Movement Disorders
Vol. 19, No. 9, 2004, pp. 1020–1028

Movement Disorder Society Task Force Report on the Hoehn and Yahr Staging Scale: Status and Recommendations

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

Christopher G. Goetz,[1]* Werner Poewe,[2] Olivier Rascol,[2] Cristina Sampaio,[2] Glenn T. Stebbins,[2] Carl Counsell,[3] Nir Giladi,[3] Robert G. Holloway,[3] Charity G. Moore,[3] Gregor K. Wenning,[3] Melvin D. Yahr,[3†] and Lisa Seidl[4]

[1]*Chairperson;* [2]*Writing Committee;* [3]*Expert Consultant;* [4]Movement *Disorder Society Secretariat Staff*

Abstract: The *Movement* Disorder Society Task Force for Rating Scales for Parkinson's disease (PD) prepared a critique of the Hoehn and Yahr scale (HY). Strengths of the HY scale include its wide utilization and acceptance. Progressively higher stages correlate with neuroimaging studies of dopaminergic loss, and high correlations exist between the HY scale and some standardized scales of motor impairment, disability, and quality of life. Weaknesses include the scale's mixing of impairment and disability and its non-linearity. Because the HY scale is weighted heavily toward postural instability as the primary index of disease severity, it does not capture completely impairments or disability from other motor features of PD and gives no information on nonmotor problems. Direct clinimetric testing of the HY scale has been very limited, but the scale fulfills at least some criteria for reliability and validity, especially for the midranges of the scale (Stages 2–4). Although a "modified HY scale" that includes 0.5 increments has been adopted widely, no clinimetric data are available on this adaptation. The Task Force recommends that: (1) the HY scale be used in its original form for demographic presentation of patient groups; (2) when the HY scale is used for group description, medians and ranges should be reported and analysis of changes should use nonparametric methods; (3) in research settings, the HY scale is useful primarily for defining inclusion/exclusion criteria; (4) to retain simplicity, clinicians should "rate what you see" and therefore incorporate comorbidities when assigning a HY stage; and (5) because of the wide usage of the modified HY scale with 0.5 increments, this adaptation warrants clinimetric testing. Without such testing, however, the original five-point scales should be maintained.

Key words: Hoehn and Yahr scale; Parkinson's disease; rating scales; clinimetrics

The *Movement* Disorder Society (MDS) Task Force for Rating Scales in PD was established in 2001 and has a three-fold mission: to critique existing scales, to identify clinical areas that are not rated adequately, and to make recommendations on maintaining or modifying currently available scales. Specifically, the charge does not include changing any existing scales or developing new scales. The first critique concerned the Unified Parkinson's Disease Rating Scale (UPDRS) with an evaluation of its strengths and weaknesses, along with recommendations for future adaptations.[1] The current report is a critique of similar format, but focuses on the Hoehn and Yahr (HY) staging scale (see Fig. 1).[2]

Whereas the HY scale has been adapted for many different uses and even applied to disorders other than PD, it was designed originally to be a simple descriptive staging scale that provided a general estimate of clinical function in PD, combining functional deficits (disability) and objective signs (impairment). Though over 30 years old and developed originally in the pre-levodopa (1-dopa) era, the scale has continued to be used widely. Originally designed as a five-point scale (1–5), during the 1990s, 0.5 increments were introduced for some clinical trials (Table 1).[3] The scale is based on the two-

*Correspondence to: Christopher G. Goetz, Rush University Medical Center, 1725 W. Harrison Street, Chicago, IL 60612.
E-mail: cgoetz@rush.edu
†Deceased.
Received 20 January 2004; Revised 18 March 2004; Accepted 21 March 2004
Published online 16 June 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/mds.20213

FIGURE 1. Melvin D. Yahr and Margaret M. Hoehn, 2004. Photo courtesy of M.M. Hoehn.

fold concept that the severity of overall parkinsonian dysfunction relates to bilateral motor involvement and compromised balance/gait. Increasing parkinsonian motor impairment therefore can be charted from unilateral (Stage 1) to bilateral disease (Stage 2) without balance difficulties, to the presence of postural instability (Stage 3), loss of physical independence (Stage 4), and being wheelchair- or bed-bound (Stage 5). In this simple scale that staged patients at their current level of function, the original authors never presumed or suggested that patients generally started PD as Stage I and declined sequentially to Stage 5 or death.

