UPDRS

	ICC sumscores	ICC items	Cronbach $\alpha$	Item total*
SPES/SCOPA motor	0.86	0.27–0.83	0.74	0.15–0.65
UPDRS motor	0.90	0.00†–0.83	0.88	0.03–0.65
SPES/SCOPA ADL	0.89	0.58–0.82	0.81	0.38–0.72
UPDRS ADL	0.93	0.61–0.91	0.75	–0.02‡–0.65
SPES/SCOPA dyskinesias	0.89	0.74–0.92	0.92	NA§
UPDRS dyskinesias	0.94	0.39–0.96	0.58	0.11–0.63
SPES/SCOPA fluctuations	0.72	0.65–0.69	0.95	NA§
UPDRS fluctuations	0.75	0.41–0.71	0.74	0.48–0.66

\*Corrected item-total correlation; †tremor R and L leg – estimates unreliable because of insufficient dispersion; ‡sensory symptoms; §not applicable (only two items).  
Values obtained by clinical assessment in 85 patients.  
ADL, activities of daily living; ICC, intra-class correlation coefficient; UPDRS, Unified Parkinson's Disease Rating Scale.

clearer instructions and descriptions. However, knowledge on the degree of reliability if patients are assessed at separate occasions, either by the same or different assessors, provides additional information because it reflects the routine of studies and clinical practice. To the best of our knowledge, only one study assessed the reproducibility of the UPDRS over separate clinical examinations.<sup>3</sup> Because this study evaluated intra-rater reliability over a 2-week interval, whereas we have assessed inter-rater reliability in immediate succession, the different designs of the studies do not allow a direct comparison.

Future studies that aim to improve the performance of the items with lower reliability (postural tremor right, rapid alternating movements left, rigidity right) are recommended. Postural tremor and rapid alternating movements may benefit from a shorter time window. The longer the assessment time the greater chance that the performance becomes inconsistent ("destabilises"). Another point that may improve the reliability is to state more clearly that the response option that best reflects the patient's performance should be scored. Some raters may be in doubt whether to rate the best, worst, or mean performance. A shorter assessment period and clearer instructions may be helpful. Whether these considerations will indeed be helpful must first be assessed. Special attention must be paid to whether an improvement in reliability is not obtained at the expense of validity, as some problems typically become apparent after longer performance.

Improving the item on rigidity will probably be difficult. One possibility is to assess whether higher reliability can be obtained if rigidity is assessed separately for either the wrist or the elbow instead of a combination of both.

It is unfortunate that we did not include more patients in the more extreme stages of disease severity—that is, H&Y stages 1 and 5. It is difficult to indicate whether our results can be extrapolated to patients in the end ranges, but it is usually more difficult to rate the intermediate states (mildly, moderately affected) than the more extreme ones (normal, severely affected). More variation in the severity of the patients would probably have resulted in larger differences between individuals and, hence, in a larger between-subject variance. As the ICC is calculated as the ratio of the variance of true differences between individuals to the total variance this probably would have resulted in a higher ICC, which would indicate that our findings probably rather underestimate than overestimate the results.

**Table 6** Reliability of SPES/SCOPA items assessed by video

	Intra-rater*	Inter-rater†
Rest tremor R	0.95	0.83
Rest tremor L	0.87	0.71
Postural tremor R	0.91	0.71
Postural tremor L	0.93	0.87
Rapid alt mov R	0.86	0.80
Rapid alt mov L	0.81	0.70
Rise from chair	0.86	0.72
Postural instability	0.87	0.87
Gait	0.84	0.70

\*Assessed by weighted  $\kappa$  (quadratic weights) (14 raters); †assessed by Kendall's coefficient of concordance W (14 raters).  
alt mov, alternating movement; L, left; R, right.

