Movement Disorders
Vol. 18, No. 7, 2003, pp. 738–750

State of the Art Review

The Unified Parkinson's Disease Rating Scale (UPDRS): Status and Recommendations

Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease[†]*

Abstract: The Movement Disorder Society Task Force for Rating Scales for Parkinson's Disease prepared a critique of the Unified Parkinson's Disease Rating Scale (UPDRS). Strengths of the UPDRS include its wide utilization, its application across the clinical spectrum of PD, its nearly comprehensive coverage of motor symptoms, and its clinimetric properties, including reliability and validity. Weaknesses include several ambiguities in the written text, inadequate instructions for raters, some metric flaws, and the absence of screening questions on several important non-motor aspects of PD. The Task Force recommends that the MDS sponsor the development of a new version of the UPDRS and encourage efforts to establish its clinimetric properties, especially addressing the need to define a Minimal Clinically Relevant Difference and a Minimal Clinically Relevant Incremental Difference, as well as testing its correlation with the current UPDRS. If developed, the new scale should be culturally unbiased and be tested in different racial, gender, and age-groups. Future goals should include the definition of UPDRS scores with confidence intervals that correlate with clinically pertinent designations, "minimal," "mild," "moderate," and "severe" PD. Whereas the presence of non-motor components of PD can be identified with screening questions, a new version of the UPDRS should include an official appendix that includes other, more detailed, and optionally used scales to determine severity of these impairments. © 2003 Movement Disorder Society

Key words: Unified Parkinson's Disease Rating Scale (UPDRS); Parkinson's disease; rating scales; clinimetrics

The International Executive Council of the Movement Disorder Society (MDS) has the authority and responsibility to develop task forces that address special topics of Society interest. In 2001, based on the increasing reliance on standardized rating scales for Parkinson's disease (PD) and concerns that currently available scales may under-represent many elements of PD impairment and disability, the society developed the Task Force for Rating Scales in PD. The mission of this Task Force is three-fold: to critique existing scales, to identify clinical areas that are not adequately rated, and to make recommendation on maintaining or modifying currently available scales. Specifically, the mission does not include the enactment of any official changes in existing scales or the development of new scales. The first critique concerns the Unified Parkinson's Disease Rating Scale [UPDRS].

The UPDRS is a scale that was developed as an effort to incorporate elements from existing scales to provide a comprehensive but efficient and flexible means to monitor PD-related disability and impairment.[1] Prior to its development, multiple scales, including the Webster,[2] Columbia,[3] King's College,[4] Northwestern University Disability,[5] New York University Parkinson's Disease Scale,[6] and UCLA Rating Scales,[7] were used in different centers, making comparative assessments difficult. The development of the UPDRS involved multiple trial versions, and the final published scale is officially known as UPDRS version 3.0.[1] The scale itself has four components, largely derived from preexisting scales that were reviewed and modified by a consortium of movement disorders specialists (Part I, Mentation, Behavior and Mood; Part II, Activities of Daily Living; Part III, Motor; Part IV, Complications). The original concept of the scale was to provide a core assessment tool that could be accompanied by additional measures to focus on global impairment or specific elements in more detail (Stanley

[†]For a full list of contributors, see Appendix 1.

*Correspondence to: Dr. Christopher G. Goetz, Rush-Presbyterian-St. Luke's Medical Center, 1725 W. Harrison Street, Chicago, IL 60612. E-mail: cgoetz@rush.edu

Received 10 January 2003; Accepted 17 January 2003

Fahn, personal communication). For example, whereas the UPDRS is often accompanied by and reported with such scales as the Schwab and England and Hoehn and Yahr scales, these latter scales are not part of the UPDRS per se.[8,9]

As part of the background work to develop the Task Force on Rating Scales for PD, the MDS secretariat staff (Caley Kleczka, Director) sent a questionnaire on UPDRS utilization patterns to all MDS members (see Appendix 2). Of 1,593 (1,405 e-mails and 188 fax communications), 185 members from all continents responded (12%). Ninety-six percent of responders had personal experience with the UPDRS, 87% using it in clinical trials, 70% in clinical practice, and 69% in other research venues. The sub-components of the UPDRS were used at different frequencies, 60% of responders regularly using Part I (Mentation), 80% Part II (Activities of Daily Living), 98% Part III (Motor), 65% Part IV (Complications). Combination scores based on sums of different parts were often used, I+II+III (37%) and II+III (41%). The questionnaire results demonstrated that responding MDS members use the UPDRS frequently and for multiple purposes, including research and clinical practice. Parts II and III are most widely used for both clinical and research purposes. These results confirmed the decision to prioritize the UPDRS as the first Task Force assessment.

METHODS

Administrative Organization

Under the Chairperson's direction, a Writing Committee was identified to draft the UPDRS critique and to remain on the Task Force in ongoing manner for future scale assessments. In addition, a series of movement disorder or statistical specialists with specific experience using the UPDRS participated as Expert Consultants. These specialists were recruited to serve on the Task Force for this critique only, with plans to rotate Expert Consultants for each future scale critique. The third group was the administrative staff of the MDS secretariat, assigned the organization of the review process, integration of materials, and editorial review.

Critique Process

Through Medline search and familiarity with the UPDRS literature, the chairperson supplied each Task Force member with a series of articles related to the UPDRS. Questions compiled by the Writing Committee were addressed to all Expert Consultants with the request to furnish a written document to respond to each point with suitable references from the articles provided or other data sets (Appendix 3). All unpublished data that were used for the critique were shared with the entire Task Force membership. An audio-taped meeting assembled all participants for open discussion of the UPDRS. A draft of the report was prepared by the Writing Committee and Secretariat staff and circulated to the Expert Consultants. The final document was assembled by the chairperson and approved by all Task Force members before submission to the MDS International Executive Committee. This Task Force document has been approved by the MDS Scientific Issues Committee prior to submission for journal peer-review in *Movement Disorders*.

