Evaluation of three different ways of assessing tremor in multiple sclerosis

S H Alusi, J Worthington, S Glickman, L J Findley, P G Bain

Abstract

Objectives—To examine the comparative reliability and validity of three simple ways of rating upper limb tremor in patients with multiple sclerosis.

Methods—Three examiners independently rated severity of upper limb tremor in patients with multiple sclerosis on a 0–10 scale by studying videotape recordings of patients' examinations, spiral drawings, and handwriting samples. The correlations of the tremor severity scores with scores from arm dexterity tests and a tremor related disability scale were also assessed.

Results—Rating tremor on posture had a good intrarater and interrater reliability. However, these reliabilities decreased when kinetic tremor was assessed, in part because dysmetria was a confounding factor. The intrarater reliabilities of rating tremor from spirals and handwriting were also good but the interrater reliabilities were only fair to moderate. Tremor severity scored by all three methods correlated highly with scores obtained from the nine hole peg test, finger tapping test, and a tremor related activities of daily living (ADL) questionnaire, indicating that all three methods were valid ways of assessing tremor in multiple sclerosis.

Conclusion—Multiple sclerosis tremors in posture can be scored using a clinical rating scale in a valid and reliable way, and from spirals and handwriting samples if the ratings are carried out by the same examiner. However, scoring kinetic tremor was less reliable. In addition, the nine hole peg and finger tapping tests provide useful objective assessments of upper limb function in tremulous patients with multiple sclerosis.

(J Neurol Neurosurg Psychiatry 2000;68:756–760)

Keywords: tremor; multiple sclerosis; scale

Upper limb tremor in multiple sclerosis was present in 55 of 100 randomly selected patients in a study conducted from a multiple sclerosis unit in north west London and was disabling in at least one third of these.¹ Another study had found that moderate and severe tremor were present in 32% and 6% of patients respectively.² Multiple sclerosis tremor manifests on action; including posture (postural), during movement (kinetic) or both and can be embedded in a complex ataxic movement disorder making accurate grading of tremor difficult.^{2,3} Various different tremor rating

scales have been deployed in therapeutic trials involving patients with multiple sclerosis tremor but these have not been tested for their reliability and sensitivity to change. Most of these tremor scales are simple with four or five points.^{4–6} A valid, precise, and reliable scale (that ideally takes into account the various components of the complexities of multiple sclerosis tremor) is required if comparisons between multiple centres and efficacy of different interventions are to be assessed accurately.

The 0–10 clinical tremor severity score devised by Bain *et al* has been shown to be a reliable and valid method of measuring essential and dystonic tremors.⁷ This 10 step grading system has the advantages of providing a precise scale that can be easily used in a clinical setting to assess tremor in a specific body part during posture and movement.⁷ However, its use for scoring impairment caused by multiple sclerosis tremors, in which other ataxic elements may complicate the picture, has not been previously examined. This study evaluates the construct validity, intrarater, and interrater reliability of this scale when used in three different ways to assess upper limb tremor in patients with multiple sclerosis. The scale was applied by three raters scoring the severity of tremor in the upper limbs (1) during action (posture and movement), (2) in writing, and (3) in spiral drawing. The comparative reliability of rating multiple sclerosis tremor in these three ways was determined.

Patients and methods

Ethical approval for the project was obtained from the Riverside research ethics committee, London. Patients with a definite diagnosis of multiple sclerosis (Poser, 1983) and associated upper limb tremors were recruited for this study from outpatient clinics at Charing Cross Hospital and the multiple sclerosis unit at Central Middlesex Hospital. Patients with other known neurological problems and those with tremor associated with other medical problems were excluded from this study. Patients with profound weakness of the upper limbs (power grade <3/5 MRC scale) were also excluded. Twenty six (65%) of the patients studied had normal power in the arms. The rest had mild to moderate weakness (power 3–4/5 MRC scale). Sensory impairments were demonstrated in one or both arms of half (20/40) the patients in the study (14 patients had abnormal light touch and pinprick sensation, 11 reduced vibration sense, and five impaired joint position sense).

Department of
Neurosciences,
Imperial College
School of Medicine,
London, UK
S H Alusi
S Glickman
P G Bain

Multiple Sclerosis
Unit, Central
Middlesex Hospital,
London, UK
J Worthington
S Glickman

Essex Neurosciences
Unit, Oldchurch
Hospital, Essex, UK
L J Findley

Correspondence to:
Dr Sundus Alusi,
Department of Psychological
Medicine, Morrison
Hospital, Swansea SA6 6NL,
UK
sundus@alusi.fsnet.co.uk

Received 2 July 1999 and in
revised form
18 January 1999
Accepted 18 January 1999

Table 1 Interpretation of κ scores

κ Coefficient	Strength of agreement
<0	Poor
0–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

Table 2 The intrarater and interrater reliability for rating tremor in two postures (P1 and P2) and during movement (M)

Tremor scoring	Intrarater reliability (weighted κ values)		Interrater reliability (weighted κ values)		
	Examiner 1	Examiner 2	1 v 2	1 v 3	2 v 3
P1	0.76*	0.82*	0.80*	0.32*	0.35*
P2	0.70*	0.69*	0.81*	0.49*	0.53*
M	0.37*	0.77*	0.76*	0.39*	0.42*

* $p < 0.001$.