**Table 7** Scale scores grouped by H&Y stages ("known groups" comparisons)

	H&Y*			ANOVA	Post-hoc <i>t</i> tests*			Ordinal regression		
	2	3	4		2-3	3-4	2-4	Trend†	Fit‡	R-sq§
SPES/SCOPA MI	11.1	13.4	18.7	<0.001	0.11	<0.001	<0.001	<0.001	0.57	0.26
UPDRS ME	27.7	31.0	43.1	<0.001	0.51	<0.001	<0.001	<0.001	0.78	0.25
SPES/SCOPA ADL	7.2	7.9	11.8	<0.001	0.84	<0.001	<0.001	<0.001	0.33	0.30
UPDRS ADL	13.4	15.0	21.6	<0.001	0.72	<0.001	<0.001	<0.001	0.44	0.26

\**p* Values of differences in scores of patients in different H&Y stages; †*t* significant trend indicates that the trend differs significant from zero—that is, no trend; ‡goodness of fit: *p*>0.05 indicates that the model does not differ significantly from a model with a good fit; §R-sq: proportion of the variance accounted for by the model.

ADL, activities of daily living; ANOVA, analysis of variance; ME, motor examination; MI, motor impairment; UPDRS, Unified Parkinson's Disease Rating Scale.

To summarise, the SPES/SCOPA is a reliable, valid, and conceptually clear scale that is completed in half the time it takes to administer the UPDRS. These advantages may favour the use of the SPES/SCOPA in evaluating motor function in patients with PD.

### ACKNOWLEDGEMENTS

Professor RAC Roos is gratefully acknowledged for reviewing the manuscript. The authors thank the following persons for their participation in evaluating the videotapes: Dr S Agostini, Dr S Bernardini, Dr C Berti, Dr D Caneparo, Dr E Cubo, Dr G Gambaccini, Dr C Klein, Dr T van Laar, Dr C Lucetti, and Dr T Prokhorov.

### Authors' affiliations

**J Marinus, M Visser, J J van Hilten**, Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands

**A M Stiggelbout**, Department of Medical Decision Making, Leiden University Medical Center, Leiden, The Netherlands

**J Martin Rabey**, Department of Neurology, Assaf Harofeh Medical Center, Zerefin, Israel

**P Martínez-Martín**, Neuroepidemiology Unit, Department of Applied Epidemiology, National Center for Epidemiology I.S. Carlos III, Madrid, Spain

**U Bonuccelli**, Department of Neurology, University Hospital, Pisa, Italy

**P H Kraus**, Department of Neurology, St Josef University Hospital, Bochum, Germany

This study was financially supported by The Netherlands Organization for Scientific Research (NWO; project no. 0940-33-021).

Competing interests: none declared

## APPENDIX – SPES/SCOPA SCALE

### A. MOTOR EVALUATION

#### Clinical examination

##### 1. Rest tremor

Assess each arm separately during 20 seconds; hands rest on thighs; if tremor is not evident at rest, try to keep the patient attentive—for example, by having them count backwards with eyes closed

0 = absent

1 = small amplitude (<1 cm) occurring spontaneously, or obtained only while keeping patient attentive (any amplitude)

2 = moderate amplitude (1–4 cm), occurring spontaneously

3 = large amplitude (≥4 cm), occurring spontaneously.

##### 2. Postural tremor

Check with arms outstretched, pronated and semipronated, and with index fingers of both hands almost touching each other (elbows flexed); assess each position during 20 seconds

0 = absent

1 = small amplitude (<1 cm)

2 = moderate amplitude (1–4 cm)

3 = large amplitude (≥4 cm).

##### 3. Rapid alternating movements of hands

Rapid alternating pronation/supination movements of upper hand, each time slapping the palm of the horizontally held lower hand during 20 seconds; each hand separately

0 = normal

1 = slow execution, or mild slowing and/or reduction in amplitude; may have occasional arrests

2 = moderate slowing and/or reduction in amplitude or hesitations in initiating movements or frequent arrests in ongoing movements

3 = can barely perform task.

##### 4. Rigidity

Assess passive movements of elbow and wrist over full range, with the patient relaxed in sitting position; ignore cogwheeling; check each arm separately

0 = absent

1 = mild rigidity over full range, no difficulty reaching end positions

2 = moderate rigidity, some difficulties reaching end positions

3 = severe rigidity, considerable difficulties reaching end positions.