As part of the background work to develop the Task Force on Rating Scales for PD, the MDS secretariat staff (see Acknowledgments) sent a questionnaire on HY scale utilization patterns to all MDS members. This questionnaire was an adaptation of one sent previously to all members regarding usage of the UPDRS.[1] Of 1,593 (1,405 e-mails and 188 fax communications), 236 members from all continents responded (15%). Of those who responded, 96% had personal experience with the HY scale, 89% using it in clinical trials, 70% in clinical practice, and 63% in other research venues. Whereas the officially published scale has integer options of 1 to 5 only, 64% of those using the scale actually chose the modified form that includes 0.5 increments. Despite the wide usage, 65% of responders considered that the HY scale did not adequately describe their patients; specifically, 76% responded that the staging categories did not encompass all components of PD disability, 69% considered the staging categories to be too broad, and 65% were concerned that progression from one category to the next was not linear through all stages. Most responders who did not use the HY scale but used other scales for rating PD favored the UPDRS, and fewer numbers used the Northwestern University Disability, Webster, and Clinical Global Impression scales.[4–7] Although the questionnaire results document the usage pattern of over 200 people specifically interested in movement disorders, the low percentage of responders limits the ability to extrapolate the information to the entire MDS. Nonetheless, the observation that responders use the HY scale for multiple purposes, including research and clinical practice, confirmed the decision to prioritize the HY scale as the second assessment project of the MDS Task Force.

MATERIALS AND METHODS

Administrative Organization

Under the chairperson's direction, a Writing Committee was identified to draft the HY 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 HY scale 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

TABLE 1. *Comparison between the original and modified Hoehn and Yahr scale*

Hoehn and Yahr scale	Modified Hoehn and Yahr scale
1: Unilateral involvement only usually with minimal or no functional disability	1.0: Unilateral involvement only
	1.5: Unilateral and axial involvement
2: Bilateral or midline involvement without impairment of balance	2.0: Bilateral involvement without impairment of balance
	2.5: Mild bilateral disease with recovery on pull test
3: Bilateral disease: mild to moderate disability with impaired postural reflexes; physically independent[a]	3.0: Mild to moderate bilateral disease; some postural instability; physically independent
4: Severely disabling disease; still able to walk or stand unassisted	4.0: Severe disability; still able to walk or stand unassisted
5: Confinement to bed or wheelchair unless aided	5.0: Wheelchair bound or bedridden unless aided

[a]Stage 3 is a summary of the authors' original, more narrative description.

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 HY scale literature, the chairperson supplied each Task Force member with a series of articles related to the HY scale. 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. The questions followed the same format used previously for the MDS-sponsored UPDRS critique.[1] No unpublished data were used for the critique. 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 before submission for journal peer-review in *Movement Disorders*.

RESULTS

Strengths of the HY Scale

Ever since its introduction, the HY scale has remained the most commonly and most widely used scale to describe severity of PD worldwide.[8] It is the standard staging system used to describe patient populations enrolled in clinical trials of antiparkinsonian interventions and the second most frequently used outcome measure after the UPDRS in all randomly ordered drug trials for PD published between 1966 and 1998.[8] It provides an overall assessment of severity based on clinical features and functional disability.[9] Easy to apply, quick to complete, and practical to research and patient care settings, the HY scale has been successfully used by raters without movement disorder expertise as well as specialists.[10] Other clinimetric strengths are discussed below.

Given its historical stature, the HY scale has been used commonly as a reference standard for testing more recently developed rating scales for impairment and disability in PD.[11] There are significant correlations between the HY stage of PD and measures of quality of life, both with respect to general health-based scales[12] as well as with disease-specific instruments like the PD questionnaire-39 (PDQ-39)[13] and the PD Quality of Life scale (PDQUALIF).[14] Studies with both types of scales have found worse quality of life with more advanced HY stages.[14] Studies of objective and quantitative motor impairment tests and objective assessments of tasks involved in daily living have identified significant correlations between objective motor performance and early versus late HY stages.[15,16]

The HY scale has an inconsistent relationship with self-care measures, some disability ratings, and the Webster score,[17] but high Spearman's correlations with standard PD rating scales like the UPDRS, Columbia University Rating Scale, the Northwestern University Disability scale, and the Extensive Disability Scale.[18,19] In longitudinal studies, the HY scale progresses as positron emission tomography (PET) indices of dopaminergic activity decline.[20]

The HY scale has been used extensively in natural history studies of PD and for the description of large populations of PD patients. Studies from both the pre-L-dopa and L-dopa eras involving large cohorts of PD patients have found similar percentages of cases assigned to the different stages of the HY scale.[2,15,21–23] In these studies, Stage 1 and Stage 5 account for the smallest number of subjects, followed by Stage 4, and the bulk of patients, ranging from 52 to 77%, fall into Stages 2 or 3. In the pre-L-dopa era, longitudinal follow-up of cohorts documented an ever-growing progression toward higher stages or death.[2,21,22]