RESULTS

Utilization of the UPDRS

One of the core advantages of the UPDRS is that it was developed as a compound scale to capture multiple aspects of PD. It assesses both motor disability (Part II: Activities of daily living) and motor impairment (Part III: Motor section). In addition, Part I addresses mental dysfunction and mood, and Part IV assesses treatment-related motor and non-motor complications. Of all available clinical scales for the assessment of parkinsonian motor impairment and disability, the UPDRS is currently the most commonly used.[10] Sixty-nine percent of 1994–1998 articles using a PD-rating scale relied on the UPDRS as the standard tool.[11] This trend is an international one, and the UPDRS predominates as the primary scale in published studies from both US and other geographical regions. It has been applied with equal frequency in studies of early and late PD.[11] The wide usage and global acceptance of the scale has resulted in its use for numerous multicenter studies. Further, the standardized ratings allow for summary scores used to communicate global severity of impairment and disability.

Another unique feature of the UPDRS is the availability of a teaching-videotape standardising the practical application of the scale and thereby serving as an important asset to enhance inter-rater reliability.[12] This feature is particularly relevant to the training of new raters and to the conduct of multicenter therapeutic trials in PD.

Despite its multidimensional approach with 4 different sections, the UPDRS has proven an easy-to-use instrument in clinical practice with an average time requirement for administration of the full scale between 10 and 20 minutes.[13] This time can be further shortened by self-administration of the Mentation and ADL sections by patients in the waiting room. There is good inter-rater reliability between patients who self-complete the historical sections (Part I, Mentation, and Part II, ADL) of the

UPDRS and the treating neurologist who interviews them.[14] When patients self-complete Parts I and II, the physician's time investment can focus primarily on Parts III and IV. When physicians are not available, the UPDRS has been effectively taught to other medical personnel for such applications as community-based neurological examinations.[15]

The UPDRS is increasingly used as a gold standard reference scale. The motor section has repeatedly been employed in attempts to develop surrogate markers for disease progression like beta-CIT-SPECT or 18-F-Dopa-PET.[16,17] The UPDRS is also the common reference scale in studies of instrument development for rating specific aspects of PD.[18–20] US and European regulatory agencies rely on the scale for new drug approvals,[21] and the UPDRS has also been used to define the placebo response in PD.[22] Almost all recent trials of surgical interventions for PD, both related to intracerebral transplantation and deep brain surgery, have employed the UPDRS. It is a key component of the Core Assessment Programs for Intracerebral Transplantation and Surgical Interventional Therapies for PD (CAPIT/CAPSIT).[23,24] Although specifically developed to assess PD, the UPDRS has been utilized to rate parkinsonian features of other conditions, including normal aging, progressive supranuclear palsy, and Lewy body dementia.[15,25,26]

Scale Application Across the Spectrum of PD

The UPDRS has been used in studies of early, mild PD, moderate but stable PD, and severe disease and motor fluctuations.[27,28] Prior studies have demonstrated that the scale favors the assessment of moderate and severe impairments, and may not be ideally configured to assess very mild disease-related signs and symptoms.[29] Although the UPDRS has been extensively applied to clinical trials of early PD to test the concept of neuroprotection or the impact of therapies on reducing the need to start levodopa (L-dopa) therapy, "floor" effects potentially limit the scale's utility in the early stages of the illness where impairment is subtle. To address this issue, some studies have permitted the inclusion of 0.5 ratings or other designations based on such anchors as "may be normal for healthy elderly subjects."[30] Modification of the UPDRS with such new wording or rating options, however, has not been validated.

Several longitudinal studies of PD have demonstrated that the UPDRS increases over time and scores are higher at key clinical decision-making points like the need to introduce symptomatic therapy.[27,31–33] Numerous studies indicate that the UPDRS is responsive to therapeutic interventions. Significant improvements in total UPDRS scores, individual subscales (Parts II and III), and averages of subscale scores obtained during *on* and *off* scores among fluctuators have been documented in comparison with placebo.[34] UPDRS improvements have been seen in patients with dose-finding studies new treatments of advanced disease as well as in studies focusing on mildly disabled patients.[28,35] Published reports using the UPDRS, however, have focused almost exclusively on Caucasians, and the UPDRS characteristics have not been extensively investigated in different racial or ethnic minorities.[36] Furthermore, the effects of gender and age on UPDRS ratings during treatment interventions have not been specifically examined.

Insufficient information is available on the ability of the UPDRS to discriminate between disease categories of clinical pertinence. To date, operative definitions of "minimal," "mild," "moderate" and "severe" stages of PD have not been explicitly defined. UPDRS scores, however, correlate with the Hoehn and Yahr scale and the Schwab and England scale.[13] Furthermore, within the UPDRS, the objective, physician-derived Motor section (Part III) correlates well with the subjective, patient-derived Activities of Daily Living (Part II) section.[13] If a measure such as the Schwab and England scale were used to define "minimal," "mild," "moderate," and "severe" PD, UPDRS scores could be tested to see how consistently they increase as the disease advances clinically. With this analysis, UPDRS scores could be developed to define numerically these clinical categories, expressed as ranges with 95% confidence intervals.

Clinimetric Issues

Of all available PD rating scales, the UPDRS has the additional advantage that it is the most thoroughly tested instrument from a clinimetric point of view. Almost one-third of all studies assessing clinimetric properties of impairment and disability scales for PD identified in a recent systematic review were targeted on the UPDRS.[10] Clinimetric scale evaluation usually assesses a scale's reliability and validity. Reliability evaluations assess the amount of measurement error in a scale, while validity evaluations assess the degree to which a scale measures what it is purported to measure. Reliability and validity are not independent: a scale cannot be valid if it is not reliable, but it can be reliable without being valid.

Reliability can be divided into two major domains: internal consistency (the degree to which scale items measure similar constructs) and rater consistency or stability (the level of rating agreement across multiple raters or in a single rater across time). The UPDRS has shown excellent internal consistency across multiple studies[13,37,38] and retains this consistency across stages of disease severity as measured by the Hoehn and Yahr

staging system.[38,39] This high degree of internal consistency may be artificially inflated due to redundancy in the large number of items in Parts II and III of the UPDRS.