##### 5. Rise from chair

Patient is instructed to fold arms across chest; use straight back chair

0 = normal

1 = slowly; does not need arms to get up

2 = needs arms to get up (can get up without help)

3 = unable to rise (without help).

##### 6. Postural stability

Stand behind the patient and pull patient backwards, while patient is standing erect with eyes open and feet spaced slightly apart; patient is not prepared

0 = normal, may take up to two steps to recover

1 = takes three or more steps; recovers unaided

2 = would fall if not caught

3 = spontaneous tendency to fall or unable to stand unaided.

##### 7. Gait

Assess gait pattern; use walking aid or offer assistance, if necessary

0 = normal

1 = mild slowing and/or reduction of step height or length; does not shuffle

2 = severe slowing, or shuffles, or has festination

3 = unable to walk.

##### 8. Speech

0 = normal

1 = slight loss of expression, diction, and/or volume

2 = slurred; not always intelligible

3 = unintelligible always or most of the time.

**Historical information****9. Freezing during "on"**

*Freezing is characterised by hesitation when trying to start walking or being "glued" to the ground while walking*

0 = absent

1 = start hesitation only, occasionally present

2 = frequently present, may have freezing when walking

3 = severe freezing when walking.

**10. Swallowing**

0 = normal

1 = some difficulty or slow; does not choke; normal diet

2 = sometimes chokes; may require soft food

3 = chokes frequently; may require soft food or alternative method of food intake.

**B. ACTIVITIES OF DAILY LIVING****11. Speech**

0 = normal

1 = some difficulty; may sometimes be asked to repeat sentences

2 = considerable difficulty; frequently asked to repeat sentences

3 = unintelligible most of the time.

**12. Feeding (cutting, filling cup, etc.)**

0 = normal

1 = some difficulty or slow; does not need assistance

2 = considerable difficulty; may need some assistance

3 = needs almost complete or complete assistance.

**13. Dressing**

0 = normal

1 = some difficulty or slow; does not need assistance

2 = considerable difficulty; may need some assistance—for instance, buttoning, getting arms into sleeves

3 = needs almost complete or complete assistance.

**14. Hygiene (washing, combing hair, shaving, brushing teeth, using toilet)**

0 = normal

1 = some difficulty or slow; does not need assistance

2 = considerable difficulty; may need some assistance

3 = needs almost complete or complete assistance.

**15. Changing position (turning over in bed, getting up out of bed, getting up out of a chair, turning around when standing)**

0 = normal

1 = some difficulty or slow; does not need assistance with any change of position

2 = considerable difficulty; may need assistance with one or more changes of position

3 = needs almost complete or complete assistance with one or more changes of position.

**16. Walking**

0 = normal

1 = some difficulty or slow; does not need assistance or walking aid

2 = considerable difficulty; may need assistance or walking aid

3 = unable to walk, or walks only with assistance and great effort.

**17. Handwriting**

0 = normal

1 = some difficulty—for instance, slow, small letters; all words legible

2 = considerable difficulty; not all words legible; may need to use block letters

3 = majority of words are illegible.

**C. MOTOR COMPLICATIONS****18. Dyskinesias (presence)**

0 = absent

1 = present some of the time

2 = present a considerable part of the time

3 = present most or all of the time.

**19. Dyskinesias (severity)**

0 = absent

1 = small amplitude

2 = moderate amplitude

3 = large amplitude.

**20. Motor fluctuations (presence of "off" periods)**

*What proportion of the waking day is patient "off" on average?*

0 = none

1 = some of the time

2 = a considerable part of the time

3 = most or all of the time.

**21. Motor fluctuations (severity of "off" periods)**

0 = absent

1 = mild end-of-dose fluctuations

2 = moderate end-of-dose fluctuations; unpredictable fluctuations may occur occasionally

3 = severe end-of-dose fluctuations; unpredictable on-off oscillations occur frequently.