The time course of progression from early to late stages of the HY scale has been modified by L-dopa, the primary pharmacologic treatment for PD. Treatment prolongs latencies to successive stages by about 3 to 5 years. Furthermore, in one study, L-dopa treatment reduced the number of patients in Stage 4/5 (or death) per 5-year period of disease duration by 30 to 50%.[22] A patient's change in HY stage carries prognostic significance and influences clinician-based interventions.[24] Progression to Stage 3 was the main reason for initiation of L-dopa in a series of 100 consecutive untreated patients.[25]

Stage 3 is defined conceptually by the emergence of balance problems, but has been found to carry further prognostic implications in a number of studies. Goetz and colleagues[26] found that Stage 2 subjects maintained stable UPDRS scores with increased medication doses, whereas once Stage 3 developed, UPDRS impairment scores increased despite of medication adjustments. Reaching HY Stage 3 is associated with subsequent higher risk of dementia in one study and reduced survival.[27] Progression of patients on the HY scale was correlated with decreased performance on cognitive test batteries in a series of nondemented patients in another study.[28] A functional neuroimaging study using β-CIT-SPECT as a marker of progressive nigrostriatal terminal dysfunction over time found significant correlations be-

tween the degree of reduction of dopamine transporter (DAT)-ligand binding and HY Stages 1 to 3.[29]

Taken together, these findings strongly support the utility of the HY scale to categorize PD subjects and to capture important aspects of PD progression.

Weaknesses of the HY Scale: Ambiguities and Limitations

Whereas most PD scales focus on measuring impairment (objective signs on examination) or disability (impact of disease on the patient), the HY scale merges the two phenomena, introducing a potential for ambiguity. For example, Stage 1 is defined as "unilateral involvement only, usually with minimal or no functional impairment," thereby combining two concepts that are not necessarily equivalent. Because of the wording, unilaterality is the hallmark of Stage 1, and must therefore encompass even subjects who are unemployable because of severe unilateral tremor and bradykinesia of the dominant hand. Formulated at a time that preceded the full development of clinimetric expertise for scale development, the HY scale is inconsistent in several other descriptive anchors. Stage 2 is defined by the lack of "impairment of balance," but this wording is not in precise parallel language to the descriptive words defining Stage 1. Furthermore, the progressive stages of the HY scale are based on two different indices of severity, unilateral versus bilateral signs and absence or presence of gait and balance impairments. As such, the HY scale is a categorical scale that describes clinical status, but each increment does not necessarily represent a higher degree of overall motor dysfunction. Whereas the progression from Stage 1 to Stage 5 in most clinical instances reflects progressive global decline, the scale clearly does not capture several clinical situations where disability from the two indices of focus do not coincide. Already mentioned are the patients with unilateral PD who must be assigned Stage 1, even those who are highly disabled and unable to hold a job because of marked, unilateral tremor and severe bradykinesia of the dominant hand. These patients may be markedly more disabled than subjects who are Stage 2 or even Stage 3 with bilateral signs. Stage 3 must be assigned to patients with postural reflex impairment even if they experience only very mild bilateral bradykinesia and tremor, whereas a lower rating (Stage 2) is given to subjects with severe bradykinesia, rigidity, and tremor without balance difficulties. The scale's non-linearity is exemplified by the reality that a unilaterally affected PD subject (Stage 1) who develops postural instability before developing bilateral signs in his extremities must be rated as Stage 3, having never been Stage 2. Clearly, many patients present for their first medical evaluation and are already Stage 2. The observation suggests for such patients that Stage 1 is not clinically pertinent or that patients actually begin PD at Stage 2. Whereas all these possibilities exist and confound the clarity of the HY scale, the high correlations between HY and other indices of function across all stages suggest that such problems are not sufficiently common enough to impact seriously on the scale when applied to large populations.

A second limitation is that the scale has only five options and therefore a large variety of impairment severities is collapsed together. For instance, Stage 2 applies equally well to the mildly but bilaterally bradykinetic patient without balance difficulties and a severely bradykinetic subject who cannot work but still has retained postural reflexes. The scale likewise does not consider the possibility of unilateral signs with mild midline dysfunction without contralateral involvement or the presence of bilateral signs with suggestive, but not diagnostic, postural instability. An attempt to address this latter problem has introduced new ambiguities, because two versions are used currently, the original and an adaptation that allows two additional rating options, Stages 1.5 and 2.5 (Table 1).[3] These added options allow for midline involvement with unilateral signs (1.5) and very mild postural impairment (suggestive, but not diagnostic, usually one or two steps before recovery from a postural threat). This adaptation is variably used in reports, sometimes under the designation "modified HY." As indicated by the MDS survey, the modified version was in fact used by the responding MDS members more frequently than the original scale was. This adaptation, however, has never undergone clinimetric testing, and no clinimetric conclusions from the original five-point scale can be extrapolated directly to the modified version.

