Movement Disorders
Vol. 9, No. 4, 1994, pp. 390–394

Utility of an Objective Dyskinesia Rating Scale for Parkinson's Disease: Inter- and Intrarater Reliability Assessment

C. G. Goetz, G. T. Stebbins, *H. M. Shale, †A. E. Lang, *D. A. Chernik, T. A. Chmura, ‡J. E. Ahlskog, and *E. E. Dorflinger

*Department of Neurological Sciences, Rush Presbyterian St. Luke's Medical Center, Chicago, Illinois, U.S.A.; *Hoffmann-La Roche Inc., Nutley, New Jersey, U.S.A.; †Morton and Gloria Shulman Movement Disorder Centre, The Toronto Hospital, Toronto, Ontario, Canada; and ‡Department of Neurology, Mayo Clinic, Rochester, Minnesota, U.S.A.*

Summary: Although dyskinesia is a frequent and important problem in Parkinson's disease (PD), a reliable assessment measure has not been thoroughly developed and tested. We modified the Obeso dyskinesia scale to create an objective rating scale for dyskinesia assessment during activities of daily living. Thirteen physicians and 15 study coordinators involved in a clinical trial independently reviewed videotape segments of PD patients performing three tasks: walking, putting on a coat, and lifting a cup to the lips for drinking. Raters evaluated the severity of worst dyskinesia seen, the types of all dyskinesias seen, and the type of dyskinesia most associated with motoric disability. For all assessments, the total group showed statistically significant inter- and intrarater reliability. Physicians had a higher consistency than did coordinators, but for most measures the difference was not statistically significant. Physicians and coordinators found the scale easy to use and especially practical for rating dyskinesia severity and for identifying the most disabling dyskinesia. Dyskinesias can be assessed in clinical trials and warrant regular documentation. **Key Words:** Parkinson's disease—Dyskinesia—Rating scale.

Dyskinesias in Parkinson's disease (PD) can be phenomenologically variable, associated with marked disability and be responsive to therapy (1,2). The Unified Parkinson's Disease Rating Scale (UPDRS) is widely used in the evaluation of PD (3,4), but includes only a limited assessment of dyskinesias. Other available dyskinesia scales for PD are limited because they treat akinesia and dyskinesia as opposites along a continuum (5–7), have been adapted from tardive dyskinesia or Huntington's disease ratings (8), or are based largely on subjective information (9). Recently, a group of movement disorder physicians and study coordinators participated in a multicenter trial of a new antiparkinsonian agent. In preparing for this trial, we modified the Obeso dyskinesia scale (9) to create an objective scale based on specific motor tasks. Our aim was to develop a practical and readily usable scale that would reliably do the following:

1. Assess severity of overall dyskinesias based on interference in activities of daily living.
2. Distinguish chorea from dystonia, the two major types of dyskinesia in PD.
3. Identify the single most disabling form of dyskinesia.

A videotape segment accompanies this article.

Address correspondence and reprint requests to Dr. C. G. Goetz, 1725 West Harrison, Chicago, IL, U.S.A.

The results to this study were presented at the seventh annual symposium on etiology, pathogenesis and prevention of Parkinson's disease, Boston, MA, October 17, 1993.

METHODS

Forty patients with PD on chronic dopaminergic therapy and with varying types and severity of dyskinesias were videotaped during three tasks: putting on and buttoning a coat, taking a cup from a table top to the lips, and walking. Twenty segments were randomly edited onto a tape and sent to 13 physicians and 15 study coordinators involved in a clinical trial of a new antiparkinsonian agent. The raters were not familiar with any of the patients videotaped. The physicians were a mixture of university-based movement disorder specialists and practice neurologists with access to large populations of patients with PD. The coordinators were nurses regularly involved in neurologic care. Ninety-two percent of the physicians and 80% of the coordinators had previously been involved in clinical trials for PD.

Raters viewed the tapes and marked the severity of the worst dyskinesias seen using a 5-point severity rating scale (Table 1). Next, they identified the different types of dyskinesia seen (chorea, dystonia, other). Finally, they rated the most disabling dyskinesia seen on the tape, taking into account the three activities and all motor components of each of the complex tasks performed. Worse function was rated. Raters at each institution evaluated the videotapes independently and without discussion. After they returned their responses, a second tape of 20 patients (70% repeats from videotape 1 and the remaining new patients was sent 2 weeks later.