**REFERENCES**

- Mitchell SL, Harper DW, Lau A, et al. Patterns of outcome measurement in Parkinson's disease clinical trials. *Neuroepidemiology* 2000;19:100-8.
- Ramaker C, Marinus J, Stiggebout AM, et al. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. *Mov Disord* 2002;17:867-76.
- Siderowf A, McDermott M, Kieburtz K, et al. Test-Retest reliability of the Unified Parkinson's Disease Rating Scale in patients with early Parkinson's disease: results from a multicenter clinical trial. *Mov Disord* 2002;17:758-63.
- Martinez-Martin P, Gil-Nagel A, Morlan Gracia L, et al. Unified Parkinson's Disease Rating Scale characteristic and structure. *Mov Disord* 1994;9:76-83.
- van Hilten B, van der Zwan AD, Zwiderman AH, et al. Rating impairment and disability in Parkinson's disease: evaluation of the Unified Parkinson's Disease Rating Scale. *Mov Disord* 1994;9:84-8.
- Rabey JM, Bass H, Bonuccelli U, et al. Evaluation of the Short Parkinson's Evaluation Scale: a new friendly scale for the evaluation of Parkinson's disease in clinical drug trials. *Clin Neuropharmacol* 1997;20:322-37.
- Werber EA, Rabey JM. The beneficial effect of cholinesterase inhibitors on patients suffering from Parkinson's disease and dementia. *J Neural Transm* 2001;108:1319-25.
- Reichmann H, Brecht HM, Kraus PH, et al. [Prampipexole in Parkinson disease. Results of a treatment observation]. *Nervenarzt* 2002;73:745-50.
- Klein C, Zoldan J, Korcyn AD, et al. Pergolide as an adjunct therapy in Parkinson's disease evaluated using the SPES. *Rev Med Hosp Gen (Mex)* 2000;63:155-64.
- Rabey JM, Klein C, Molochnikov A, et al. Comparison of the Unified Parkinson's Disease Rating Scale and the Short Parkinson's Evaluation Scale in patients with Parkinson's disease after levodopa loading. *Clin Neuropharmacol* 2002;25:83-8.
- World Health Organization. The international classification of functioning, disability, and health. [www3.who.int/icf/onlinebrowser/icf.cfm](http://www3.who.int/icf/onlinebrowser/icf.cfm).
- Verbrugge LM, Jette AM. The disablement process. *Soc Sci Med* 1994;38:1-14.
- SCOPA homepage. [www.lumc.nl/2050/research/scopa\\_homepage.html](http://www.lumc.nl/2050/research/scopa_homepage.html).
- Guttman M. Double-blind comparison of prampipexole and bromocriptine treatment with placebo in advanced Parkinson's disease. International Prampipexole-Bromocriptine Study Group. *Neurology* 1997;49:1060-5.
- Hubble JP, Koller WC, Cutler NR, et al. Prampipexole in patients with early Parkinson's disease. *Clin Neuropharmacol* 1995;18:338-47.