Because the HY scale relies on the determination of parkinsonism on one or both sides of the body and postural reflex testing, a standard procedural assessment protocol could allow all raters to test and judge responses similarly. Despite over 30 years of use, the HY scale does not have a teaching manual or teaching videotape with examples of each stage as conceived by the original writers.

Because the HY scale is weighted heavily toward postural instability assessments, progressive disability and impairments due to other features of PD, both motor and nonmotor, are not captured by the HY scale. Autonomic nervous system dysfunction and cognitive decline are known to influence overall clinical independence but are not captured specifically with this scale. Treatment-induced disability in patients with severe L-dopa-induced dyskinesias cannot be taken easily into account by the

HY scale, and if balance is impaired because of severe dyskinesia, it may be impossible to assign these patients to an accurate HY stages that reflects their parkinsonism. Finally, the scale is influenced heavily by balance and motility of the lower limbs, whereas disability due to upper limb dysfunction is not detected well.

Despite studies showing prolonged latencies to higher HY stages with L-dopa treatment,[21] the scale seems relatively insensitive to treatment-induced change, particularly in the lower categories.[9] The MDS-sponsored "Management of Parkinson's Disease: an Evidence-Based Review"[30] summarized treatment results from a large series of clinical trials with an emphasis on randomly ordered controlled trials (Level I). Several agents deemed efficacious in this report effected statistically significant changes in efficacy measures of parkinsonism, but failed to change HY scales.[30] As only one example, among six randomly ordered double-blind studies on the efficacy of an agonist for PD, despite statistically significant changes in the UPDRS, only one reported significant improvements in HY stage.[31] The others showed no change in the HY scores,[32,33] did not report the HY findings despite collecting the data[34,35] or did not report using the HY scale as an efficacy outcome.[36,37] Furthermore, even with drug treatment of PD that otherwise leads to clinically pertinent improvements, Stage 2 patients do not revert regularly to Stage 1. Some studies of modern dopaminergic therapies find the percentages of patients reaching the higher stages of the HY scale over 10 years to be similar to figures from the pre-L-dopa era.[38] Similarly, latencies to reach successive HY stages in a recent postmortem series of PD brains were comparable to those in untreated patients.[39]

Clinimetric Issues

Despite its widespread use and acceptance, few formal clinimetric examinations of reliability and validity for the HY scale have been conducted. Reliability testing assesses the scale's measurement error whereas validity testing assesses the scale's ability to measure its designated domains. The HY scale assesses PD disability and impairment at the ordinal level of measurement. One thus assumes that Stage 2 reflects greater disability/impairment than does Stage 1. Because the HY scale does not presume equal interval level measurements, however, one cannot assume that the difference between Stage 1 and Stage 2 is the same as that between Stage 2 and Stage 3. The assessment of clinimetric properties of ordinal level measures is therefore limited to nonparametric analytic techniques and assessments of internal consistency (a measure of reliability) and factor structure (a measure of validity) are not possible.

Most clinimetric examinations of the HY scale have been limited to the assessment of reliability. Scale reliability is usually tested by assessing: (1) its internal consistency, or the degree to which each item is consistent with the overall scale; and (2) the stability or consistency of ratings conducted by different raters (inter-rater reliability) or the same rater over time (test–retest or intra-rater reliability). All published reports of the HY scale reliability have assessed the latter source of measurement error, specifically inter-rater reliability. These studies document a moderate to significant level of inter-rater reliability, with nonweighted and weighted Kappa scores ranging between 0.44 and 0.71.[10,40,41] The stability of inter-rater reliability has been shown across experienced movement disorder specialists and inexperienced neurology residents.[41] No formal assessments of test–retest reliability (intra-rater reliability) have been conducted.

Clinimetric examinations of validity have been limited in scope. Validity assessments are typically carried out for face or content validity (does the scale seem to measure what it is designed to measure?), criterion validity (does the scale provide results comparable to a gold-standard?), and construct validity (does the scale adequately assess the domains of interest?). Some limitations to validity assessment of the HY scale are due to its ordinal level of measurement, but others are due to its construction and wording.

Because the scale combines disability and impairment criteria for each stage, establishing face validity of the HY scale is difficult. By combining these elements, certain clinical situations are difficult to assign (e.g., severe unilateral disease). Additionally, there is a lack of assessment of specific disease-related motor impairments (dyskinesia, tremor) and nonmotor impairments (cognitive impairment, depression).