TABLE 1. *Dyskinesia rating scale*

Directions:

1. View the patient walk, drink from a cup, put on a coat and button clothing.
2. Rate the severity of dyskinesias. These may include chorea, dystonia, and other dyskinetic movements in combination. Rate the patient's worst function.
3. Check which dyskinesias you see (more than one response possible).
4. Check the type of dyskinesia that is causing the most disability on the tasks seen on the tape (only one response is permitted).

Severity rating code: 0, absent; 1, minimal severity, no interference with voluntary motor acts; 2, dyskinesias may impair voluntary movements but patient is normally capable of undertaking most motor acts; 3, intense interference with movement control and daily life activities are greatly limited; 4, violent dyskinesias, incompatible with any normal motor task.

Severity of worst dyskinesia observed	Dyskinesias present (more than one choice possible)			Most disabling dyskinesia (choose one)		
	Chorea (C)	Dystonia (D)	Other (list)	C	D	Other

The severity rating was modified from a scale developed by Obeso (9). The original scale evaluated patient perception of duration and an overall assessment of the intensity of dyskinesia by history and observation. To develop an objective scale, we eliminated the duration portion and modified the instructions for the intensity scale so that the rating was performed by an observer. Although the original scale included patient consciousness of dyskinesia as a component of severity, we based the rating only on objective signs. The three tasks were chosen by the investigators to be examples of activities of daily living that involved large and small muscles of all extremities as well as trunk and neck control.

Data Analyses

Inter-rater reliability for the severity of dyskinesia ratings was assessed using Kendall's coefficient of concordance (W) calculated for physicians and coordinators combined, as well as for each group separately. Rates of inter-rater agreement on types of dyskinesia observed and for most disabling dyskinesia were assessed using a κ coefficient of agreement, again calculated for physicians and coordinators combined and for each group separately. Comparison of physician agreement rates with those from coordinators was assessed by unpaired t tests. Consistency of ratings across the repeated videotapes (intrarater agreement) for severity of dyskinesia was calculated using a Spearman's Rank Order Correlation Coefficient for physicians and study coordinators, both combined and each separately. Intrarater agreement for type of dyskinesia and most disabling dyskinesia across the two videotapes was calculated using a Cramér Coefficient for physicians and study coordinators combined and as separate groups. After the second tape was rated, physicians and coordinators completed a questionnaire assessing how easily the scale could be used for rating severity, type, and most disabling dyskinesia. A 5-point scale was used for these ratings (1, very hard; 5, very easy). They also rated how well the severity rating scale and the three tasks chosen for the study reflected patient disability on a comparable 5-point scale (1, very poorly; 5, very well).

RESULTS

Severity of Dyskinesia

The examiners' ratings covered the scale's full range of dyskinesia severity (severity rating 0 =

19%, 1 = 32%, 2 = 33%, 3 = 8%, 4 = 8%) (Table 2). Combined physician and coordinator ratings showed significant interrater reliability for severity of dyskinesia on both tapes (first tape W = 0.760, df = 19, $p < 0.001$; second tape W = 0.876, df = 19, $p < 0.001$). Inter-rater reliability was higher for physicians than for coordinators, but the difference was not statistically significant for either the first tape ($t = 1.32$, $p > 0.05$) or the second tape ($t = 1.20$, $p > 0.05$). Intrarater consistency was high for the whole group ($r_s = 0.855$, $p < 0.001$), and physicians were significantly more consistent than were coordinators ($r_s = 0.908$ vs. $r_s = 0.826$, $t = 2.44$, $p < 0.05$).

