Interrater Reliability of the Unified Parkinson's Disease Rating Scale Motor Examination

*‡§Marcus Richards, *‡§Karen Marder, *§Lucien Cote, and *†‡§Richard Mayeux

Departments of *Neurology, †Psychiatry, the ‡Gertrude H. Sergievsky Center, and §the Center for Alzheimer's Disease Research, College of Physicians and Surgeons, Columbia University, New York, U.S.A.

Summary: Interrater reliability of the Unified Parkinson's Disease Rating Scale (UPDRS) motor examination was assessed by three neurologists experienced in the administration of this scale. Intraclass correlation coefficients indicated good-to-excellent agreement for speeded repeated movements, resting tremor, arising from a chair, and gait; moderate agreement for action tremor, rigidity, posture, postural stability, and bradykinesia; and poor agreement for speech disorder and facial immobility. Overall, these results indicate that satisfactory interrater reliability is attainable with the UPDRS motor examination. **Key Words:** Parkinson's disease—Rating scale—Motor examination—UPDRS-Interrater reliability.

The motor examination of the Unified Parkinson's Disease Rating Scale (UPDRS) (1), which derives from the Columbia University Rating Scale (2), enables the quantification of type, number, and severity of extrapyramidal signs (EPS) and is widely used for the clinical evaluation of Parkinson's disease (PD). Because this quantification is based on standardized criteria, it is axiomatic that disagreement among raters in the interpretation of these criteria will introduce measurement error. This in turn will result in imprecise clinical characterization of the patient and in loss of statistical power when UPDRS values are used in research. It is therefore necessary to assess interrater reliability of the UPDRS, i.e., extent of agreement between different raters on quantification of EPS in the same patient. Two such studies have been undertaken (3,4), both reporting only moderate agreement. This might be expected in the second of these studies, because raters were inexperienced in the use of the UPDRS and did not undertake prior standardization training over its administration and interpretation.

On the other hand, raters in the first study were expert in the use of this scale. To determine whether high reliability estimates could be attained for the UPDRS, we undertook an interrater reliability assessment of this scale using three raters experienced in its administration.

METHODS

Subjects

Subjects were 24 patients (13 male, 11 female) with idiopathic PD randomly selected from the Washington Heights-Inwood Columbia Aging Project (WHICAP), a community-based prospective investigation of neurodegenerative disease in a geographically defined district of New York City. All PD patients in this study were identified by means of a community registry, whose structure is described in detail elsewhere (5). All subjects gave informed consent.

The Unified Parkinson's Disease Rating Scale Motor Examination

The UPDRS motor examination allows the quantification of speech, facial mobility, resting tremor (face and each limb), action tremor (upper limbs), rigidity (neck and each limb), finger taps (number of

seconds to complete 15 rapid taps of thumb with index finger, a nonstandard measure), hand movements (rapid opening-closing), rapid alternating movements (rams: pronation–supination), leg agility (rapid heel tapping), arising from a chair, posture, gait, postural stability, and body bradykinesia/hypokinesia. For each sign, a 5-step severity gradation was employed, with 0 representing absence and 4 representing maximum severity of that sign.

Design

A balanced incomplete block design (6) was employed, with all 24 patients double-rated by 2 of 3 attending neurologists (K.M., L.C., R.M.), each with extensive experience in movement disorders and highly familiar with the UPDRS. Rater combination (L.C. with K.M., K.M. with R.M., L.C. with R.M.) was counterbalanced. Within-pair order (L.C. eliciting, K.M. observing; K.M. eliciting, L.C. observing, etc.) was also counterbalanced. Each neurologist therefore rated a total of 16 patients and lead the examination for 8 of these. Rigidity ratings were performed by each rater in immediate succession. All other signs were rated by observation, with one rater eliciting patient responses. Raters did not discuss their ratings with each other during or after individual assessments, and thus remained unaware of each others' scores until the completion of the study.

A total EPS score was obtained by summing all the individual scores. Interrater reliability of this total and of each sign was assessed by intraclass correlation, estimated by the residual maximum likelihood method (7,8) using raters as a fixed effect. This method assesses the proportion of variability in the ratings that can be attributed to differences between patients and not to rater differences or measurement error. For signs requiring multiple measures (e.g., neck and all 4 limbs for rigidity), a mean value was calculated.

RESULTS

Demographic and clinical characteristics of the sample are presented in Table 1.

For each sign, mean ratings for the first and second rating are presented in Table 2. Reliability coefficients are also presented in Table 2. Analysis of variance (taking into account the balanced incomplete block design) revealed no difference between the three raters in the mean scores for the total UPDRS motor score and for each individual sign.