Assessments of rater consistency have examined both inter-rater reliability and intra-rater reliability. Inter-rater reliability appears adequate for the total UPDRS[13,18] as well as the Activities of Daily Living[14] and the Motor Examination[30,40] sections. There are two reports of unacceptably low inter-rater reliability for selected items assessing speech and facial expression on the Motor Examination section of the UPDRS.[30,40] Other studies, however, reported acceptable inter-rater reliability estimates for these items.[13,18] There are three published reports examining intra-rater reliability. One study used the UPDRS and the other used a modified version of the scale applied to elderly community subjects without the specific diagnosis of PD.[15,40] Both of these studies showed low to medium intra-rater reliability. Among 400 early-stage PD subjects examined on two occasions, separated by approximately 2 weeks, the intraclass correlation coefficients were very high: total score, 0.92; Mentation, 0.74; Activities of Daily Living, 0.85; Motor, 0.90.[41]

Validity can be divided into three major domains: face or content validity, criterion validity, and construct validity. The UPDRS has adequate face validity and samples important and typical domains associated with PD. In addition, its construction was guided by experts in the field and based on previous scales. Criterion validity has not been established because there is no absolute "gold-standard" that can be used for this assessment. The majority of validation studies have assessed the construct validity of the UPDRS. These studies have generally found satisfactory results regarding convergent validity with other instruments assessing PD, such as the Hoehn and Yahr or Schwab and England scales or timed motor tests.[20,38,39,42] Divergent validity, or the degree to which the scale does not measure domains unrelated to PD, has not been well established. However, one study found a significant correlation between the UPDRS and measures of mental status and depression.[13] This finding may not indicate poor divergent validity, but rather the association of mental status and mood changes associated with PD.

Multiple studies have examined construct validity of the UPDRS through factor analysis. These studies have found between three and six factors that account for a significant proportion of the total scale variance.[13,18,37–39] The resultant factors form rational groupings of the items, and suggest that the scale has a valid multidimensional assessment format. One factor structure, composed of six factors, axial/gait bradykinesia, right bradykinesia, left bradykinesia, rigidity, rest tremor, and postural tremor, has been shown to be stable across *on* and *off* stages.[38,39] and to have a similar factor structure when used in other movement disorders.[25]

Additional validity studies have been conducted to assess the ability of the UPDRS to detect changes in function in either untreated or treated states. In general, these studies have demonstrated that the UPDRS is sensitive to change in clinical status.

Minimal Clinically Relevant Difference and Minimal Clinically Relevant Incremental Difference

Implicit to the strength and utility of a rating scale is the determination of increases or decreases that represent clinically relevant changes in the disease under consideration. Identifying the threshold or smallest difference between two assessments that has an impact on disability or handicap in a disease is known clinimetrically as the *Minimal Clinically Relevant Difference* (MCRD). However simple and straightforward the MCRD may be conceptually, its operational definition is difficult to establish, and very few scales are associated with a well-defined MCRD. Several factors complicate the establishment of a MCRD for the UPDRS. First, PD signs vary throughout the day in parkinsonian patients even without motor fluctuations. The natural moment-to-moment or visit-to-visit variation in the UPDRS among patients considered to be stable in overall function has not been extensively studied.[41] Second, because the four subscales of the UPDRS measure different aspects of PD and rely on physician-based and patient-based assessments, a single MCRD may not exist. Third, for a disease like PD, different MCRD values may apply at different disease severities. Specifically, a smaller MCRD may likely apply to groups of patients with mild disease, whereas a larger differential value would be expected for groups with more severe illness. MCRD has been particularly well studied for pain assessment scales and to a lesser degree in assessment measures for asthma and chronic obstructive pulmonary disease.[43–46] To date, no MCRD has been established for the UPDRS, but appropriate studies based on methods extrapolated from the pain literature could be performed with the aim of establishing MCRD score ranges with confidence intervals.[46]

The MCRD concept is applicable in two settings, individual and group. At the individual level, though a MCRD is not statistically enumerated, intervention decisions within each physician/patient relationship are guided by this concept of a minimal change from the prior visit that warrants clinical recognition. For the design of clinical trials involving groups of patients, a

uniform MCRD definition is desirable both for analyses of efficacy and futility. In some studies using the UPDRS, a 30% improvement in the Part III score has been applied to define "responders."[35] This empirically determined figure is often used in clinical medicine, based largely on the erroneous assumption that placebo effects occur in 30% of patients, regardless of disease, scale, study duration, or impairment under consideration.[47,48] The 30% UPDRS change from baseline used in clinical studies has not been experimentally derived, and furthermore, does not specifically presume to represent a minimal change of clinical significance.

One method to establish a MCRD would be to follow patients with both sequential UPDRS scores and a global estimate of change, for example, the Clinician Interview-based Impresssion of Change scale.[49,50] This scale ranges from 1 ("very much improved") to 4 ("no change") to 7 ("very much worse") relative to either a prior visit or a determined baseline. The key anchors, 3 ("minimally improved") or 2 ("moderately improved") could potentially be examined relative to the corresponding UPDRS scores to calculate a MCRD expressed as appropriate UPDRS ranges and confidence intervals. If such ratings are obtained from investigator and patient, these values could be examined against the total UPDRS as well as the specific scale sections.