Type of Dyskinesias Observed

Several types of dyskinesia were identified by the raters (chorea = 39%, dystonia = 21%, chorea plus dystonia = 33%, other = 7%). Analysis of combined ratings resulted in significant levels of agreement as to type of dyskinesia on both tapes (first tape K = 0.394, $p < 0.001$; second tape K = 0.422, $p < 0.001$). Interrater reliability was higher for physicians than for coordinators, but the difference was not statistically significant for either the first tape ($t = 1.05$, $p > 0.05$) or second tape ($t < 1.0$, $p > 0.05$). Intrarater consistency was high for the whole group ($C = 0.705$, $p < 0.01$), and although physicians were slightly more consistent than coordinators, the difference was not statistically significant ($C = 0.732$ vs. $C = 0.699$, $t < 1$, $p > 0.05$).

Most Disabling Dyskinesia

Raters identified the most disabling dyskinesia to be chorea in 61% of cases, dystonia in 37%, and other forms in 2%. Analysis of combined ratings resulted in significant levels of agreement as to the most disabling dyskinesia for both the first and second tapes (first tape K = 0.419, $p < 0.001$; second tape K = 0.378, $p < 0.001$). There were no significant differences between physicians' and coordinators' level of agreement for either the first tape ($t < 1.0$) or second tape ($t < 1.0$). Intrarater consistency was high for the whole group ($C = 0.837$, $p < 0.01$), and there were no statistically significant differences between physicians and coordinators.

Utility of the Scale

On the 5-point evaluation scale for ease of application, the median rating was high (rating of 4) for the assessment of the scale's ease for capturing severity and disability from dyskinesia. Raters considered the chosen tasks as appropriate for reflecting disability (median rating 4). Ease of identifying types of dyskinesia and most disabling dyskinesia

TABLE 2. *Ratings of interrater and intrarater reliability*

Rating	Tape	Rater	Interrater agreement[a]	p	Intrarater agreement[b]	p
Severity	1	Combined	0.760	<0.001		
		Physicians	0.821	<0.001		
		Coordinators	0.710	<0.001		
	2	Combined	0.876	<0.001	0.855	<0.001
		Physicians	0.897	<0.001	0.908	<0.001
		Coordinators	0.843	<0.001	0.826	<0.001
Type of dyskinesia	1	Combined	0.394	<0.001		
		Physicians	0.404	<0.001		
		Coordinators	0.377	<0.001		
	2	Combined	0.422	<0.001	0.705	<0.01
		Physicians	0.423	<0.001	0.732[c]	<0.01
		Coordinators	0.344	<0.001	0.699	<0.01
Most disabling dyskinesia	1	Combined	0.419	<0.001		
		Physicians	0.481	<0.001		
		Coordinators	0.338	<0.001		
	2	Combined	0.378	<0.001	0.837	<0.01
		Physicians	0.483	<0.001	0.840	<0.01
		Coordinators	0.279	<0.001	0.836	<0.01

[a] Kendall's coefficient of agreement for severity measure, and κ coefficient for type of dyskinesia and most disabling dyskinesia.

[b] Spearman rank order correlation for severity measure, and Cramér coefficient for type of dyskinesia and most disabling dyskinesia.

[c] Physicians showed significantly greater consistency on severity measure than did coordinators ($p < 0.05$). No other significant differences between physicians and coordinators were found.

were rated lower (median rating 3). The median rating for the overall utility of the scale was 4, indicating better than average utility.

DISCUSSION

In preparation for this program, we reviewed the available dyskinesia assessments, including a variety of hypo/hyperkinesia continuum scales (5–7), the Abnormal Involuntary Movement Scale (AIMS) (8), and the Obeso scale used in the Core Assessment Protocol for Intracerebral Transplantation (CAPIT) (9). Continuum scales treat tremor and other parkinsonian signs as polar opposites of dyskinesia, with 0 representing normal function, negative numbers representing parkinsonism, and positive numbers representing dyskinesia. Clinically, however, patients can be simultaneously parkinsonian and dyskinetic with disability related to each (1,2). Such scales have no capacity to integrate this information. The AIMS was developed primarily for tardive dyskinesia, and although it stresses intensity and anatomic distribution, it does not assess phenomenology or specific disabilities. Furthermore, the rating entries (none, minimal, mild, moderate, severe) have no explanatory phrases to help anchor the rating uniformly among raters. The CAPIT committee recommended the Obeso scale (9), but this assessment is primarily a subjective one based on timing and severity of dyskinesia by history. Its major advantages are its specific development for PD and its incorporation of functional descriptions for each numerical rating.

