

Task Force Report: Scales for Screening and Evaluating Tremor Critique and Recommendations

Rodger Elble, MD, PhD,^{1*} Peter Bain, MD,² Maria João Forjaz, PhD,³ Dietrich Haubenberger, MD,⁴ Claudia Testa, MD, PhD,⁵ Christopher G. Goetz, MD,⁶ Albert F. G. Leentjens, MD, PhD,⁷ Pablo Martinez-Martin, MD, PhD,⁸ Anne Pavy-Le Traon, MD, PhD,⁹ Bart Post, MD, PhD,¹⁰ Cristina Sampaio, MD, PhD,^{11,12} Glenn T. Stebbins, PhD,⁶ Daniel Weintraub, MD,¹³ and Anette Schrag, PhD, FRCP¹⁴

¹Department of Neurology, Southern Illinois University School of Medicine, Springfield, Illinois, USA

²Department of Neurology, Imperial College London, Charing Cross Hospital, London, United Kingdom

³National School of Public Health and Red de Investigación en Servicios Sanitarios y Enfermedades Crónicas (REDISSEC), Carlos III Institute of Health, Madrid, Spain

⁴Department of Neurology, Medical University of Vienna, Vienna, Austria

⁵Department of Neurology, Virginia Commonwealth University, Richmond, Virginia, USA

⁶Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois, USA

⁷Department of Psychiatry, Maastricht University Medical Center, Maastricht, the Netherlands

⁸Alzheimer Center Reina Sofia Foundation and Centro de Investigación Biomédica en Red Sobre Enfermedades Neurodegenerativas (CIBERNED), Carlos III Institute of Health, Madrid, Spain

⁹Department of Neurology, Hôpital Purpan, University Hospital of Toulouse, Toulouse, France

¹⁰Department of Neurology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands

¹¹CHDI Foundation, Princeton, New Jersey, USA

¹²Instituto de Medicina Molecular, University of Lisbon, Portugal

¹³Department of Psychiatry, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA

¹⁴University College London Institute of Neurology, London, United Kingdom

ABSTRACT: The Movement Disorder Society established a task force to review rating scales for the assessment of tremor. Screening instruments used in identifying patients with tremor were also reviewed. Seven tremor severity scales, six activities of daily living (ADL)/disability scales, four quality-of-life scales, and five screening instruments were identified by searching PubMed.gov. The availability, use, acceptability, reliability, validity, and sensitivity to change were reviewed for each scale; and each scale was classified as recommended, suggested or listed based on whether 3, 2, or 1 of the following criteria were met: (1) used in the assessment of tremor (yes/no), (2) used in published studies by people other than the developers (yes/no), and (3) successful clinimetric testing (yes/no). Five tremor severity scales (the Fahn-Tolosa-Marin Tremor Rating Scale, the Bain and Findley Clinical Tremor Rating Scale, the Bain and Findley Spirography Scale, the Washington Heights-Inwood Genetic Study of Essential

Tremor Rating Scale, and the Tremor Research Group Essential Tremor Rating Assessment Scale), one ADL/disability scale (the Bain and Findley Tremor ADL Scale), one quality-of-life scale (the Quality of Life in Essential Tremor Questionnaire), and one screening instrument (the Washington Heights-Inwood Genetic Study of Essential Tremor Rating Scale, version 1) are recommended using these criteria. However, all scales need a more comprehensive analysis of sensitivity to change in order to judge their utility in clinical trials and individual patient assessments. The task force recommends that further work with existing recommended scales be performed as opposed to the development of new tremor scales. © 2013 International Parkinson and Movement Disorder Society

Key Words: tremor; rating scales; reliability; validity; sensitivity

*Correspondence to: Dr. Rodger Elble, Southern Illinois University School of Medicine, PO Box 19643, Springfield, IL 62794-9643; relble@siu.edu

Relevant conflicts of interest/financial disclosures: Nothing to report.

Full financial disclosures and author roles may be found in the online version of this article.

Received: 22 April 2013; **Revised:** 25 June 2013; **Accepted:** 17 July 2013

Published online 3 September 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.25648

Tremor is the most common movement disorder in adults, and essential tremor (ET) is the most common cause of tremor; however, before 1993, there were no tremor rating scales with documented reliability and validity.² Since then, many investigators have addressed the need for validated tremor scales. The Movement Disorder Society (MDS) Committee on Rating Scales in Movement Disorders asked one of us (R.J.E.) to form and lead a task force for this review of tremor scales.

Materials and Methods

Selection of Scales

Rating scales were considered if they were designed specifically for the assessment of tremor amplitude or the impact of tremor on activities of daily living (ADL) and quality of life. We did not review scales that included tremor in a broader assessment of Parkinson disease, dystonia, or ataxia. Screening instruments used in identifying patients with tremor were also reviewed, and we reviewed scales that were not designed specifically for tremor if they had been used repeatedly in the assessment of tremor by more than one group of investigators.