Allied to the concept of MCRD is the *Minimal Clinically Relevant Incremental Difference* (MCRID). Rather than comparing two assessments within a patient or group (pre- vs. post-treatment), this terms refers to the difference between two groups at the end of a comparable period. In the case of a clinical trial, the MCRID would determine the relevant expected difference at the end of a treatment between a placebo-group and the patients receiving the treatment in question. Knowing the MCRID would allow clinicians to determine the threshold UPDRS value that would *discriminate* two treatments. So far, there is no experimentally generated or systematically analysed data on a MCRID for the total or sub-component scores of the UPDRS. There is, however, limited experience with this concept based on expert opinion or reliance on differences found in previous trials. In these cases, opinions or data widely accepted by the scientific community are used to determine estimates of an end of treatment score associated with clinically important differences in patient function between two interventions and thereby to estimate a MCRID. Among the few examples of an empirically used MCRID based on experience and literature reviews, in one pallidotomy trial that enrolled advanced PD patients with high preoperative UPDRS motor scores, a MCRID for Part III of the UPDRS motor was established at 10 points.[51] In a randomised trial that compared L-dopa to pergolide for 3 years, an a priori stopping rule for established superiority of one treatment over the other was set as a between-group difference greater than 4 points in the Part III UPDRS score at one year.[52]

Ambiguities of the UPDRS

In the context of marked strengths and wide usage of the UPDRS, a number of limitations nonetheless exist. First, as a composite scale, the UPDRS is uneven in the type of information it gathers. For example, Section I is conceptually different from parts II and III, and as a screening assessment for the presence of depression, dementia or psychosis, it cannot be used as an adequate severity measure of any of these behaviours. In cases of interventions, targeting such non-motor problems of PD, specific additional scales are generally used.[53–55] An appendix to the UPDRS with a series of recommended scales for more detailed measurement of all screening questions would enhance consistency of data collection among researchers. At present, such appendices do not exist. Likewise, section IV is constructed differently than the rest of the UPDRS with a mixture of 5-point options and dichotomous (yes/no) ratings that are difficult to analyse together. As such, though this portion is sometimes used in clinical trials, most intervention studies for dyskinesias or motor-fluctuations currently rely primarily on other scales. A number of additional dyskinesia scales have been proposed to supplement the UPDRS[56–58] and most studies of patients with motor fluctuations have used self-scoring *on-off* diaries.[34] Similarly dichotomous, yes/no questions for the presence of gastrointestinal complaints, orthostatic hypotension, or sleep problems (Part IV, items 40–42) can only be used as screening items to assess presence or absence of select clinical problems. The Task Force members considered these items as insufficient to assess severity of impairment or disability related to non-motor domains of PD.

Some items of the motor section have relatively poor inter-rater reliability, including speech, facial expression, posture, body bradykinesia, action tremor, and rigidity.[13,30,59] Although the UPDRS teaching tape for Part III provides visual anchors to improve inter-rater reliability, such tapes for the rest of the UDPRS have not been developed.[12] A specific example of a key testing problem is the assessment of postural stability in Part III. Because the response of the patient and the assigned rating depend directly on the force of the postural threat, standardized instructions and application of the test are essential for consistent ratings. These instructions are not part of the UPDRS.

Additionally, there is some redundancy of items in both the ADL and motor sections. While duplication of material enhances the internal consistency of the scale, some critics consider such enhancement a spurious inflation.[13] Redundancy also increases the time required to administer the scale. Efforts to reduce redundancy have led to the Short Parkinson's Evaluation Scale (SPES), based directly on the UPDRS, but with fewer items and reduced rating options of 0–3.[18] Whereas the elimination of redundancy and the enhancement of inter-rater reliability are overall positive goals for scales, the shrinkage of numerical options clinimetrically diminishes the capacity to discriminate change. The majority of the Task Force considered 0–4 ratings to be preferred over 0–3. Allied to duplication of information is the concern that aspects of parkinsonian motor impairment in Part III are not necessarily reflective of the impact of each cardinal feature on overall function. For example, bradykinesia-related items are overrepresented in terms of the number of assessment items in comparison to tremor and postural stability.

The allocation of items to specific sections of the UPDRS is not altogether consistent, leading to potential ambiguity of interpretation. Part II, titled "Activities of Daily Living," includes a mixture of items that directly relate to daily activities (e.g., dressing, eating), but also examine patient perceptions of primary disease manifestations (e.g., tremor, salivation). Items that overlap these two categories include the gait items that assess primary parkinsonian features (freezing, falls), but impact on walking as an activity of daily living. Renaming Part II as "Historical" or "Patient Perceptions" would semantically, but not conceptually, resolve this ambiguity. Items assessing function outside activities of daily living could alternatively be reassigned to another section of the scale.

The UPDRS Part II is culturally biased, and the anchoring descriptions for some item ratings are not applicable to all ethnic environments. For example, Dressing (Item 10) describes difficulty with buttons, even though many traditional cultures do not use them; Cutting Food/Handling Utensils (Item 9) presumes that food is regularly cut for eating and that utensils are used, although some cultures serve food in bite-size portions and some do not use eating utensils. Although the scale was considered applicable to most international urban settings, the UPDRS may be limited by ambiguities when applied in epidemiological research efforts that involve field work to rural and geographically isolated cultures. Even within Western cultures, the UPDRS has been primarily used in studying Caucasians with Parkinson's disease, and it has not been examined extensively in other ethnic or racial groups.[36]

Comorbidities and the UPDRS

PD is more prevalent in subjects over age 50 years, and the co-existence of other diseases like diabetes, stroke, and arthritis can confound the evaluation of PD-related impairment and disability. Furthermore, short-term disabilities resulting from a fracture or the exacerbation of rheumatic disease may increase overall disability without altering PD severity itself. Conversely, correction of cataracts may improve overall patient function and facilitate the execution of activities of daily living without directly affecting PD. Finally, common co-existent disorders like depression can potentially affect the speed of a patient's movement, alter motivation, and enhance perceptions of disability even when PD itself is stable. The question of how the UPDRS should accommodate these various issues of co-morbidities is not specifically detailed in the scale instructions. Two different prototypic paradigms could be used: the first, a concerted attempt to disregard all components of impairment or disability due to conditions unrelated to PD, using the UPDRS as a scale of PD-related dysfunction in its purest sense; a second strategy involves a "rate-as-you-see" approach, using the UPDRS to describe a patient's functional impairment regardless of the direct relationship to PD. The first approach has the advantage of focusing on PD itself, but it is highly subject to investigator and patient bias. For accurate assessment of Part III, the rater would need to have a list of co-morbidities to maximize appropriate interpretation of signs. The second strategy deals with the reality of the patient status without interpretation of underlying cause for impairment, but will likely inflate ratings that have minimal or no direct relationship to PD. Standardized instructions for rating PD in the context of co-morbidities do not exist in the current UPDRS, and the Task Force members agreed that clarification of data handling for co-morbidities would be an important asset of a future scale modification. The rate-as-you-see method for Part III (Motor Examination) would likely reduce inter-rater variability, but the inclusion of "open fields" in the margins of the scale document for raters to note the contribution of other medical conditions would be needed for full interpretation. Lang[60] suggested open-field marginal denotations to indicate when dyskinesias (D), excessive parkinsonian tremor (T), or apraxia (A) confound the execution of rated tasks. Similar denotations could be included in the margins of the scale for confounding co-morbidities that cause weakness, orthopedic, or non-parkinsonian coordination deficits.