To eliminate subjective information that could not be validated, we modified and expanded the Obeso scale. We added a clinical protocol that was easy to administer and involved the use of readily available props, a lab coat, cup, and chair. We relied on objective observation as the source of data collection and therefore eliminated one intensity rating from the original scale that had distinguished between two equally severe dyskinesias, one of which the patient had no consciousness and the other of which the patient was fully aware. In our experience, subjective distress does not necessarily correlate with the objective severity of dyskinesia. Some patients are very bothered or conscious of mild dyskinesias and others are seemingly unaware of very striking, abnormal involuntary movements. Otherwise, we maintained the clinical descriptions developed in the original scale because we felt that they were well adapted to patients with PD. We added questions that focused on the distinction between dystonia and chorea, because these dyskinesias are phenomenologically and anatomically distinct from one another and are often associated with different disabilities (1,2).

Despite the utility and high inter- and intrarater reliability results from this scale, future modifications could be considered. Filling the cup with water would allow small spills to be documented as a source of disability and using a coat that fits each patient well would reduce variability based on differences in patient size. More importantly, we recommend small changes in the descriptive phrases in the severity rating scale in order to keep the scale focused on the tasks under observation, to emphasize the rating of the most intense dyskinesia seen, and to separate the middle categories. As follows, we refer to this adaptation as the Modified Dyskinesia Scale (version 2.0):

Severity Rating Code: Rate worst dyskinesia seen

0 = Absent
1 = Minimal severity: no interference with voluntary motor acts involved in the rated task.
2 = Dyskinesias impair voluntary movements but patient is capable of efficiently completing the motor acts involved in the rated task.
3 = Intense interference with movement control so that completing the rated motor task is greatly limited.
4 = Violent dyskinesias, incompatible with completion of the rated motor task.

These changes more clearly distinguish categories and remove qualifying language that could confuse the evaluator (e.g., "may," "most," "normally"). Rating 1 is differentiated from rating 2 by the impairment of normal voluntary control in rating 2; rating 2 is differentiated from rating 3 because successful execution of tasks is compromised in the latter.

The scale was aptly used both by physicians and study coordinators and both groups found it easy to apply. This finding suggests its easy incorporation into clinical trials where multiple assessments occur and where nonphysician professional nurses or coordinators are likely to collect data. As epidemiologic studies related to PD increase, scales that can be reliably administered by nonphysician field workers are also of practical value. We acknowledge, however, that the study participants were already familiar with movement disorder patients and that their participation in the clinical trial linked to this project was based in part on their experience in PD. Training sessions to focus on tremor, chorea,

and dystonia with specific examples of each would be necessary before initiating wide usage of this scale to less experienced investigative teams.

Study Raters and Affiliations (listed alphabetically)

C. Adler, Scottsdale, Arizona, U.S.A.; J. E. Ahlskog, Rochester, Minnesota, U.S.A.; P. Argatoff, Calgary, Alberta, Canada; P. Bedard, Quebec City, Quebec, Canada; M. Bergen, New Brunswick, New Jersey, U.S.A.; P. Blanchette, Quebec City, Quebec, Canada; M. Brewer, Phoenix, Arizona, U.S.A.; C. Gagnon, Quebec City, Quebec, Canada; L. Golbe, New Brunswick, New Jersey, U.S.A.; P. Gray, Ottawa, Ontario, Canada; S. Hatchard, London, Ontario, Canada; J. Hodder, Edmonton, Alberta, Canada; J. Hubble, Kansas City, Kansas, U.S.A.; S. Imke, Phoenix, Arizona, U.S.A.; A. Lang, Toronto, Ontario, Canada; A. Lieberman, Phoenix, Arizona, U.S.A.; W. Martin, Edmonton, Alberta, Canada; T. Mendis, Ottawa, Ontario, Canada; M. Muenter, Scottsdale, Arizona, U.S.A.; T. Roberts, Edmonton, Alberta, Canada; S. Rope, Kansas City, Kansas, U.S.A.; L. B. Scantlebury, Toronto, Ontario, Canada; K. Shannon, Chicago, Illinois, U.S.A.; A. J. Stoessl, London, Ontario, Canada; O. Suchowersky, Calgary, Alberta, Canada; J. Thelen, Chicago, Illinois, U.S.A.; K. Wheeler, Scottsdale, Arizona, U.S.A.; K. Wright, Rochester, Minnesota, U.S.A.