Medline on PubMed.gov was searched in February 2012 using Boolean expressions consisting of tremor AND each of the following: measurement, treatment, scale, rating, quality of life, quantification, disability, handicap, sensitivity, reliability, and validity. The titles and abstracts of references retrieved by this search were screened for relevance, and potentially relevant articles and references cited in those articles were reviewed. Only scales described completely in the literature were considered. We did not include volumetric³ or timed⁴ measures of tremor, which are not true rating scales, and we did not include scales requiring laryngoscopy⁵ or other sophisticated instrumentation.

Method of Review

An evaluation form was developed by the Task Force, according to published MDS guidelines,⁶ to facilitate the consistent, structured, comprehensive review of all scales (see online Appendix 1). This form addressed the content, availability, use, acceptability, reliability, validity, and sensitivity to change; and it concluded with an overall impression. All members used this form to evaluate the Fahn-Tolosa-Marin Tremor Rating Scale⁷ to test the appropriateness of the form. Subsequently, the chair and one or two other members of the Task Force completed the form for each scale. These evaluations were summarized (see online Appendix 2) and circulated to all Task Force members for final review.

The final recommendations were based on three criteria, which were used in previous MDS reviews⁶ of

this type: (1) used in the assessment of tremor (yes/no), (2) used in published studies by individuals other than the developers (yes/no), and (3) successful clinimetric testing (yes/no). Clinimetric testing was considered successful if it has been demonstrated that a scale was reliable, valid, and sensitive to change. Scales were classified as *recommended*, *suggested*, or *listed*, respectively, based on whether 3, 2, or 1 criterion were met.

Successful clinimetric testing was defined as follows. For reliability, we required high test-retest reliability for self-administered scales and questionnaires, and we required high test-retest reliability or inter-rater reliability for observer-rated scales. Internal consistency testing was considered desirable for all scales. For validity, good face or content validity and high correlations with other measures of tremor (convergent validity) were required. For sensitivity to change (responsiveness), the detection of change in at least one clinical trial was required, and good test-retest reliability was desirable.^{8,9}

Results

Seventeen scales (Table 1) and five screening instruments were selected for review. Most were designed for ET or cerebellar intention tremor. Recommended scales are reviewed here, and reviews of suggested and listed scales are available online in Appendix 3.

Tremor Severity Scales

The Fahn-Tolosa-Marin Tremor Rating Scale

The Fahn-Tolosa-Marin scale was first published in 1988¹⁰ and was later revised in 1993.⁷ This scale contains sections for assessing rest, postural and kinetic/intention tremor amplitude in specific anatomic locations (part A); tremor in writing, drawing, and pouring (part B); activities of daily living (part C); and global assessments by the patient and examiner, with each item rated on a scale from 0 to 4. The main differences in the initial and revised versions of this scale are (1) the assessment of orthostatic tremor (tremor in trunk and lower limbs when standing), (2) an item assessing the impact of tremor on social activities, and (3) the definition of severe extremity tremor as being >4 cm rather than >2 cm.

Clinimetric properties. For the first version of the scale, the scores of two raters for 10 patients correlated highly for part A ($\rho = 0.93$), part B ($\rho = 0.94$), part C ($\rho = 0.99$), the total score ($\rho = 0.97$), and global assessment by the examiner ($\rho = 0.99$).⁷ Two experienced raters using the revised scale in a study of 15 patients with ET had a test-retest reliability (intra-class correlation [ICC]) of 0.859 for the total score and ICCs of 0.882, 0.825, and 0.671 for parts A, B, and C, respectively.¹¹ In patients with ET and multiple

TABLE 1. Overview of the scales assessed and their classification

Scale	Applied in Tremor	Applied Beyond Original Developers	Clinimetric Testing			Scale Designation ^a
			Reliability	Validity	Sensitivity to Change	
Tremor severity scales						
Fahn-Tolosa-Marin Tremor Rating Scale	X	X	X	X	X	Recommended
Bain and Findley Clinical Tremor Rating Scale	X	X	X	X	X	Recommended
Bain and Findley Spirography Scale	X	X	X	X	X	Recommended
WHIGET Tremor Rating Scale, version 2	X	X	X	X	X	Recommended
TETRAS	X	X	X	X	X	Recommended
Bain and Findley handwriting scale ^{2,3}	X		X	X	X	Suggested
Matsumoto Clinical Tremor Rating Scale ⁴⁰	X		X	X	X	Suggested
Disability and ADL scales						
Bain and Findley Tremor ADL Scale	X	X	X	X	X	Recommended
TADLS ⁵²	X	X		X	X	Suggested
CADET ³¹	X		X	X		Listed
WHIGET Performance-Based Test of Function in Essential Tremor ³²	X			X		Listed
Bain and Findley Handicap Scale ³	X					Listed
Glass scale ⁵³	X		X	X		Listed
ETEA ⁵⁴	X		X	X		Listed
Quality-of-life scales						
QUEST	X	X	X	X	X	Recommended
MOS Short-Form Health Survey (SF-36) ⁵⁵	X	X		X	X	Suggested
QLS ⁵⁶	X					Listed

^aFor reviews of suggested and listed scales, see online Appendix 3.