TABLE 1. *Items not covered by the Unified Parkinson's Disease Rating Scale*

Item
Anhedonia
Bradyphrenia
Anxiety
Hypersexuality
Sleep disorders (insomnia, excessive daytime sleepiness)
Fatigue
Dysautonomia (urinary dysfunction, constipation, impotence, sweating)
Dysregulation
Health-related quality of life

Whereas these denotations cannot be handled in a simple statistical manner, they would be potentially useful in clinical practice.

In contrast, Part II (Activities of Daily Living) has instructions asking patients specifically to rate disability that *they* attribute to PD.[1] The Task Force members acknowledged that most individual patients are comfortable with this introspective process, although patient-bias may be unstable over time and subject to educational efforts and patient experience that clarify the contribution of other illnesses to overall disability.

In clinical trials, the problems related to co-morbidities and their impact on UPDRS scores can be minimized by excluding patients with medical conditions that confound interpretation of the primary rating measure. Alternatively, some studies permit co-morbid conditions, but only when they are stable and of long duration. If these conditions do not overly elevate the baseline scores and introduce concerns of "ceiling effects," the UPDRS can still measure change from the intervention under study.

Important Elements Not Covered

Several key elements of PD are not covered by the UPDRS. When the scale was formulated in the mid-1980s, the developers were well aware of this limitation, but they made choices to delete questions on some parkinsonian impairments, mainly to create a scale that was reasonably simple and short (S. Fahn, personal communication). After more than a decade of scale utilization, however, the Task Force members considered that these initial choices should be re-considered. In view of the anatomical, neurochemical, physiological, and conceptual evolution in thinking about PD, clinical neuroscientists may currently need to have a scale that adequately reflects the multifaceted elements of PD and assesses additional non-motor symptoms and signs that contribute to disability and quality of life. Several areas of concern exist (Table I), though changes in all items may be impractical if the UPDRS is to remain a standard scale to be applied in both clinical and research domains. The concept that screening questions could cover these topics was favored, but an appendix should be added to the UPDRS that would include officially recognized scales to assess each of the screened areas in further depth.[61] Another option would be the development of multiple UPDRS versions of different lengths and different levels of comprehensiveness, leaving the choice of scale to the physician (daily practice, in-depth evaluation, clinical trials).[62] Multiple UPDRS versions, however, would have the potential to cause reporting ambiguity and undermine the "unified" concept that anchors the UPDRS. The Task Force members favored the maintenance of a single UPDRS that has sufficient screening questions to capture problems related to all aspects of PD with official appendices that recommend scales to assess each of the screened areas in further depth when needed.

CONCLUSIONS AND RECOMMENDATIONS

The UPDRS is the most widely used clinical rating scale for PD. The data cited in this critique highlight the well-established and respected status of the UPDRS within the movement disorders, scientific, and regulatory communities. The Task Force members unanimously considered the concept of a single clinical rating scale to be an important tool for clear and consistent communication among movement disorder colleagues. The strengths of the UPDRS are many, and the scale provides a relatively comprehensive assessment of motoric aspects of PD. Extensive clinimetric analyses have already been conducted on the UPDRS, providing it both scientific and clinical credibility. Some items, however, have poor inter-rater reliability, and individual text anchors or instructions for data acquisition are ambiguous. The UPDRS is less comprehensive in its assessment of non-motor features of the disease.

The Task Force members agreed that the identified weaknesses were substantive, but amenable to correction. They could be reduced or eliminated if the scale was retained in its basic structure and core elements, but modified and expanded to reflect the growing knowledge of PD-related impairments. Modifications should focus on clarity, resolution of ambiguity, and also provide ratings for the very mild impairments seen in early PD. These changes were considered important for optimising PD evaluation across the spectrum of clinical severities. Expansion, however, should be limited to essential items that are currently not covered by the UPDRS, so that the scale would retain its ready utility in both clinical and research settings. It was emphasized that any change in the place of a given existing item, any modification in the

content or definition of a given item, or any addition of a new item would necessarily require new validation testing. Although a new official version of the UPDRS would compromise already operative longitudinal studies based on the current UPDRS version, statistical methods readily exist to compare the current version and a new version and could be conducted for total, section, and individual item scores to address this dilemma. If developed, the new version could be tested in the context of two particular scale issues that have not been studied with the current UPDRS version: 1) the definition of UPDRS scores that define clinically pertinent categories, "minimal," "mild," "moderate," and "severe" PD; and 2) the definition of UPDRS scores for a MCRD and MCRID. Based on these considerations, the Task Force on Rating Scales in PD made the following recommendations:

1. The Movement Disorder Society should organize and sponsor an official new version of the UPDRS.
2. The new version should:
   a. Retain the basic structure of the current UPDRS, with sections that include both physician-derived and patient-derived data and retain at least 0–4 ratings for motor items assessing severity of impairment or disability.
   b. Provide specific instructions for the acquisition of data on each item.
   c. Eliminate ambiguities in all text anchors.
   d. Cover the full spectrum of PD, especially providing ratings for mild parkinsonism.
   e. Be responsive to concerns of cultural bias.
   f. Have specific instructions on a uniform manner of dealing with co-morbidities.
   g. Better characterize and rate dyskinesias, capturing peak-dose and end of dose dyskinesia, both choreic and dystonic forms.
   h. Include additional screening questions on uncovered non-motor items that impact on the overall function of PD subjects.
   i. Incorporate an appendix that lists officially recommended additional scales applicable to each screening question. These optionally used scales would assess severity of impairment related to non-motor elements of PD and be applied when the screening questions identify the presence of deficits in a specific function.
   j. Be tested by factor analysis to determine core impairments that potentially could lead to a reduced number of items that would serve to form a short version of the scale for everyday clinical practice.
3. The MDS should encourage efforts to conduct comprehensive clinimetric testing of the new scale to include not only standard assessments of reliability and validity, but also:
   a. Analysis of the new scale's correlation with the current UPDRS.
   b. Tests to establish UPDRS score ranges with confidence intervals that define clinically pertinent categories, such as "minimal," "mild," "moderate," and "severe" PD.
   c. Evaluations of MCRD and MCRID.
   d. Assessments of the scale's validity across race, gender, and age.
4. If the International Executive Committee designates a panel to develop a new version of the UPDRS, the panel of experts selected for this project should include movement disorder specialists that represent academic and practice settings and statistical experts. All current members of the Task Force would be willing to serve on this panel. The development and testing of proposed changes should include input from the entire MDS membership.
5. Until a new UPDRS version is developed, clinical and research efforts should retain primary reliance on the UPDRS version 3.0, which is still considered an overall strong and useful assessment tool for evaluating impairment and disability in PD.

APPENDIX 1

Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease

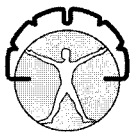
Chairperson: Christopher G. Goetz, MD

Writing Committee: Werner Poewe, MD, Olivier Rascol, MD, Cristina Sampaio, MD, and Glenn T. Stebbins, PhD

Expert Consultants: Stanley Fahn, MD, Anthony E. Lang, MD, Pablo Martinez-Martin, MD, PhD, Barbara Tilley, PhD, and Bob Van Hilten, MD, PhD

MDS Secretariat staff: Caley Kleczka, Lisa Seidl

APPENDIX 2



THE *MOVEMENT* DISORDER SOCIETY

To all MDS Members:

The MDS has initiated a Task Force on the Development of Rating Scales for Parkinson's Disease. The first charge of this task force will be a critique of the Unified Parkinson's Disease Rating Scale (UPDRS). Of pivotal interest is the frequency and manner in which MDS members use the scale. Please take a few minutes to complete the questionnaire below and return it so that your experience can be incorporated into the assessment. Thank you for participating in this important MDS initiative.

Sincerely,

Christopher Goetz (Chair), Werner Poewe, Olivier Rascol, Cristina Sampaio
MDS Task Force on the Development of Rating Scales for Parkinson's Disease

---

QUESTIONNAIRE ON UTILIZATION OF THE UNIFIED PARKINSON'S DISEASE RATING SCALE (UPDRS)

1. For the following questions, please check yes or no:
   - a. Do you have personal experience using the UPDRS? ❑ Yes ❑ No
   - b. Do you use the UPDRS as part of clinical practice? ❑ Yes ❑ No
   - c. Do you use the UPDRS as part of clinical trials? ❑ Yes ❑ No
   - d. Do you use the UPDRS as part of other research? ❑ Yes ❑ No
   - e. Do you use the UPDRS as part of a Database? ❑ Yes ❑ No

If you use the UPDRS for any reason, please continue; if not, go to question 4.

2. If you use the UPDRS, do you regularly use the following:
   - a. Part I (mentation)? ❑ Yes ❑ No
   - b. Part II (ADL)? ❑ Yes ❑ No
   - c. Part III (Motor)? ❑ Yes ❑ No
   - d. Part IV (Complications)? ❑ Yes ❑ No
   - e. Total sum of I+II+III? ❑ Yes ❑ No
   - f. Total sum of II+III? ❑ Yes ❑ No

3. For those situations where you use the UPDRS, please rate its utility by completing this table with one of the following in each space.

NA = not applicable, I do not use in this situation
VU = very useful in this situation
U = useful in this situation
SL = slightly useful in this situation
NU = not useful in this situation

UPDRS	Part I	Part II	Part III	Sum I+II+III	Sum II+III	Part IV
Clinical Practice						
Clinical Trials						
Other Research						
Building a Database						
Other Uses (please specify)						

4. If you do not use the UPDRS, what is the primary method by which you rate motor impairment and disability in your PD patients? (Please check yes or no)
   - a. Do you use another rating scale? ❑ Yes ❑ No If yes, please explain: _____
   - b. Do you use written descriptions of impairment and disability? ❑ Yes ❑ No If yes, please explain: _____
   - c. Do you use another method? ❑ Yes ❑ No If yes, please explain: _____

APPENDIX 3

MDS Task Force for the Development of Rating Scales for Parkinson's Disease

EXPERT CONSULTANT QUESTIONNAIRE

(If more space is required, please attach additional pages.)

1. What are the strengths of the UPDRS? Comments may be general and specific. Make certain that data are cited in support for each advantage discussed.

2. What are the weaknesses? Comments may be general and specific. Make certain that data are cited in support for each advantage discussed.

3. What clinimetric criteria have been met by the UPDRS; reliability, validity, capacity to detect change from interventions, others? Again, cite data. What is the quality of these data? On which version were the various tests performed? If the criteria were met for a version that is different than the one currently used, are those data on earlier versions applicable?

4. A key concept in defining an intervention response is the Minimally Clinically Relevant Incremental Difference (MCRID), meaning the change score by number or % in a scale that represents a clinically meaningful effect. Are there data to provide this measure for the UPDRS?

5. What is your concept of the UPDRS relative to comorbidities? Do you consider this scale an index of parkinsonism, so one would rate the findings relative to PD and for all parts (I-IV) exclude the impact of other conditions unrelated to PD? Or, do you rate the findings purely as you see them and therefore incorporate impairment or disability from other conditions (orthopedic, rheumatological, even psychiatric—anxiety etc.)? For example, in testing finger taps, should the rater ignore the impact of problems like weakness and slowness from arthritis that could elevate the rating or should the clinician rate just what is seen according to the scale's descriptive

APPENDIX 3 (CONTINUED)

anchors? If you are a clinician, comment on the impact of this decision in terms of rating PD severity and changes after an intervention. If you are a statistician, comment on the statistical issues of either decision as part of your response.