TABLE 1. *Mean and SD for demographic and clinical characteristics of the sample*

Variable	Mean	SD
Age (yr)	70.7	9.4
Education	10.2	6.5
PD duration (yr)	5.6	4.5
L-dopa (% taking)	66.7	—
Selegeline (% taking)	29.2	—
DA agonist (% taking)	8.3	—
Anticholinergics (% taking)	4.2	—

PD, Parkinson's disease; DA, dopamine.

DISCUSSION

This study indicated moderate-to-excellent agreement between raters on the severity of most EPS items on the UPDRS. In particular, agreement was good-to-excellent for signs involving speeded, repeated movements (RAMS, taps, hand movements, and leg agility); for resting tremor; for arising from a chair; and for gait. Agreement was moderate for action tremor, rigidity, posture, postural stability, and bradykinesia. Agreement was poor, however, for speech disorder and facial mobility.

It should be noted that extrapyramidal signs in the present sample (as indicated by the means in Table 2) were relatively mild and so these results cannot be automatically generalized to severely impaired patients.

In general, interrater reliability of the UPDRS was higher in the present study than previously reported (3,4). In one of these studies (4), disagreement between raters may have resulted from lack of

TABLE 2. *Mean ratings and intraclass correlation coefficients for the UPDRS items*

Sign	Mean score		Coefficient
	1st rating	2nd rating	
Total motor score[a]	29.37	27.87	0.82
Speech	1.04	1.04	0.29
Facial mobility	1.25	1.50	0.07
Resting tremor (mean)	0.52	0.48	0.84
Action tremor (mean)	1.27	1.00	0.67
Rigidity (mean)	1.53	1.37	0.49
Taps (mean)	15.10	16.02	0.7
Hand movements (mean)	1.52	1.31	0.75
RAMS (mean)	1.23	1.35	0.89
Leg agility (mean)	1.17	1.04	0.92
Arising from chair	1.00	1.33	0.75
Posture	1.62	1.33	0.49
Gait	1.12	1.12	0.77
Postural stability	1.17	1.33	0.54
Bradykinesia	1.50	1.50	0.64

UPDRS, Unified Parkinson's Disease Rating Scale.
[a] Sum of all individual UPDRS motor items.

familiarity with the scale. This was clearly not the case in the study by Ginanneschi et al. (3).

Concerning the poor reliability for facial immobility, it should be noted that severity criteria for this sign, as well as for posture and bradykinesia, allow a rating of 1 (i.e., the most mild degree of sign presence) to be considered within normal limits for an older person. We have previously argued that such ambiguity over the distinction between normal and abnormal can hinder agreement between raters (9). It might therefore be advisable to standardize severity criteria across all signs so that a score of ≥1 on any item clearly represents abnormality. This source of variance would not account for the moderate or poor reliability for speech, rigidity, and postural stability. However, it is likely that lack of agreement over the severity of these signs could be remedied by further standardization training involving comparison of scores during rating with a view to reducing interrater discrepancy. As Yule and Taylor (10) point out, "good inter-rater agreement *can* be achieved, but . . . it must be planned for. Methods must be adequately standardized and examiners must be properly trained."

REFERENCES

1. Stern MB. The clinical characteristics of Parkinson's disease and parkinsonian syndromes: diagnosis and assessment. In: Stern MB, Hurting HI, eds. *The comprehensive management of Parkinson's disease.* New York: PMA Publishing, 1978:3–50.

2. Duvoisin RC. The evaluation of extrapyramidal disease. In: de Ajuriaguerra G, ed. *Monoamines, noyaux gris centraux et syndrome de Parkinson.* Symposium Bel-Air. IV. 1971: 313–325.

3. Ginanneschi A, Degl'Innocenti F, Magnolfi S, et al. Evaluation of Parkinson's disease: reliability of three rating scales. *Neuroepidemiology* 1988;7:38–41.

4. Geminiani G, Cesana BM, Tamma F, et al. Interobserver reliability between neurologists in training of Parkinson's disease rating scales. *Mov Disord* 1991;6:330–335.

5. Mayeux R, Denaro J, Hemenegildo N, et al. A population-based investigation of Parkinson's disease with and without dementia: relationships to age and gender. *Arch Neurol* 1992;49:492–497.

6. Fleiss JL. *The design and analysis of clinical experiments.* New York: John Wiley, 1986.

7. Patterson HD, Thompson R. The recovery of inter-block information when block sizes are unequal. *Biometrica* 1971; 58:545–554.

8. Patterson HD, Thompson R. Maximum likelihood estimation of components of variance. *Proceedings of the 8th International Biometric Conference* 1975;197–207.

9. Richards M, Marder K, Bell K, Dooneief G, Mayeux R, Stern Y. Interrater reliability of extrapyramidal signs in a group assessed for dementia. *Arch Neurol* 1991;48:1147–1149.

10. Yule W, Taylor E. Classification of soft signs. In: Tupper DE, ed. *Soft neurological signs.* New York: Grune and Stratton, 1987:25.