- 16 Lieberman A, Ranhosky A, Korts D. Clinical evaluation of pramipexole in advanced Parkinson's disease: results of a double-blind, placebo-controlled, parallel-group study. *Neurology* 1997;**49**:162–8.
- 17 Pinter MM, Pogarell O, Oertel WH. Efficacy, safety, and tolerance of the non-ergoline dopamine agonist pramipexole in the treatment of advanced Parkinson's disease: a double blind, placebo controlled, randomised, multicentre study. *J Neural Neurosurg Psychiatry* 1999;**66**:436–41.
- 18 Gruber RA, Marinus J, Visser M, et al. Inter- and intra-rater reliability and discriminative ability of five measures of bradykinesia in subjects with and without Parkinson's disease. *Mov Disord* 2002;**17**[suppl 5]:S119.
- 19 Visser M, Marinus J, Bloem BR, et al. Clinical tests for postural instability in patients with Parkinson's disease. *Mov Disord* 2002;**17**[suppl 5]:S120.
- 20 Gibb WR, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *J Neural Neurosurg Psychiatry* 1988;**51**:745–52.
- 21 Fahn S, Elton RL, and members of the UPDRS Development Committee. Unified Parkinson's Disease Rating Scale. In: Fahn S, Marsden CD, Goldstein M, et al, eds. *Recent developments in Parkinson's disease*. Florham Park, NJ: Macmillan Healthcare Information, 1987:153–63.
- 22 Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. *Neurology* 1967;**17**:427–42.
- 23 Schwab RS, England Jr AC. Projection technique for evaluating surgery in Parkinson's disease. In: Gillingham FJ, Donaldson IML, eds. *Third symposium on Parkinson's disease*. Edinburgh: E & S Livingstone, 1969:152–7.
- 24 Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973;**33**:613–9.
- 25 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.
- 26 Prochazka A, Bennett DJ, Stephens MJ, et al. Measurement of rigidity in Parkinson's disease. *Mov Disord* 1997;**12**:24–32.
- 27 Richards M, Marder K, Cote L, et al. Interrater reliability of the Unified Parkinson's Disease Rating Scale for motor examination. *Mov Disord* 1994;**9**:89–91.
- 28 Bland JM, Altman DG. Cronbach's alpha. *BMJ* 1997;**314**:572.

### Clinical Evidence—Call for contributors

*Clinical Evidence* is a regularly updated evidence based journal available worldwide both as a paper version and on the internet. *Clinical Evidence* needs to recruit a number of new contributors. Contributors are health care professionals or epidemiologists with experience in evidence based medicine and the ability to write in a concise and structured way.

#### Currently, we are interested in finding contributors with an interest in the following clinical areas:

Altitude sickness; Autism; Basal cell carcinoma; Breast feeding; Carbon monoxide poisoning; Cervical cancer; Cystic fibrosis; Ectopic pregnancy; Grief/bereavement; Halitosis; Hodgkins disease; Infectious mononucleosis (glandular fever); Kidney stones; Malignant melanoma (metastatic); Mesothelioma; Myeloma; Ovarian cyst; Pancreatitis (acute); Pancreatitis (chronic); Polymyalgia rheumatica; Post-partum haemorrhage; Pulmonary embolism; Recurrent miscarriage; Repetitive strain injury; Scoliosis; Seasonal affective disorder; Squint; Systemic lupus erythematosus; Testicular cancer; Varicocele; Viral meningitis; Vitiligo

However, we are always looking for others, so do not let this list discourage you.

#### Being a contributor involves:

- Appraising the results of literature searches (performed by our Information Specialists) to identify high quality evidence for inclusion in the journal.
- Writing to a highly structured template (about 2000–3000 words), using evidence from selected studies, within 6–8 weeks of receiving the literature search results.
- Working with *Clinical Evidence* Editors to ensure that the text meets rigorous epidemiological and style standards.
- Updating the text every eight months to incorporate new evidence.
- Expanding the topic to include new questions once every 12–18 months.

If you would like to become a contributor for *Clinical Evidence* or require more information about what this involves please send your contact details and a copy of your CV, clearly stating the clinical area you are interested in, to Claire Folkes (cfolkes@bmjgroup.com).

### Call for peer reviewers

*Clinical Evidence* also needs to recruit a number of new peer reviewers specifically with an interest in the clinical areas stated above, and also others related to general practice. Peer reviewers are health care professionals or epidemiologists with experience in evidence based medicine. As a peer reviewer you would be asked for your views on the clinical relevance, validity, and accessibility of specific topics within the journal, and their usefulness to the intended audience (international generalists and health care professionals, possibly with limited statistical knowledge). Topics are usually 2000–3000 words in length and we would ask you to review between 2–5 topics per year. The peer review process takes place throughout the year, and our turnaround time for each review is ideally 10–14 days.

If you are interested in becoming a peer reviewer for *Clinical Evidence*, please complete the peer review questionnaire at [www.clinicalevidence.com](http://www.clinicalevidence.com) or contact Claire Folkes (cfolkes@bmjgroup.com).