Reliability of a New Scale for Essential Tremor

Rodger Elble, MD, PhD,[1]* Cynthia Comella, MD,[2] Stanley Fahn, MD,[3] Mark Hallett, MD,[4] Joseph Jankovic, MD,[5] Jorge L. Juncos, MD,[6] Peter LeWitt, MD,[7] Kelly Lyons, PhD,[8] William Ondo, MD,[9] Rajesh Pahwa, MD,[8] Kapil Sethi, MD,[10] Natividad Stover, MD,[11] Daniel Tarsy, MD,[12] Claudia Testa, MD, PhD,[13] Ron Tintner, MD,[14] Ray Watts, MD,[11] and Theresa Zesiewicz, MD[15]

[1]Southern Illinois University School of Medicine, Springfield. Illinois, USA; [2]Rush University Medical Center, Chicago, Illinois, USA; [3]Columbia University Medical Center, New York, New York, USA; [4]National Institute of Neurological Disorders and Stroke, Bethesda, Maryland, USA; [5]Baylor College of Medicine, Houston, Texas, USA; [6]Emory University, Atlanta, Georgia, USA; [7]Henry Ford Health Systems, Bloomfield, Michigan, USA; [8]University of Kansas Medical Center, Kansas City, Kansas, USA; [9]University of Texas Health Science Center, Houston, Texas, USA; [10]Georgia Health Sciences University, Augusta, Georgia, USA; [11]University of Alabama at Birmingham, Birmingham, Alabama, USA; [12]Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA; [13]Virginia Commonwealth University, Richmond, Virginia, USA; [14]The Methodist Hospital, Houston, Texas, USA; [15]University of South Florida, Tampa, Florida, USA

ABSTRACT

Background: The objective of this study was to determine the reliability of a new scale for the clinical assessment of essential tremor. The Essential Tremor Rating Assessment Scale contains 9 performance items that rate 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.

Methods: Videos of 44 patients and 6 controls were rated by 10 specialists on 2 occasions 1–2 months apart. Inter- and intrarater reliability was assessed with a 2-way random-effects intraclass correlation, using an absolute agreement definition.

Results: Inter- and intrarater intraclass correlations for head and upper-limb tremor ranged from 0.86 to 0.96, and intraclass correlations for total score were 0.94 and 0.96. The intraclass correlations for voice, face, trunk, and leg were less robust.

Conclusions: This scale is an exceptionally reliable tool for the clinical assessment of essential tremor.

Key Words: essential tremor; rating scale; reliability

Before 1993, there were no rating scales for essential tremor (ET) with documented reliability and validity. Many different ad hoc scales had been used in numerous small clinical trials of essential tremor, making the results of these trials difficult or impossible to compare.[1]

The Fahn-Tolosa-Marín scale has been used extensively in clinical trials of ET, at times with modifications to enhance relevance to ET.[2] There is still no comprehensive item-by-item analysis of this scale, and interrater reliability has been fair to poor, especially for writing and drawing.[3–5] Furthermore, severe extremity tremor is defined as >4 cm, which is much less than advanced ET.[6]

The tremor rating scales of Bain and coworkers assess tremor in the head, limbs, Archimedes spirals, and handwriting using 0–10 ratings.[7,8] Ratings are defined subjectively or by examples in a published manual, making good reliability difficult to achieve.[7]

The original tremor rating scale of Louis and coworkers[9] was subsequently revised "to broaden the applicability of this scale to clinical trials."[10] This scale measures only upper-extremity tremor and requires video training to achieve high interrater reliability.[10]

Recognizing the limitations of existing scales, the Tremor Research Group (TRG) met several times over a period of 9 years to develop the TRG Essential Tremor Rating Assessment Scale (TETRAS; see Appendix 1). TETRAS has a 12-item activities-of-daily-living (ADL) subscale that addresses many of the activities assessed in the ADL scales of Fahn,[6] Louis,[11] Bain,[7] and their coworkers, and TETRAS also has a 9-item performance subscale that quantifies tremor in the head, face, voice, limbs, and trunk. TETRAS was developed with 3 mandates: (1) rapid clinical assessment of ET, (2) no equipment other than pen and paper, and (3) objective metric anchors for each 0–4 rating whenever possible, so as to reduce experiential rating bias and uncertainty. The performance subscale takes less than 10 minutes. Ratings of head tremor, and limb tremor are defined by specific amplitude ranges in centimeters (Table 1), so raters must first estimate the maximum amplitude of tremor and then assign the corresponding rating. The performance subscale allows 0.5-point increments in scoring. The 0.5-point increments in upper-limb tremor ratings are

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Dr. Rodger Elble, Department of Neurology, Southern Illinois University School of Medicine, 751 North Rutledge, PO Box 19643, Springfield, IL 62794-9643, USA; relble@siumed.edu

Funding agencies: This study was supported by a grant from GlaxoSmithKline.