6. Using your knowledge base and perspective (clinical neurologist, statistician), summarize and critique the data related to factor analysis, issues of repetitive data, items that are particularly strong and particularly weak from a clinimetric perspective.

7. How successful is the UPDRS rating impairment and disability across the clinical spectrum of PD? Mild, moderate, and advanced disease?

8. Are there key elements of parkinsonian impairment and disability that are not covered in the UPDRS?

9. If you are a clinician, what version of the UPDRS do you use currently? _____

10. Based on your experience, review of the literature, and all data you have received, do you favor continuing the use the UPDRS as the standard PD rating scale for impairment and disability?

REFERENCES

1. Fahn S, Elton RL, UPDRS program members. Unified Parkinsons Disease Rating Scale. In: Fahn S, Marsden CD, Goldstein M, Calne DB, editors. Recent developments in Parkinsons disease, vol 2. Florham Park, NJ: Macmillan Healthcare Information; 1987. p 153–163.
2. Webster DD. Critical analysis of the disability in Parkinson's disease. Mod Treat 1968;5:257–258.

3. Duvoisin RC. The evaluation of extrapyramidal disease. In: de Ajuriaguerra J, Gauthier G, editors. Monoamines noyaux gris centraux et syndrome de Parkinson. Paris: Masson; 1971.
4. Parkes JD, Zilka KJ, Calver DM, Knill-Jones RP. Controlled trial of amantadine hydrochloride in Parkinson's disease. Lancet 1970; 1:259–262.
5. Canter CJ, de la Torre R, Mier M. A method of evaluating disability in patients with Parkinson's disease. J Nerv Ment Dis 1961;133:143–147.
6. Lieberman A, Le Brun Y, Boal D. The use of a dopaminergic receptor stimulating agent (Piribendil ET-495) in Parkinson's disease. In: Usdin E, editor. Advances in biochemical spychopharmacology, Vol 12. New York: Raven Press, 1973; p 415–425.
7. Diamond SG, Markham CH, Treciokas LJ. A double-blind comparison of levodopa, madopar and Sinemet in Parkinson's disease. Ann Neurol 1978;3:267–272.
8. Schwab JF, England AC. Projection technique for evaluating surgery in Parkinson's disease. In: Billingham FH, Donaldson MC, editors. Third Symposium on Parkinson's Disease. Edinburgh: Livingstone; 1968. p 152–157.
9. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. Neurology 1967;17:427–442.
10. Ramaker C, Marinus J, Stiggelbout AM, et al. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease: a systematic review. Mov Disord 2002;17:867–876.
11. Mitchell SL, Harper DW, Bhalla R. Patterns of outcome measurement in Parkinson's disease clinical trials. Neuroepidemiology 2000;19:100–108.
12. Goetz CG, Stebbins GT, Chmura TA, et al. Teaching tape for the motor section of the unified Parkinson's disease rating scale. Mov Disord 1995;10:263–266.
13. Martinez-Martin P, Gil-Nagel A, Gracia LM, et al. Unified Parkinson's disease Rating Scale characteristics and structure. The Cooperative Multicentric Group. Mov Disord 1994;9:76–83.
14. Louis ED, Lynch T, Marder K, et al. Reliability of patient completion of the historical section of the Unified Parkinson's disease Rating Scale. Mov Disord 1996;11:185–192.
15. Bennett DA, , Shannon KM, Beckett LA, Goetz CG, Wilson RS. Metric properties of nurses' ratings of parkinsonian signs with a modified UPDRS. Neurology 1997;49:1580–1587.
16. Morrish PK, Sawle GV, Brooks DJ. An [18F]dopa-PET and clinical study of the rate of progression in Parkinson's disease. Brain 1996;119:585–591.
17. Seibyl JP, Marek KL, Quinlan D, et al. Decreased single-photon emission computed tomographic [123I]beta-CIT striatal uptake correlates with symptom severity in Parkinson's disease. Ann Neurol 1995;38:589–598.
18. Rabey JM, Bass H, Bonuccelli U, Brooks D, 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 Neuropharm 1997;20:322–337.
19. Hogan T, Grimaldi R, Dingemanse J, et al. The Parkinson's symptom inventory (PDSI): a comprehensive and sensitive instrument to measure disease symptoms and treatment side-effects. Parkinson Relat Disord 1999;5:93–98.
20. Martinez-Martin P, Garcia Urra D, del Ser Quijano T, et al. A new clinical tool for gait evaluation in Parkinson's disease. Clin Neuropharm 1997;20:183–194.
21. Committee for proprietary medicinal products. Guidance on clinical investigation of medicinal products in the treatment of Parkinson's disease. 2002. Online at http://www.emea.eu.int/pdfs/human/ewp/056395en.pdf.
22. Goetz CG, Leurgans S, Raman R et al. Placebo-associated improvements in motor function: comparison of subjective and objective sections of the UPDRS in early Parkinson's disease. Mov Disord 2002;17:283–288.
23. Langston JW, Widner H, Goetz CG, et al. Core assessment program for intracerebral transplantations (CAPIT). Mov Disord 1992;7:2–13.
24. Defer GL, Widner H, Marie RM, et al. Core assessment program for surgical interventional therapies in Parkinson's disease (CAPSIT-PD). Mov Disord 1999;14:572–84.
25. Cubo E, Stebbins GT, Golbe LI, Nieves AV 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.
26. Ballard C, McKeith I, Burn D, et al. The UPDRS scale as a means of identifying extrapyramidal signs in patients suffering with dementia with Lewy bodies. Acta Neurol Scand 1997;96:366–371.
27. Parkinson Study Group. Effects of tocopherol and deprenyl on the progression of disability in early Parkinson's disease. N Engl J Med 1993;328:176–183.
28. Lieberman A, Ranhosky A, Korts D. Clinical evaluation of pramipexole in advance PD. Neurology 1997;49:162–168.
29. Vieregge P, Stolze H, Klein C, Heberlein I. Gait quantification in Parkinson's disease: locomotor disability and correlation to clinical rating scales. J Neural Transm 1997;104:237–248.
30. Richards M, Marder K, Cote L, et al. Interrater reliability of the Unified Parkinson's disease Rating Scale motor examination. Mov Disord 1994;9:89–91.
31. McDermott MP, Jankovic J, Carter J, et al. Factors predictive of the need for levodopa therapy in early, untreated Parkinson's disease. Arch Neurol 1995;52:565–570.
32. Poewe WH, Wenning GK. The natural history of Parkinson's disease. Neurology 1996;47 (Suppl. 3):S146–S152.
33. Oakes D and the Parkinson Study Group. Progression of early Parkinson's disease until levodopa therapy is required. Neurology 1990;40(Suppl. 1):370.
34. Rascol O, Goetz C, Koller W, et al. Treatment interventions for Parkinson's disease: an evidence based assessment. Lancet 2002; 359:1589–1598.
35. Adler CH, Sethi KD, Hauser RA, et al. Ropinirole for the treatment of early PD. Neurology 1997;49:393–399.
36. Tanner CM, Parkinson Study Group. Pramipexole in ethnic minority with Parkinson's disease. Mov Disord 1999;14(Suppl. 5): 901.
37. Van Hilten JJ, van der Zwan AD, Zwinderman AH, et al. Rating impairment and disability in Parkinson's disease: evaluation of the Unified Parkinson's disease Rating Scale. Mov Disord 1994;9:84–88.
38. Stebbins GT, Goetz CG. Factor structure of the Unified Parkinson's disease Rating Scale: Motor Examination section. Mov Disord 1998;13:633–636
39. Stebbins GT, Goetz CG, Lang AE, et al. Factor analysis of the motor section of the Unified Parkinson's disease Rating Scale during the off-state. Mov Disord 1999;14:585589
40. Camicioli R, Grossmann SJ, Spencer PS, et al. Discriminating mild parkinsonism: methods for epidemiological research. Mov Disord 2001;16:33–40.
41. Siderowf A, McDermott M, Kieburtz K, et al. Test-retest reliability of the UPDRS in patients with early Parkinson's disease: results of a multicenter clinical trial. Mov Disord 2002;17:758–763.
42. Martinez-Martin P, Fontan C, Frades Payo B, et al. Parkinson's disease: quantification of disability based on the Unified Parkinson's disease Rating Scale. Neurologia 2000;15:382–387.
43. Todd KH, Funk KG, Funk JP, et al. Clinical significance of reported changes in pain severity. Ann Emerg Med 1996;27:485–489.
44. Gallagher EJ, Liebman M, Bijur PE. Prospective validation of clinically important changes in pain severity measured on a visual analog scale. Ann Emerg Med 2001;38:633–638.
45. Bird SB, Dickson EW. Clinically significant changes in pain along the visual analog scale. Ann Emerg Med 2001;38:639–643.
46. Jones PW. Interpreting thresholds for a clinically significant change in health status in asthma and COPD. Eur Respir J 2002; 19:398–404.