No direct tests of the HY scale criterion validity have been conducted. Most studies have used the HY scale as the gold standard against which the validity of other scales is assessed. Although these studies cannot be considered examinations of HY scale validity, they do provide some assessment of the relationship between HY staging and other measures of PD impairment/disability. Most studies report significant correlations between HY stage and the other measures.[18,19,23,42–44] Related findings have been reported for significant correlations between HY stage and imaging measure of PD pathology, such as β-CIT SPECT scanning[29] and [^{18}F]fluorodopa PET scanning.[45] Taken together, these results suggest adequate criterion validity, or at least convergent validity. One problem with most of the above-mentioned studies, however, is a relative paucity of patient representation from

extreme HY stages. Most participants in these studies fall within Stage 2 to Stage 4, weakening the statistical confidence in relationships between HY stage and other measures for mild and advanced patients.

Although 0.5 intervals have been added in a "modified HY" scale, this adaptation has never been evaluated clinimetrically.

Because of clinimetric assumptions that underlie non-continuous scales, reports of group HY scores should be expressed with medians and interquartile ranges, not means and standard deviations (e.g., "The 60 subjects had a median HY stage of 3 [range 2–5]"). Analysis of differences between groups or change scores should involve the use of nonparametric methods. For parametric and survival analyses, the HY scale can be incorporated legitimately as an index used in the primary outcome. For example, because the development of postural instability is considered an important hallmark of clinical progression, the time to HY Stage 3 could be used as a primary variable for Kaplan-Meier curve presentation and subsequent analyses.

Comorbidities and the HY Staging System

PD is more prevalent in subjects over 50 years of age, and the coexistence of other diseases like diabetes, stroke, and arthritis can confound the evaluation of PD-related impairment and disability. The question of how the HY staging system should accommodate these various issues of comorbidities is not addressed specifically in the original scale description or later explanatory articles. Two different approaches could be applied: the first would consider the HY scale as dedicated to PD-related dysfunction in its purest sense and make a concerted attempt to disregard all components of impairment or disability due to conditions unrelated to PD; the second strategy would involve a "rate-as-you-see" approach, using the HY system to describe a PD patient's stage based on all the clinical impairments seen, regardless of their 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. A further argument against this method is that each stage conceptually represents a picture of overall function of the individual patient. The attempt to separate what is attributable to PD and what is attributable to comorbidities thus largely invalidates the concept of a global summary, which is the core of the HY scale's ease and utility. Given that standardized instructions for rating PD in the context of comorbidities do not exist for the HY system, all members of the Task Force except one agreed on the "rate-as-you-see" method that incorporates comorbidities into the staging choice.

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). Due to insufficiencies in their validation processes, very few neurological scales are associated with a well-defined MCRD.[46,47]

As already described, the HY scale is a categorical rather than a continuous scale, but it is constructed on the presumption that high scores represent more severe disease than do lower scores. MCRD values for categorical scales are potentially possible if certain criteria are met or certain assumptions are accepted. First, the progressive ratings within the scale must be sequentially relevant to the natural course of clinical decline in the disease under question. Each rating must be mutually exclusive of the others and in combination, the categories must cover the full spectrum of the disease. When MCRD values are established for a scale related to clinical status, patients move up and down the scale according to therapeutic responses. If a scale has excellent clinimetric properties for inter- and intra-rater reliability, the value of an intervention could be assessed at an individual level, defining each subject who changes by at least one unit as a responder, and at a group level by establishing the percentage of such responders.

In regards to the HY scale, the individual stages are relatively distinct and clearly relevant to PD, but they have been criticized as too broad and therefore insensitive to clinical change. The perceived need to use 0.5 increments and the wide usage of this adaptation in clinical practice (see above) suggests that a MCRD is likely smaller than one categorical increment in the original scale. Based on the original data presented in Hoehn and Yahr's report and other findings cited above, there is substantial evidence that a change from one stage to another in the HY scale is usually clinically relevant. The necessary assumptions needed for MCRD, however, are not met by the HY scale, and each stage increment is unlikely to represent the minimal difference implicit to the concept of a MCRD.

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 term refers to the difference between two groups at the end of the clinical

trial period. In the case of a placebo-controlled 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. In contrast to MCRD, MCRID applies only to groups, but is particularly relevant to the planning of clinical trials and to sample size calculations when establishing equivalence or non-inferiority to a standard treatment.[48,49] Again, although not specifically tested, the broad categorical ranges of HY and the lack of difference in HY after treatment with otherwise established treatments suggests that the implicit minimal difference of MCRID cannot be met by this scale.