Abbreviations: WHIGET, Washington Heights-Inwood Genetic Study of Essential Tremor; TETRAS, Tremor Research Group Essential Tremor Rating Assessment Scale; ADL, activities of daily living; TADLS, Tremor Activities of Daily Living Scale; CADET, Columbia University Assessment of Disability in Essential Tremor; ETEA, Essential Tremor Embarrassment Assessment; QUEST, Quality of Life in Essential Tremor Questionnaire; MOS, Medical Outcomes Study; QLS, Questions on Life Satisfaction.

sclerosis, intra-rater reliabilities (the same videotaped examinations presented twice to the same raters) for upper extremity items in parts A and B were good to excellent, but inter-rater reliability can be fair to poor for some items when examiners are not trained.^{12,13} The scale has good face validity and correlates strongly with transducer measures of tremor.^{14,15} It has been sensitive to change in treatment trials^{11,16,17} and in a longitudinal study of ET.¹⁸

Strengths and weaknesses. This scale is well known and has been used widely in clinical trials. It includes both clinician-based and patient-based ratings as well as disability aspects. The anchors for face, tongue, voice, head, trunk, and orthostatic tremor are ambiguously defined as none (0), slight (1), moderate (2), marked (3), and severe (4). Raters must be trained to achieve adequate reliability. The definition of severe extremity tremor (>4 cm) could have a problematic ceiling effect in a study of patients with advanced tremor, as in trials of deep brain stimulation, and the instructions on where the rater is to measure tremor (ie, index finger, great toe) could be stated more clearly.

Conclusion. This scale fulfills criteria for a recommended scale in the assessment of tremor severity, because it has been used in multiple studies of tremor and has demonstrated good overall clinimetric properties. However, good inter-rater reliability requires

training, and there is a potential ceiling effect for upper extremity tremor.

Bain and Findley Clinical Tremor Rating Scale

In 1993, Bain and Findley introduced a battery of tremor rating scales for assessing rest and postural and kinetic/intention tremor severity in the head, voice, and four limbs using ratings on a scale from 0 to 10.^{2,3} Most individuals cannot discriminate beyond 7 levels,¹⁹ so Bain and coworkers redefined 0 as no tremor, 1 to 3 as mild tremor, 4 to 6 as moderate tremor, 7 to 9 as severe tremor, and 10 as extremely severe tremor, resulting in a 5-level scale with intermediate gradations.^{2,3,20}

Clinimetric properties. Inter-rater reliability (κ values) ranged from 0.55 to 0.81 for upper limb postural tremor and from 0.58 to 0.84 for head tremor in patients with ET and tremor associated with cervical dystonia.² Intra-rater reliability for these two assessments ranged from 0.62 to 0.85 and from 0.71 to 0.78, respectively. Reliability for kinetic (finger-nose) tremor was less, and the reliability for voice tremor was poor.² Similar reliability was observed for upper extremity tremor in individuals with multiple sclerosis.²¹ For validity testing, the right upper extremity postural tremor score correlated with accelerometry

($r = 0.655$), an ADL scale ($r = 0.628$), spiral ratings ($r = 0.804$), and handwriting ratings ($r = 0.762$) in patients with ET²; and upper extremity postural tremor in patients with multiple sclerosis correlated with a finger-tapping task ($r = -0.61$), a pegboard task ($r = -0.62$), and an ADL questionnaire ($r = 0.58$).²¹ In patients with Parkinson's disease, the ratings from 0 to 10 for rest tremor correlated well ($r = 0.67$) with rest tremor item 20 from the Unified Parkinson Disease Rating Scale.²² This scale was sensitive to change in a trial of dopaminergic agonists for Parkinson's disease rest tremor^{22,23} and in a trial of stereotactic surgery for tremor in multiple sclerosis.²⁴

Strengths and weaknesses. This scale has been used widely and can be done quickly at the bedside. No cups, paper, pen, or other props are needed. However, the scoring mild/moderate/severe is based on subjective impressions. Only upper limb postural tremor and head tremor have acceptable reliability, and there are inconsistencies among raters on the definition and interpretation of kinetic versus intention tremor.

Conclusion. This scale fulfills criteria for a recommended scale in the assessment of tremor severity in the head and upper limbs, because it has been used in several settings with adequate psychometric properties demonstrated, but it can be difficult for raters who are naive to tremor assessments and could be improved by a video library that provides examples of different tremor severities.

Bain and Findley spirometry scale

Bain and Findley used ratings from 0 to 10 for the assessment of Archimedes spirals,² and illustrated examples have been published.³

Clinimetric properties. Inter-rater reliability (κ values) ranged from 0.56 to 0.9, and intra-rater reliability ranged from 0.58 to 0.91 in patients with ET and tremor associated with cervical dystonia.^{2,25} The inter-rater ICC for seven raters of 54 spirals was 0.93 for patients with ET.¹⁵ Reliability was not as good in a cohort of patients with multiple sclerosis.²¹ The ratings from 1 to 4 of the spiral scale are less reliable, because it is difficult to distinguish mild tremor from other irregularities in movement. According to Bain and Findley,³ 95% of normal controls had a score of 1 or lower; however, in a population-based screening study, mean spiral scores of 2.3 and 3.7 were reported in individuals with no tremor.²⁶ This scale has a potential ceiling effect in severely affected patients, because a score of 10 is given when tremor "completely disrupts an attempt to draw a spiral."³