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: 23 April 2012; **Revised:** 10 July 2012; **Accepted:** 27 July 2012

Published online 2 October 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.25162

Table 1. Metric anchors for TETRAS performance measures of head and extremity tremor

Head tremor	Upper limb tremor	Lower limb tremor
0 = no tremor	0 = no tremor	0 = no tremor
1 = <0.5 cm	1 = barely visible	1 = barely visible
2 = 0.5 to <2.5 cm	1.5 = <1 cm	2 = <1
3 = 2.5–5 cm	2 = 1 to <3 cm	3 = 1–5
4 = >5 cm	2.5 = 3 to <5 cm	4 = >5 cm
	3 = 5 to <10 cm	
	3.5 = 10 to <20 cm	
	4 = ≥ 20 cm	

defined by specific ranges of tremor amplitude. For all other items, the 0.5-point increments may be used when the 0–4 integer rating is uncertain.

In a preliminary study, 10 members of TRG simultaneously rated 10 ET patients during the live administration of the TETRAS performance subscale. Excellent interrater reliabilities were found for head and upper-limb tremor.[12] In another preliminary study, 10 members of TRG simultaneously rated 3 patients with mild, moderate, and severe ET and rated the videos of these exams 1 month later. The correlations between video scores and live exam scores were greater than 0.87 for all items except face (0.67) and voice (0.63) tremor.[13] We now estimate the inter- and intrarater reliabilities of the TETRAS performance subscale and the correlation of this subscale with the ADL subscale, using 50 videotaped exams.

Patients and Methods

All studies were performed with the signed written informed consent of the patients and controls and approved by the institutional review board of each institution. Patients with ET and controls with no history of tremor were recruited from the authors' clinics. The patients were diagnosed using Tremor Investigation Group criteria.[14]

Nine of the authors, all movement disorder specialists, videotaped TETRAS exams of 1 control and at least 4 patients. Each specialist was asked to videotape patients with mild, moderate, and severe ET so that the videos were evenly distributed over these levels of severity. The TETRAS ADL and performance subscales were performed during each videotaping. Fifty videos (44 patients and 6 controls) were compiled in random order to a set of DVDs and mailed to the same specialists and 1 other. Each specialist rated all 50 videos. The same videos in a different order were rated by the same specialists 1–2 months after the first rating.

Because of omissions in some of the videos, some of the test items could not be scored for every video. Four of the 10 video raters had no experience or training in TETRAS, and nearly all video omissions came from these 4 raters. Nevertheless, all raters scored 31–46 videos for each item, except the standing item, for which only 19 were scored.

Inter- and intrarater reliability of the performance subscale was assessed with 2-way random-effects intraclass correlations (ICCs), using an absolute agreement definition.

Results

The patients (mean age, 67 years; range, 35–80 years) and controls (mean age, 50 years; range, 27–82 years) had comparable ages, and 27 of 50 participants were men. The average duration of tremor in the 44 patients was 30 years (range, 6–72 years). The distribution of total TETRAS performance scores for the 50 participants was fairly uniform (Appendix 2).

The inter- and intrarater ICCs were greater than 0.85 for all items except face tremor, voice tremor, lower-limb tremor, and trunk (standing) tremor (see Appendices 3 and 4 for details).

The 6 experienced and 4 inexperienced raters did not differ statistically (repeated-measures ANOVA) in their inter- and intrarater reliabilities for any of the test items, but the biggest differences were for the face, lower-extremity, and trunk (standing) items.

The raters performed live TETRAS assessments during the videotaping of the 50 TETRAS exams. The Pearson correlation between total ADL score and total performance score was 0.887 ($P < .0001$; Fig. 1).

Cronbach alpha for the live exams performed during the videotapings was 0.951, and it was 0.968 after removal of the face, lower-limb, and trunk items. Statistically identical results were obtained when Cronbach alpha was computed using the video ratings of each rater. The item-to-total score correlations ranged from 0.88 to 0.95 (mean, 0.91) for all upper-limb items except right upper-limb postural tremor with the limb extended forward (0.75). Item-to-total correlations for the head (0.69), face (0.45), voice (0.68), lower limb (0.60), and trunk (0.46) were lower.

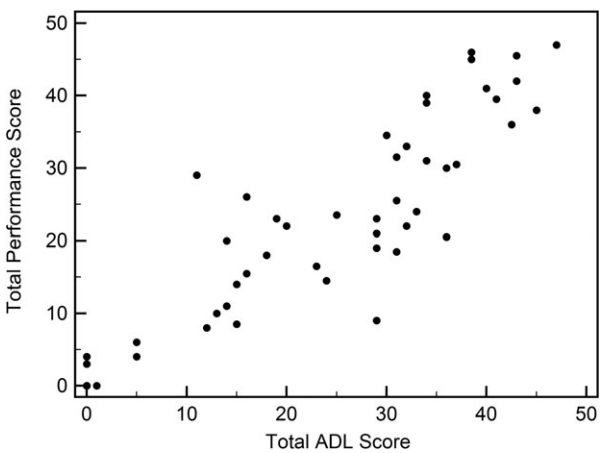

FIG. 1. Total ADL and performance scores for the 50 patients and controls.