Capturing the Clinical Spectrum of Parkinsonism: Application to Other Conditions

Many movement disorders specialists consider that the HY scale does not capture suitably the clinical spectrum of PD: 65% percent of respondents to the MDS questionnaire answered "No" to the question, "Do you think the Hoehn and Yahr scale represents well the patients you see?" The main arguments for this answer were that categories do not encompass all components of PD disability and that the categories are too broad and do not progress linearly.

Nonetheless, the HY scale has been used frequently to characterize broad descriptive categories of PD patients. In these instances, mild or early PD generally has been defined operationally as HY Stages 1 and 2, and advanced disease has likewise been defined as HY Stages 4 and 5. It is less clear that Stage 3 operationally defines moderate PD, because the hallmark of Stage 3, the occurrence of balance difficulties, has never been defined in a standardized manner and no systematic application of the objective pull test and teaching tape have been developed.

Although never developed for evaluating parkinsonism-plus syndromes, the HY scale has been utilized in several studies of progressive supranuclear palsy, multiple system atrophy, and dementia with Lewy bodies. Latencies to HY Stage 3 are significantly shorter in these patients compared with those in PD. In one cohort of subjects diagnosed with neuropathological criteria and clinical signs, no patient with PD developed HY Stage 3 within 1 year of onset of motor symptoms in contrast to more than two-thirds of the atypical parkinsonian disorder patients. Because these parkinsonism-plus syndromes generally have early balance deficits with postural instability, the HY scale is restricted often to only Stage 3, 4, and 5.[39] Whereas this limitation exists, it is not an implicit scale problem because the scale developers never suggested that their scale be adapted or applied to these conditions. In a practical sense, clinicians can use this observation for diagnostic purposes, because rapid progression to HY Stage 3 places the diagnosis of PD in serious doubt. It is a tribute to the strength of the scale that it has withstood so many applications and adaptations and remains used as a global summary of clinical status across several forms of parkinsonism.

CONCLUSIONS AND RECOMMENDATIONS

The HY scale is a widely used clinical rating scale, describing broad categories of motor dysfunction in PD. Among its advantages is that it is simple and applied easily. It captures typical patterns of PD progression with and without dopaminergic therapy. Progression in HY stage correlates with motor decline and deterioration in quality of life. The limited clinimetric analyses conducted to date support its scientific and clinical credibility. On the other hand, because of its simplicity, the scale is not comprehensive and by focusing on the issues of unilateral versus bilateral disease and the presence or absence of postural reflex impairment, it leaves other aspects of PD unassessed. By combining disability and impairment, ambiguities exist, and all clinical presentations of PD are not covered. The broad categories of the scale do not permit consistent detection of effective interventions, and the establishment of MCRD and MCRID indices is not feasible. Attempts to rectify weaknesses have included the introduction of widely used 0.5 increments to the scale, but this adaptation has not been tested clinimetrically and introduces unresolved analytic problems. Although still used frequently as an outcome measure in clinical trials, the HY scale has been replaced largely by the UDPRS as a primary outcome measure of treatment efficacy. Time to the development of a given HY stage has been used successfully to distinguish patients with PD from other parkinsonism-plus syndromes, and this measure could be potentially incorporated into interventional studies designed to test delay in clinical progression. Based on these considerations, the Task Force on Rating Scales in PD made the following recommendations:

1. The HY scale remains an important descriptive categorical scale in PD, and should continue to be used in its original form for demographical presentation of patient groups
2. When used to describe group characteristics, medians and ranges should be reported, not means with standard deviations, and comparisons between groups should involve nonparametric methods.
3. In research settings, the scale is useful primarily for

defining inclusion/exclusion criteria at baseline and as a validation standard for other rating instruments.

4. To retain the principle of simplicity that is at the core of the HY scale, when reporting HY stages, clinicians should "rate what you see" and therefore incorporate comorbidities.
5. Because of the wide usage of the "modified HY" scale with 0.5 increments, this adaptation warrants clinimetric testing. Without such testing, however, the original five-point scale should be maintained and the "modified HY" scale should not be favored.

Acknowledgments The institutional affiliations of the Task Force members are as follows: Dr. Goetz, Rush University Medical Center, Chicago, IL; Dr. Poewe, Innsbruck University Hospital, Innsbruck, Austria; Dr. Rascol, Laboratoire de Pharmacologie Medicale et Clinique, Toulouse, France; Dr. Counsell, Hospital de Santa Maria, Lisbon, Portugal; Dr. Stebbins, Rush University Medical Center, Chicago, IL; Dr. Sampaio, University of Aberdeen, Aberdeen, United Kingdom; Dr. Giladi, Tel Aviv Sourasky Medical Center, Tel-Aviv, Israel; Dr. Holloway, University of Rochester Medical Center, Rochester, NY; Dr. Moore, University of South Carolina, Columbia, SC; Dr. Wenning, Innsbruck University Hospital, Innsbruck, Austria; Dr. Yahr, Mount Sinai Medical Center, New York, NY; Ms. Seidl, The *Movement* Disorder Society, Milwaukee, WI.