The spiral scale correlated well ($r > 0.8$) with ET that was quantified with a digitizing tablet and the Fahn-Tolosa-Marin scale,¹⁵ and it also correlated with accelerometry ($r = 0.406$), an ADL questionnaire ($r = 0.659$), RUE postural tremor rating ($r = 0.804$),

and a handwriting scale ($r = 0.917$).² This scale also correlated well with an ADL questionnaire ($r = 0.77$), a finger-tapping task ($r = -0.68$), and a pegboard task ($r = -0.74$) in patients with multiple sclerosis.²¹ The spiral scale was sensitive to change in a study of stereotactic surgery for multiple sclerosis²⁴ and in studies of ethanol for ET^{15,27} and botulinum toxin for tremor in multiple sclerosis.²⁵

Strengths and weaknesses. Reliability is good if raters are trained using the examples from the original publication,³ and this measure of action tremor has good face and construct validity. However, the ratings are subjective, and there are some floor and ceiling effects.

Conclusion. This scale fulfills criteria for a recommended scale in the assessment of tremor severity, but its sensitivity at the lower and higher end is probably limited.

Washington Heights-Inwood Genetic Study of Essential Tremor Tremor Rating Scale

The original Washington Heights-Inwood Genetic Study of Essential Tremor (WHIGET) Tremor Rating Scale was developed for the purpose of distinguishing abnormal upper extremity tremor from normal tremor in population studies of ET.^{28,29} This scale was later revised so that kinetic tremor is rated from 0 to 4 instead of from 0 to 3, although postural tremor retains a rating from 0 to 3.³⁰ This was done "to broaden the applicability of this scale to clinical trials."³⁰

Clinimetric properties. Inter-rater and intra-rater reliability (weighted κ) of at least 0.97 was achieved for version 2 using a teaching videotape.³⁰ The total score on the original 0 to 3-point scale correlated ($r = 0.57$) with a tremor disability questionnaire³¹ and with the WHIGET Performance-Based Test of Function in Essential Tremor ($r = 0.71$).³² Version 1 of this scale was sensitive to progression in a longitudinal study of ET.³³ Version 2 was sensitive to change in a study of sodium oxybate for ethanol-sensitive ET.³⁴

Strengths and weaknesses. This scale is limited to upper extremity tremor and was designed specifically for the assessment of ET. A training video is required to achieve high reliability.³⁰ The instructions and implements (ie, cups and spoon) in this scale require clarification and standardization.

Conclusion. This scale fulfills criteria for a recommended scale in the assessment of tremor severity. Version 2 is more appropriate for the assessment of tremor severity. However, validity testing was published only for version 1, and data supporting the validity of version 2 should be published.

Tremor Research Group Essential Tremor Rating Assessment Scale

The Tremor Research Group Essential Tremor Rating Assessment Scale (TETRAS) has a 12-item ADL

subscale with ratings from 0 to 4, and it contains a 9-item performance subscale that rates action tremor in the head, face, voice, limbs, and trunk from 0 to 4 in half-point intervals. Head and limb tremor ratings are defined by specific amplitude ranges in centimeters to reduce experiential rating bias and uncertainty.

Clinimetric properties. Reliability testing has been done only for the performance subscale.³⁵ The inter-rater and intra-rater ICCS for head and upper limb tremor ranged from 0.86 to 0.96, respectively; and the inter-rater and intra-rater ICCs for the total score were 0.94 and 0.96, respectively, even for untrained raters.³⁵ The ICCs for voice, face, trunk, and leg were less robust. TETRAS performance scores correlated well with transducer measures of tremor³⁶ and with the TETRAS ADL subscale.³⁵ Good test-retest reliability and sensitivity to change have been demonstrated for upper limb postural and kinetic ET.^{37,38}

Strengths and weaknesses. TETRAS is a short, easy to apply scale that was designed specifically for the clinical assessment of ET. The performance subscale is a highly reliable and valid assessment of action tremor in the head and upper limbs.

Conclusion. This scale fulfills criteria for a recommended scale in the assessment of tremor severity. Assessment of test-retest reliability and sensitivity to change is needed for the ADL subscale.

Disability and Activities of Daily Living Scales **Bain and Findley Tremor Activities of Daily Living Scale**

The original 25-item scale of Bain and Findley² was reduced to a 20-item self-assessment ADL scale by Lundervold and coworkers,³⁹ and these are referred to as the Bain Disability Scale⁴⁰ and the Tremor Disability Scale (TREDS), respectively.³⁹ Both mainly address upper limb function.