Discussion

Our use of performance ratings defined in terms of specific amplitude ranges (in centimeters) resulted in exceptional inter- and intrarater reliabilities, even for raters without prior experience or training with this scale. However, all raters were experienced movement disorder specialists, and it remains to be determined if raters with less expertise perform as well. Comparable interrater reliabilities were found in our earlier preliminary study, in which 10 TRG members, all developers of TETRAS, simultaneously rated 10 ET patients during the live administration of the TETRAS performance subscale.[12]

The TETRAS ADL and performance subscales have obvious content validity for ET, and there is also evidence of strong construct validity. The TETRAS ADL and performance scores are highly correlated, and the TETRAS performance items of upper-extremity function correlate strongly with transducer measures of upper-limb tremor.[15–17]

TETRAS was designed specifically for ET and has not been tested in children or other patient populations. Its metric anchors are too large for small children. TETRAS does not include an assessment of rest tremor because rest tremor is usually not present in ET.[18] Furthermore, distinguishing rest tremor in ET from postural tremor in the presence of incomplete relaxation is difficult and is best avoided when the principal goal is assessment of ET severity rather than the diagnosis or complete characterization of ET in a particular patient.

TETRAS is heavily weighted by upper-extremity tremor and is arguably not ideal when the principal interest is tremor elsewhere. The least reliable items of the TETRAS performance subscale were face tremor, lower-limb tremor, and trunk tremor while standing. This has been the experience with other tremor scales.[7] These items could be deleted with little loss of content validity because the current clinical definition of ET focuses on upper-extremity tremor and head tremor.[14] Exclusion of the face, lower-limb, and trunk (standing) items would reduce the maximum total performance score from 64 to 52.

Finally, most of us thought that the anchor ''barely perceptible'' for grade 1 upper- and lower-extremity tremor could be improved by adding ''<0.5 cm'' or ''a few millimeters.'' Most also thought that lower-extremity tremor was difficult to distinguish from other irregularities of movement and posture. Some raters performed heel-knee-shin testing with the patient seated, and others performed this test with the patient supine, which seems preferable for the specific assessment of tremor. Training would improve the consistency and quality of administering the lower-extremity exam, but it is unclear whether training can reduce the uncertainty in rating tremor.

In conclusion, TETRAS was designed specifically for the clinical measurement of ET severity, requiring no instruments other than a pen and paper. The scale is brief, valid, and highly reliable, and we believe it is ideal for ET clinical trials.

References

1. Deuschl G, Raethjen J, Hellriegel H, Elble R. Treatment of patients with essential tremor. Lancet Neurol. 2011;10:148–161.
2. Zesiewicz TA, Elble RJ, Louis ED, et al. Evidence-based guideline update: treatment of essential tremor: Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology. 2011;77:1752–1755.
3. Elble RJ, Lyons KE, Pahwa R. Levetiracetam is not effective for essential tremor. Clin Neuropharmacol. 2007;30:350–356.
4. 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.
5. 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.
6. Fahn S, Tolosa E, Marín C. Clinical rating scale for tremor. In: Jankovic J, Tolosa E, eds. Parkinson's Disease and Movement Disorders. 2nd ed. Baltimore, MD: Williams & Wilkins; 1993:225–234.
7. Bain PG, Findley LG, Atchison P, et al. Assessing tremor severity. J Neurol Neurosurg Psychiatry. 1993;56:868–873.
8. Bain PG, Findley LJ. Assessing Tremor Severity: A Clinical Handbook. London: Smith-Gordon, 1993.
9. Louis ED, Ford B, Bismuth B. Reliability between two observers using a protocol for diagnosing essential tremor. Mov Disord. 1998;13:287–293.
10. Louis ED, Barnes L, Wendt KJ, et al. A teaching videotape for the assessment of essential tremor. Mov Disord. 2001;16:89–93.
11. 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.
12. Elble R, Comella C, Fahn S, et al. The essential tremor rating assessment scale (TETRAS). Mov Disord. 2008;23(Suppl 1):S357.
13. Tintner R, Group TR. The Tremor Rating Scale (TRS). Mov Disord. 2004;19:1131–1132.
14. Deuschl G, Bain P, Brin M. Consensus statement of the Movement Disorder Society on Tremor. Ad Hoc Scientific Committee. Mov Disord. 1998;13:2–23.
15. 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.
16. 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;129:2660–2666.
17. Mostile G, Fekete R, Giuffrida JP, et al. Amplitude fluctuations in essential tremor. Parkinsonism Relat Disord. 2012 [Epub ahead of print].
18. Quinn NP, Schneider SA, Schwingenschuh P, Bhatia KP. Tremorsome controversial aspects. Mov Disord. 2011;26:18–23.