REFERENCES

1. Movement Disorder Society Task Force on Rating Scales for PD. Unified Parkinson's disease rating scale: status and recommendations. Mov Disord 2003:18:738–750.
2. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. Neurology 1967;17:427–442.
3. Jankovic J, McDermott M, Carter J, Gauthier S, Goetz C, Golbe L, et al. Variable expression of Parkinson's disease: a base-line analysis of the DATATOP cohort. Neurology 1990;50:1529–1534.
4. Fahn S, Elton RL, UPDRS program members. Unified Parkinson's disease rating scale. In Fahn S, Marsden CD, Goldstein M, Calne DB, editors. Recent developments in Parkinson's disease, Vol 2. Florham Park, NJ: Macmillan Healthcare Information; 1987. p 153–163
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. Webster DD. Critical analysis of the disability in Parkinson's disease. Mod Treat 1968;5:257–258.
7. Guy W. ECDEU assessment manual for psychopharmacology. Rockville, MD: US Department of Health, Education and Welfare; 1976.
8. Mitchell SL, Harper DW, Lau A, Bhalle R. Patterns of outcome measurement in Parkinson's disease clinical trials. Neuroepidemiology 2000;19:100–108.
9. Diamond SG, Markham CH. Evaluating the evaluations: or how to weigh the scales of parkinsonian disability. Neurology 1983;33: 1098–1099.
10. Geminiani G, Cesana BM, Tamma F, Contri P, Pacchetti C, Carella F, et al. Interobserver reliability between neurologists in training of Parkinson's disease rating scales. Mov Disord 1991;6: 330–335.
11. Ramaker C, Marinus J, Stiggelbout AM, van Hilten BJ. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. Mov Disord 2002;17:867–876.
12. Schrag A, Selai C, Jahanshahi M, Quinn NP. The EQ-5D—a generic quality of live measure—is a useful instrument to measure quality of life in patients with Parkinson's disease. J Neurol Neurosurg Psychiatry 2000;69:67–73.
13. Jenkinson C, Fitzpatrick R, Peto V, Greenhall R, Hyman N. The Parkinson's Disease Questionnaire (PDQ-39): development and validation of a Parkinson's disease summary index score. Age Ageing 1997;28:353–357.
14. Welsh M, McDermott MP, Holloway RG, Plumb S, Pfeiffer R, Hubble J, Parkinson Study Group. Development and testing of the Parkinson's disease quality of life scale. Mov Disord 2003;18:637–645.
15. Pinter MM, Pogarell O, Oertel WH. Efficacy, safety and tolerance of the non-ergoline dopamine agonist, pramipexole in the treatment of advanced PD. J Neurol Neurosurg Psychiatry 1999;66: 436–441.
16. Reynolds NC, Montgomery GK. Factor analysis of Parkinson's impairment. Arch Neurol 1987;44:1013–1016.
17. Henderson L, Kennard C, Crawford TJ, Day S, Everitt BS, Goodrich S, et al. Scales for rating motor impairment in Parkinson's disease: studies of reliability and convergent validity. J Neurol Neurosurg Psychiatry 1991;54:18–24.
18. Hely M, Wilson A, Williamson P, O'Sullivan D, Rail D, Morris G. Reliability of the Columbia scale for assessing signs of Parkinson's disease. Mov Disord 1993;8:466–472.
19. Martinez-Martin P, Gil-Nagel A, Gracia LM, Gomez JB, Martinez-Sarries J, Bermejo F. Unified Parkinson's disease rating scale characteristics and structure. The Cooperative Multicentric Group. Mov Disord 1994;9:76–83.
20. Vingerhoets FJ, Snow BJ, Lee CS, Schulzer M, Mak E, Calne DB. Longitudinal fluorodopa positron emission tomographic studies of the evolution of idiopathic parkinsonism. Ann Neurol 1994;36: 759–764.
21. Marttila RJ, Rinne UK. Disability and progression in Parkinson's disease. Acta Neurol Scand 1977;56:159–169.
22. Maier-Hoehn MM. Parkinsonism treated with levodopa: progression and mortality. J Neural Transm 1983;19:253–264.
23. 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.
24. Poewe WH, Wenning GK. The natural history of Parkinson's disease. Neurology 1996;47(Suppl.):146–152.
25. Goetz CG, Tanner CM, Shannon KM. Progression of Parkinson's disease without levodopa. Neurology 1987;37:695–698.
26. Goetz CG, Stebbins GT, Blasucci LM. Differential progression of motor impairment in levodopa-treated Parkinson's disease. Mov Disord 2000;15:479–484.
27. Roos RAC, Jongen JCF, van der Velde EA. Clinical course of patients with idiopathic Parkinson's disease. Mov Disord 1996;11: 236–242.
28. Growdon JH, Corkin S, Desclos G, Rosen TJ. Hoehn and Yahr stage predicts extent of cognitive deficits in Parkinson's disease. Neurology 1987;37:157.
29. Staffen W, Mair A, Unterrainer J, Trinka E, Ladurner G. Measuring the progression of idiopathic Parkinson's disease with 123I-β-CIT-SPECT. J Neural Transm 2000;107:543–552.
30. Goetz CG, Koller WC, Poewe W, Rascol O, Sampaio C, Task Force for Evidence-Based Review of the Movement Disorder Society. Management of Parkinson's disease: an evidence-based review. Mov Disord 2002;12(Suppl.):1–166.
31. Lieberman A, Ranhosky A, Korts D. Clinical evaluation of pramipexole in advanced PD. Neurology 1997;49:162–168.
32. Guttman M, International Pramipexole-Bromocritine Study Group. Double-blind comparison of pramipexole and bromocriptine treatment with placebo in advanced PD. Neurology 1997;49:1060–1065.