Clinimetric properties. Reliability testing has been performed on the 20-item version only. Mastumoto and coworkers reported a test-retest Spearman correlation of 0.99 in a cohort of multiple sclerosis patients.⁴⁰ Bain and Findley observed good correlations with upper extremity tremor scores ($r = 0.628$), spiral scores ($r = 0.659$), and handwriting scores ($r = 0.686$) in patients with ET and cervical dystonia.² In patients with multiple sclerosis, the ADL score correlated with upper extremity postural tremor scores ($r = 0.59$), spiral scores ($r = 0.77$), and handwriting scores ($r = 0.76$).²¹ However, ADL scores did not correlate significantly with the Matsumoto Clinical Tremor Rating Scale but did correlate with quantitative motion analysis ($\rho = 0.75$) in patients with multiple sclerosis.⁴⁰ For ET patients, the 20-item version of this scale correlated with the WHIGET Performance-Based Test of Function in Essential Tremor ($r = 0.74$).³⁹ The scale was moderately sensitive to the

effects of thalamotomy and thalamic deep brain stimulation in patients with multiple sclerosis.^{24,40}

Strengths and weaknesses. This ADL scale focuses primarily on upper limb function. It has acceptable reliability and validity. Sensitivity to change appears to be modest, at least for tremor in multiple sclerosis.

Conclusion. This scale fulfills criteria for a recommended scale in the assessment of ADL.

Quality-of-Life Scales

Quality of Life in Essential Tremor questionnaire

This quality-of-life questionnaire was designed specifically for ET.⁴¹

Clinimetric properties. Although it was intended to be “relatively brief,” in two studies, only about 40% of individuals completed the questionnaire in its entirety.^{41,42} Test-retest reliability is adequate for the Quality of Life in Essential Tremor (QUEST) questionnaire summary index (ICC = 0.77) but is lower for the dimensions work/finances and hobbies/leisure (ICC = 0.60 and ICC = 0.57, respectively). There were modest correlations with self-ratings of tremor severity.⁴¹ In one deep brain stimulation study, a 30% improvement in the QUEST was reported, whereas there was no significant change in scores on the 36-item Medical Outcomes Study short-form quality-of-life measure (SF-36).⁴³ QUEST was not sensitive to a beneficial effect of botulinum toxin on upper limb tremor in multiple sclerosis.²⁵

Strengths and weaknesses. QUEST is the first quality-of-life questionnaire designed specifically for ET. Compared with the SF-36, QUEST may be more sensitive to change in clinical treatment trials, but acceptability has been poor.

Conclusion. This scale fulfills criteria for a recommended scale in the assessment of quality of life. However, the issue of missing data should be addressed.

Screening Instruments

The screening instruments reviewed are summarized in Table 2. Reviews of the instruments that were classified as *suggested* or *listed* are contained in online Appendix 3.

WHIGET Tremor Rating Scale, version 1

The original WHIGET Tremor Rating Scale was developed for the purpose of identifying patients with ET in population studies, and this version has inclusion and exclusion criteria for distinguishing ET from other forms of action tremor.^{28,29} Postural and kinetic tremor are rated from 0 to 3 in a battery of motor tasks. Diagnostic categories include possible, probable, and definite ET based on increasingly stringent criteria. The diagnosis of *definite ET* requires at least grade

TABLE 2. Overview of the screening instruments assessed and their classification

Screening Instrument	Applied in Tremor	Applied Beyond Original Developers	Clinimetric Testing				Scale Designation ^a
			Reliability	Validity	Sensitivity and Specificity		
WHIGET Tremor Rating Scale, version 1	X	X	X	X	X	Recommended	
WHIGET Tremor Screening Questionnaire ⁵⁷	X	X		X	X	Suggested	
Kiel tremor screening instrument ²⁶	X			X	X	Listed	
Louis spirography screening tool ⁵⁸	X					Listed	
Modified finger-nose-finger screening tool ⁵⁹	X			X		Listed	

^aFor reviews of suggested and listed instruments, see online Appendix 3. Abbreviations: WHIGET, Washington Heights-Inwood Genetic Study of Essential Tremor.

2 action tremor (a clearly oscillatory tremor of 1-cm to 2-cm amplitude) in multiple tasks.

Clinimetric properties. In a cohort of 40 ET patients and 60 controls, the inter-rater reliability (weighted κ) for two neurologists was 0.85.⁴⁴ The total score correlated ($r = 0.57$) with a tremor disability questionnaire³¹ and with the WHIGET Performance-Based Test of Function in Essential Tremor ($r = 0.71$).³² Disagreement among raters is common in the diagnosis of individuals with mild or questionably abnormal tremor, but the ability to distinguish definite ET from normal or questionably abnormal individuals is essentially perfect.^{28,44} This scale was sensitive to progression in a longitudinal study of ET.³³

Strengths and weaknesses. This scale was designed specifically for population studies of ET. The stringent diagnostic criteria for definite ET will exclude many patients with mild disease.^{28,44} The published training video for this scale is for version 2, not version 1.³⁰ The instructions and implements (ie, cups and spoon) in this scale require clarification and standardization.

Conclusion. Version 1 of this scale fulfills criteria for a recommended scale for screening individuals for at least moderate ET.