Disclosure Statement

This project was supported by a study grant from Hoffmann-La Roche Inc. Drs. Goetz, Stebbins, Lang and Ahlskog and Ms. Chmura receive no direct financial support from the sponsor. Drs. Shale, Chernik and Dorflinger are employees of Hoffmann-La Roche Inc.

LEGENDS TO VIDEOTAPE

The videotape has three parts, focusing on the severity rating scale (five patients), examples of chorea without dystonia, and dystonia without chorea (two patients) and cases of mixed dyskinesias where the most disabling is either chorea or dystonia (two patients). Examples chosen had high rater agreement.

Segment 1. Dyskinesia severity rating (n = 5 patients). Patient 1: 0, dyskinesia absent. Patient 2: 1, minimal severity, no interference with voluntary motor acts. Patient 3: 2, dyskinesia may impair voluntary movements, but patient is normally capable of undertaking most motor acts. Patient 4: 3, intense interference with movement control and daily life activities are greatly limited. Patient 5: 4, violent dyskinesias, incompatible with any normal motor task.

Segment 2. Identifying different forms of dyskinesia (n = 2 patients). Patient 6: chorea only, but note that the subject has parkinsonian tremor. Patient 7: Dystonia only in the foot, but no chorea.

Segment 3. Mixed dyskinesia: identifying most disabling form (n = 2 patients). Patient 8: chorea (diffuse) and dystonia (right foot) present, but most disabling is the foot dystonia that impairs walking. Patient 9: chorea (diffuse) and dystonia (neck) present, but most disabling is the chorea that impairs all motor acts.

REFERENCES

1. Marsden CD, Parkes JD, Quinn N. Fluctuations of disability in Parkinson's disease: clinical aspects. Im: Marsden CD, Fahn S, eds. *Movement disorders*. Vol. 2. London: Butterworth. 1982:96–105.
2. Tanner CM. Drug-induced movement disorders. In: Vinken PJ, Bruyn GW, Klawans HL, eds. *Extrapyramidal disorders: handbook of clinical neurology*. Vol. 49. Amsterdam, The Netherlands: Elsevier, 1986:185–204.
3. Fahn S, Elton RL. Members of the UPDRS Development Committee. Unified Parkinson's disease rating scale. In: Fahn S, Marsden CD, Calne DB, Golstein M, eds. *Recent developments in Parkinson's disease*. Vol. 2. Florham Park, NJ: MacMillan Healthcare Information, 1987:153–163.
4. Stebbins GT, Goetz CG, Flournoy T: Unified Parkinson's Disease Rating Scale: reliability and factorial validity of the motor exam section. *Ann Neurol* 1991;30:298.
5. Nutt JG, Woodward WR, Hammerstad JP, Carter JH, Anderson JL. The "on-off" phenomenon in Parkinson's disease: relation to levodopa absorption and transport. *N Engl J Med* 1984;310:483–488.
6. Kurlan R, Rubin AJ, Miller C, Rivera-Calimlim L, Clarke A, Shoulson I. Duodenal delivery of levodopa for on-off fluctuations in parkinsonism: preliminary observations. *Ann Neurol* 1986;20:262–265.
7. Mouradian MM, Juncos JL, Fabbrini G, Schlegel J, Bartko JJ, Chase TN. Motor fluctuations in Parkinson's disease. *Ann Neurol* 1988;24:372–378.
8. Guy W. *ECDEU assessment manual for psychopharmacology*. US Washington, DC: Government Printing Office, 1976.
9. Langston JW, Widner H, Goetz CG, Brooks DB, Fahn S, Freeman T, Watts R. Core assessment program for intracerebral transplantations (CAPIT). In: Lindvall O, Bjorklund A, Widner H, eds. *Intracerebral transplantation in movement disorders*. Amsterdam, The Netherlands: Elsevier, 1991:227–241.