47. Price DD, Milling LS, Kirsch I, et al. An analysis of factors that contribute to the magnitude of placebo analgesia in an experimental paradigm. Pain 1999; 83:147–156
48. Mattocks KM, Horwitz RI. Placebos, Active Control Groups, and the unpredictability Paradox. Biol Psychiatry 2000;47:693–698.
49. Oremus M, Perrault A, Demers L, et al. Review of outcome measurement instruments in Alzheimer's disease drug trials: psychometric properties of global scales. J Geriatr Psychiatry Neurol 2000;13:197–205.
50. Knopman DS, Knapp MJ, Gracon SI, et al. The Clinician Interview-based Impression. Neurology 1994;44:2315–2321.
51. de Bie RMA, de Haan RJ, Nijssen PCG, et al. Unilateral pallidotomy in Parkinson's disease: a randomised, single-blind , multicentre trial. Lancet 1999;354:1665–1669.
52. Oertel WH et al. Pergolide versus levodopa monotherapy (PELMOPET) [abstract]. Mov Disord 2000;15(Suppl. 3):4.
53. Parkinson Study Group. Low-dose clozapine for the treatment of drug-induced psychosis in Parkinson's disease. N Engl J Med 1999;340:757–763.
54. Goetz CG, Blasucci LM, Leurgans S, et al. Olanzapine and clozapine: comparative effects on motor function in hallucinating PD patients. Neurology 2000;55:789–794.
55. Giladi N, Shabtai H, Benbunan B, et al. The effect of treatment with rivastigmin (Exelon) on cognitive functions of patients with dementia and Parkinson's disease. Neurology 2001;56:A128.
56. Luquin MR, Scipioni O, Vaamonde J, et al. Levodopa-induced dyskinesias in Parkinson's disease: clinical and pharmacological classification. Mov Disord 1992;7:117–124.
57. Marconi R, Lefebvre-Caparros D, Bonnet AM, et al. Levodopa-induced dyskinesias in Parkinson's disease, phenomenology and pathophysiology. Mov Disord 1994;9:2–12.
58. Goetz CG. Rating scales for dyskinesias in Parkinson's disease. Mov Disord 1999;14:48–53.
59. Prochazka A, Bennett DJ, Stephens MJ, et al. Measurement of rigidity in Parkinson's disease. Mov Disord 1997;12:24–32.
60. Lang AE. Teaching tape for the motor section of the Unified Parkinson's Disease Rating Scale. Mov Disord 1996;11:344–345.
61. Marinus J, Ramaker C, Van Hilten JJ, et al. Health related quality of life in Parkinson's disease: a systematic review of disease specific instruments. J Neurol Neurosurg Psychiatry 2002;72:241–248.
62. Martinez-Martin P. Rating scales in Parkinson's disease. In: Jankovic J, Tolosa E, editors. Parkinson's disease and movement disorders, 2nd ed. Baltimore: Williams and Wilkins; 1993. p 281–229.