Discussion

The MDS classifies a scale as *recommended* if the scale has been applied to the targeted patient population(s), there are published data on its use by individuals other than the developer(s) of the scale, and the scale has been studied clinimetrically and has been found valid, reliable, and sensitive to change. We reviewed seven tremor severity scales, and five (Fahn-Tolosa-Marin, Bain and Findley Clinical Tremor Rating Scale, Bain and Findley Spirography Scale, version 2 of the WHIGET Tremor Rating Scale, and TETRAS) are recommended on the basis of these criteria (Table 1). Only one of six ADL/disability scales (Bain and Findley Tremor ADL Scale) is recommended. The Fahn-Tolosa-Marin Tremor Rating Scale and TETRAS

have ADL subscales that require additional clinimetric study. One of four quality-of-life scales (QUEST) is recommended. Excluding the second criterion (use by individuals other than the developers) would change the designation of only two scales, the Bain and Findley handwriting scale and the Matsumoto Clinical Tremor Rating Scale, and neither scale has been used extensively by the original developers.

Only one of five screening instruments, version 1 of the WHIGET Tremor Rating Scale, is recommended, with the caveat that the true sensitivity and specificity of this instrument are uncertain, because there is no *gold standard* for distinguishing normal tremor from abnormal tremor or for distinguishing ET from other forms of action tremor.⁴⁵ The electrophysiologic demonstration of abnormal motor unit entrainment is not completely sensitive or specific in identifying individuals with abnormal tremor; and tremor amplitudes, measured with motion transducers, exhibit significant overlap with control values due to natural variability in tremor amplitude from one moment to another.^{46–48}

Since 1993, considerable attention has been paid to intra-rater and inter-rater reliability and to content and construct validity of published tremor scales; however, comparing scales on the basis of intra-rater and inter-rater reliability is difficult, because different statistical methods (κ , weighted κ , ICC, Spearman correlation, etc.), patient populations, and examiners have been used. Many scales have required training to achieve high intra-rater and inter-rater reliability, but little attention has been devoted to reliability and generalizability of the training methods used by the developers.

Sensitivity to change (responsiveness) has received relatively little attention in the development of all tremor scales, but responsiveness is an essential issue in comparing one scale with another.^{9,49} We considered that the requirement of being *sensitive to change* was met if it had been demonstrated that a scale was sensitive to the effects of a therapeutic intervention or disease progression. However, no scale has been

subjected to a full evaluation of sensitivity to change. Test-retest reliability is lacking for many scales, and estimates of minimum detectable change and minimum clinically significant change are needed for all tremor scales.

Test-retest reliability is the consistency of a measure from one time to another and is influenced by variability in rater performance and patient performance (eg, a patient's tremor). Test-retest reliability and intra-rater reliability often are used interchangeably; however, strictly speaking, intra-rater reliability is the consistency of a rater in repeatedly measuring one instance of a variable (eg, a video of a patient's tremor).⁵⁰ Thus, test-retest reliability can be less than perfect even if intra-rater reliability is perfect. Hence, measurements of tremor with motion transducers will not have perfect test-retest reliability, even though rater performance is not an issue.³⁷ Better test-retest reliability results in greater statistical power (lower sample size requirement) and in smaller standard error of measurement and minimum detectable change.^{8,9} More estimates of test-retest reliability and minimum detectable change⁹ are needed for all tremor scales using patients with a full range of tremor severity. Statistical power and minimum detectable change can vary, depending on the range of scores at baseline,^{37,51} and the responsiveness of rating scales for a particular disorder can vary considerably, despite comparable reliability.⁴⁹

Finally, we did not identify a need for the MDS to sponsor the development of a new scale. However, many of the existing scales are in need of additional psychometric validation, with particular emphasis on test-retest reliability and sensitivity to change. Furthermore, most of the scales have not been validated in languages other than English or in pediatric populations, and the validity of existing scales for rare tremor disorders (eg, Holmes tremor, orthostatic tremor) is unknown. ■