33. Pinter MM, Helscher RJ, Nasel COJ, Riedl E, Schnaberth G. Quantification of motor deficit in Parkinson's disease with a motor performance test series. J Neural Transm 1992;4:131–141.
34. Parkinson Study Group. Safety and efficacy of pramipexole in early PD. JAMA 1997;278:125–130.
35. Shannon KM, Bennett JP, Friedman JH, Pramipexole Study Group. Efficacy of pramipexole, a novel dopamine agonist, as monotherapy in mild to moderate PD. Neurology 1997;49:724–728.
36. Hubble JP, Koller WC, Cutler NR, Sramek JJ. Pramipexole in patients with early PD. Clin Neuropharmacol 1995;18:338–347.
37. Parkinson Study Group. Pramipexole vs. levodopa as initial treatment for PD. JAMA 2000;284:1931–1938.
38. Hely MA, Morris JGL, Traficante R, Reid WG. The Sydney multicentre study of Parkinson's disease: progression and mortality at 10 years. J Neurol Neurosurg Psychiatry 1999;67:300–307.
39. Müller J, Wenning GK, Jellinger K, McKee A, Poewe W, Litvan I. Progression of Hoehn and Yahr stages in parkinsonian disorders: a clinicopathologic study. Neurology 2000;55:888–891.
40. Ginanneschi A, Degl'Innocenti F, Maurello MT, Magnolfi S, Marini P, Amaducci L. Evaluation of Parkinson's disease: reliability of three rating scales. Neuroepidemiology 1988;7:38–41.
41. Ginanneschi A, Degl'Innocenti F, Maurello MT, Magnolfi S, Marini P, Amaducci L. Evaluation of Parkinson's disease: a new approach to disability. Neuroepidemiology 1991;10:282–287.
42. Rabey JM, Bass H, Bonuccelli U, Brooks D, Klotz P, Korczyn AD, Kraus P, Martinez-Martin P, Morrish P, Van Sauten W, Van Hilten B. 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;20:322–337.
43. Stebbins GT, Goetz CG. Factor structure of the Unified Parkinson's Disease Rating Scale: motor examination section. Mov Disord 1998;13:633–636.
44. Stebbins GT, Goetz CG, Lang AE, Cubo E. Factor analysis of the motor section of the unified Parkinson's disease ratings scale during the off-state. Mov Disord 1999;14:585–589.
45. Eidelberg D, Moeller JR, Ishikawa T, Spetsieris P, Chaly T, Robeson W, Dahl JR, Margouleff D. Assessment of disease severity in parkinsonism with fluorine-18-fluorodeoxyglucose and PET. J Nucl Med 1995;36:378–383.
46. Bird SB, Dickson EW. Clinically significant changes in pain along the visual analog scale. Ann Emerg Med 2001;38:639–643.
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. Macedo A, Nunes S, Ferreira JJ, Castro-Caldas A, Durão V, Sampaio C. A survey of the evidence on the treatment of Parkinson's disease. In: Korczyn A, editor. Fourth congress of the European society of clinical neuropharmacology. Rome: Monduzzi Editore; 1997. p 85–91.
49. Wheatley K, Stowe RL, Clarke CE, Hills RK, Williams AC, Gray R. Evaluating drug treatments for Parkinson's disease: how good are the trials? BMJ 2002;324:1508–1511.