References

- Louis ED, Ferreira JJ. How common is the most common adult movement disorder? Update on the worldwide prevalence of essential tremor. *Mov Disord* 2010;25:534–541.
- Bain PG, Findley LG, Atchison P, et al. Assessing tremor severity. *J Neurol Neurosurg Psychiatry* 1993;56:868–873.
- Bain PG, Findley LJ. *Assessing Tremor Severity: A Clinical Handbook*. London, United Kingdom: Smith-Gordon; 1993.
- Hellriegel H, Raethjen J, Deuschl G, Volkmann J. Levetiracetam in primary orthostatic tremor: a double-blind placebo-controlled crossover study. *Mov Disord* 2011;26:2431–2434.
- Bove M, Daamen N, Rosen C, Wang CC, Sulica L, Gartner-Schmidt J. Development and validation of the vocal tremor scoring system. *Laryngoscope* 2006;116:1662–1667.
- Antonini A, Martinez-Martin P, Chaudhuri RK, et al. Wearing-off scales in Parkinson's disease: critique and recommendations. *Mov Disord* 2011;26:2169–2175.
- Fahn S, Tolosa E, Marin C. Clinical rating scale for tremor. In: Jankovic J, Tolosa E, editors. *Parkinson's Disease and Movement Disorders*. 2nd ed. Baltimore, MD: Williams & Wilkins; 1993: 225–234.
- Perkins DO, Wyatt RJ, Bartko JJ. Penny-wise and pound-foolish: the impact of measurement error on sample size requirements in clinical trials. *Biol Psychiatry* 2000;47:762–766.
- Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005;19: 231–240.
- Fahn S, Tolosa E, Marin C. Clinical rating scale for tremor. In: Jankovic J, Tolosa E, editors. *Parkinson's Disease and Movement Disorders*. Baltimore, MD and Munich, Germany: Urban & Schwarzenberg; 1988:225–234.
- Elble RJ, Lyons KE, Pahwa R. Levetiracetam is not effective for essential tremor. *Clin Neuropharmacol* 2007;30:350–356.
- Hooper J, Taylor R, Pentland B, Whittle IR. Rater reliability of Fahn's tremor rating scale in patients with multiple sclerosis. *Arch Phys Med Rehabil* 1998;79:1076–1079.
- Stacy MA, Elble RJ, Ondo WG, Wu SC, Hulihan J. Assessment of interrater and intrarater reliability of the Fahn-Tolosa-Marin Tremor Rating Scale in essential tremor. *Mov Disord* 2007;22: 833–838.
- Elble RJ, Pullman SL, Matsumoto JY, Raethjen J, Deuschl G, Tintner R. Tremor amplitude is logarithmically related to 4- and 5-point tremor rating scales. *Brain* 2006;129(pt 10):2660–2666.
- Haubenberger D, Kalowitz D, Nahab FB, et al. Validation of digital spiral analysis as outcome parameter for clinical trials in essential tremor. *Mov Disord* 2011;26:2073–2080.
- Ondo WG, Jankovic J, Connor GS, et al. Topiramate in essential tremor: a double-blind, placebo-controlled trial. *Neurology* 2006; 66:672–677.
- Zhang K, Bhatia S, Oh MY, Cohen D, Angle C, Whiting D. Long-term results of thalamic deep brain stimulation for essential tremor. *J Neurosurg* 2010;112:1271–1276.
- Putzke JD, Whaley NR, Baba Y, Wszolek ZK, Uitti RJ. Essential tremor: predictors of disease progression in a clinical cohort. *J Neurol Neurosurg Psychiatry* 2006;77:1235–1237.
- Miller GA. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 1956;63:81–97.
- Bain PG, Mally J, Gresty M, Findley LJ. Assessing the impact of essential tremor on upper limb function. *J Neurol* 1993;241:54–61.
- Alusi SH, Worthington J, Glickman S, Findley LJ, Bain PG. Evaluation of three different ways of assessing tremor in multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2000;68:756–760.
- Navan P, Findley LJ, Jeffs JA, Pearce RK, Bain PG. Double-blind, single-dose, cross-over study of the effects of pramipexole, pergolide, and placebo on rest tremor and UPDRS part III in Parkinson's disease. *Mov Disord* 2003;18:176–180.
- Navan P, Findley LJ, Jeffs JA, Pearce RK, Bain PG. Randomized, double-blind, 3-month parallel study of the effects of pramipexole, pergolide, and placebo on Parkinsonian tremor. *Mov Disord* 2003; 18:1324–1331.
- Alusi SH, Aziz TZ, Glickman S, Jahanshahi M, Stein JF, Bain PG. Stereotactic lesional surgery for the treatment of tremor in multiple sclerosis: a prospective case-controlled study. *Brain* 2001;124(pt 8):1576–1589.
- Van Der Walt A, Sung S, Spelman T, et al. A double-blind, randomized, controlled study of botulinum toxin type A in MS-related tremor. *Neurology* 2012;79:92–99.
- Lorenz D, Papengut F, Frederiksen H, et al. Evaluation of a screening instrument for essential tremor. *Mov Disord* 2008;23:1006–1012.
- Knudsen K, Lorenz D, Deuschl G. A clinical test for the alcohol sensitivity of essential tremor. *Mov Disord* 2011;26:2291–2295.
- Louis ED, Ford B, Bismuth B. Reliability between two observers using a protocol for diagnosing essential tremor. *Mov Disord* 1998;13:287–293.
- Louis ED, Ottman R, Ford B, et al. The Washington Heights-Inwood Genetic Study of Essential Tremor: methodologic issues in essential-tremor research. *Neuroepidemiology* 1997;16:124–133.
- Louis ED, Barnes L, Wendt KJ, et al. A teaching videotape for the assessment of essential tremor. *Mov Disord* 2001;16:89–93.

31. Louis ED, Barnes LF, Wendt KJ, et al. Validity and test-retest reliability of a disability questionnaire for essential tremor. *Mov Disord* 2000;15:516–523.
32. Louis ED, Wendt KJ, Albert SM, Pullman SL, Yu Q, Andrews H. Validity of a performance-based test of function in essential tremor. *Arch Neurol* 1999;56:841–846.
33. Louis ED, Agnew A, Gillman A, Gerbin M, Viner AS. Estimating annual rate of decline: prospective, longitudinal data on arm tremor severity in two groups of essential tremor cases. *J Neurol Neurosurg Psychiatry* 2011;82:761–765.
34. Frucht SJ, Bordelon Y, Houghton WH, Reardan D. A pilot tolerability and efficacy trial of sodium oxybate in ethanol-responsive movement disorders. *Mov Disord* 2005;20:1330–1337.
35. Elble R, Comella C, Fahn S, et al. Reliability of a new scale for essential tremor. *Mov Disord* 2012;27:1567–1569.
36. Mostile G, Giuffrida JP, Adam OR, Davidson A, Jankovic J. Correlation between Kinesia system assessments and clinical tremor scores in patients with essential tremor. *Mov Disord* 2010;25:1938–1943.
37. Mostile G, Fekete R, Giuffrida JP, et al. Amplitude fluctuations in essential tremor. *Parkinsonism Relat Disord* 2012;18:859–863.
38. Chang WS, Chung JC, Kim JP, Chang JW. Simultaneous thalamic and posterior subthalamic electrode insertion with single deep brain stimulation electrode for essential tremor. *Neuromodulation* 2012;16:236–243.
39. Lundervold DA, Pahwa R, Ament PA, Corbin DE. Validity of clinical and patient ratings of tremor disability among older adults. *Parkinsonism Relat Disord* 2003;10:15–18.
40. Matsumoto J, Morrow D, Kaufman K, et al. Surgical therapy for tremor in multiple sclerosis: an evaluation of outcome measures. *Neurology* 2001;57:1876–1882.
41. Troster AI, Pahwa R, Fields JA, Tanner CM, Lyons KE. Quality of life in Essential Tremor Questionnaire (QUEST): development and initial validation. *Parkinsonism Relat Disord* 2005;11:367–373.
42. Martinez-Martin P, Jimenez-Jimenez FJ, Carroza Garcia E, et al. Most of the Quality of Life in Essential Tremor Questionnaire (QUEST) psychometric properties resulted in satisfactory values. *J Clin Epidemiol* 2010;63:767–773.
43. Sandvik U, Hariz GM, Blomstedt P. Quality of life following DBS in the caudal zona incerta in patients with essential tremor. *Acta Neurochir (Wien)* 2012;154:495–499.
44. Louis ED, Ford B, Pullman SL. Prevalence of asymptomatic tremor in relatives of patients with essential tremor. *Arch Neurol* 1997;54:197–200.
45. Louis ED, Ford B, Lee H, Andrews H, Cameron G. Diagnostic criteria for essential tremor: a population perspective. *Arch Neurol* 1998;55:823–828.
46. Elble RJ. Physiologic and essential tremor. *Neurology* 1986;36:225–231.
47. Elble RJ. Characteristics of physiologic tremor in young and elderly adults. *Clin Neurophysiol* 2003;114:624–635.
48. Elble RJ, Deuschl G. Tremor. In: Brown WF, Bolton CF, Aminoff M, editors. *Neuromuscular Function and Disease: Basic, Clinical and Electrodiagnostic Aspects*. Philadelphia, PA: W. B. Saunders Company; 2002:1759–1779.
49. Steffen T, Seney M. Test-retest reliability and minimal detectable change on balance and ambulation tests, the 36-item short-form health survey, and the unified Parkinson disease rating scale in people with parkinsonism. *Phys Ther* 2008;88:733–746.
50. Holmefur M, Aarts P, Hoare B, Krumlinde-Sundholm L. Test-retest and alternate forms reliability of the assisting hand assessment. *J Rehabil Med* 2009;41:886–891.
51. Donoghue D; Physiotherapy Research and Older People (PROP) Group; Stokes EK. How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. *J Rehabil Med* 2009;41:343–346.
52. Lyons KE, Pahwa R, Busenbark KL, Troster AI, Wilkinson S, Koller WC. Improvements in daily functioning after deep brain stimulation of the thalamus for intractable tremor. *Mov Disord* 1998;13:690–692.
53. Gironell A, Martinez-Corral M, Pagonabarraga J, Kulisevsky J. The Glass scale: a simple tool to determine severity in essential tremor. *Parkinsonism Relat Disord* 2010;16:412–414.
54. Traub RE, Gerbin M, Mullaney MM, Louis ED. Development of an essential tremor embarrassment assessment. *Parkinsonism Relat Disord* 2010;16:661–665.
55. Contopoulos-Ioannidis DG, Karvouni A, Kouri I, Ioannidis JP. Reporting and interpretation of SF-36 outcomes in randomised trials: systematic review [serial online]. *BMJ* 2009;338:a3006.
56. Kuehler A, Henrich G, Schroeder U, Conrad B, Herschbach P, Ceballos-Baumann A. A novel quality of life instrument for deep brain stimulation in movement disorders. *J Neurol Neurosurg Psychiatry* 2003;74:1023–1030.
57. Louis ED, Ford B, Lee H, Andrews H. Does a screening questionnaire for essential tremor agree with the physician's examination? *Neurology* 1998;50:1351–1357.
58. Louis ED, Zhao Q, Meng H, Ding D. Screening for Action Tremor in Epidemiological Field Surveys: Assessing the reliability of a Semi-Quantitative, Visual, Template-Based Scale for Rating Hand-Drawn Spirals. *Tremor Other Hyperkinet Mov (N Y)* 2012;2, pii.
59. Louis ED, Applegate LM, Borden S, Moskowitz C, Jin Z. Feasibility and validity of a modified finger-nose-finger test. *Mov Disord* 2005;20:636–639.