RATING TREMOR AT REST, ON POSTURE, AND DURING MOVEMENT

Forty two video recordings were made from 30 patients. Five patients had more than one recording (two or three) made after various interventions, at least 3 months apart. There were 14 men and 16 women, average age was 42.9 years (SD 10.3), range (23–70), and average disease duration was 17.4 years (SD 7.0). The patients' median expanded disability status score (EDSS) was 6.5 (range 1–9). Twenty four of the patients were right handed and six were left handed.

The patients were examined and videotaped in the sitting position. The rest component of tremor was examined with the arms relaxed and supported in the patient's lap; however, this was excluded from the analysis as none of the patients had rest tremor. The postural component was examined in two postures: (1) with the arms outstretched and the hands pronated (P1), and (2) with the arms flexed at the elbows, and abducted at the shoulders to 90 degrees with the hands pronated and the fingers held near to the nose ("the batwing" position) (P2). The movement component was examined during a finger-nose-finger test (M). The three clinicians were asked to rate tremor severity for each of these actions separately on the 0–10 scale. Two of the raters repeated the scoring on the same video recordings 15 months later to assess intrarater reliability. For this study only the right hand scores were used. This was done to eliminate bias when scoring left arm tremors as previous work by us had shown that the tremor severity between the hands of each patient were correlated. The associated dysmetria and dysidiadochokinesia

Table 3 The intrarater and interrater reliability of scoring dysmetria and dysidiadochokinesia on a 0–4 scale (κ values)

Ataxia scale	Intrarater reliability (κ values)		Interrater reliability (κ values)		
	Examiner 1	Examiner 2	1 v 2	1 v 3	2 v 3
Dysmetria	0.35*	0.45*	0.59*	0.41*	0.40*
Dysidiadochokinesia	0.59*	0.47*	0.58*	0.53*	0.33*

* $p < 0.001$.

Table 4 The intrarater and interrater reliability for rating tremor from spirals or handwriting

Examiners	Intrarater reliability (weighted κ values)			Interrater reliability (weighted κ values)		
	1	2	3	1 v 2	1 v 3	2 v 3
Right hand spiral	0.73*	0.82*	0.62*	0.46*	0.43*	0.52*
Left hand spiral	0.75*	0.73*	0.60*	0.47*	0.30*	0.56*
Dominant handwriting	0.80*	0.79*	0.64*	0.43*	0.28*	0.37*

* $p < 0.001$.

were also rated on a 0–4 scale by looking at patients' movements while reaching out and touching a target and performing alternating hand movement respectively (appendix 1). For the purposes of this study all rhythmic tremulous movements were considered to be the result of tremor (defined as a rhythmic oscillation of a body part). Thus the raters were asked to score the rhythmic oscillations of the upper limb or the manifestations of that movement in writing or drawing specimens.

RATING TREMOR FROM SPIRAL DRAWINGS AND HANDWRITING SAMPLES

Twenty three patients, seven men and 16 women, provided spiral drawings and handwriting samples. The patients' average age was 43.0 years (SD 11.9), range (18–67), with an average disease duration of 17.3 years (SD 10.4). Median EDSS was 6.5 (range 1–9). Twenty of the patients were right handed and three left handed.

The patients were asked to draw an Archimedes spiral and to write the phrase "Mary had a little lamb". All samples were performed with the patients seated and the forearm supported on a table. Spirals were drawn with both hands and handwriting samples were obtained from the dominant hand only. Some patients with very severe tremor were unable to perform all the tasks. Thus, 21 patients completed a spiral with the right hand, 22 with the left hand, and 20 patients provided a handwriting sample. All the samples obtained were photocopied twice, then the photocopies were individually number coded and shuffled. This produced a total of 86 spirals and 40 handwriting samples. The three raters were then asked to score the spirals and handwriting on the 0–10 scale while referencing the book "Assessing Tremor Severity" (Bain and Findley 1993).⁸ The raters were blinded to the fact that each sample had been duplicated to assess the intrarater reliability.

VALIDATION OF THE TREMOR CLINICAL GRADING SCORES

All patients underwent the following arm function tests before being scored on the tremor scale:

(1) The nine hole peg test (9HPT): Patients were asked to place nine pegs in holes with the dominant hand and then repeat the test with the non-dominant hand as previously described.⁹ The test ended when all nine pegs were placed or after a maximum of 50 seconds. The speed of the manoeuvre was then calculated as the number of pegs per second.

(2) Finger tapping test (FTT): Patients were asked to tap a key on a large calculator with their index finger with the dominant hand and then repeat this with the non-dominant hand for 10 seconds as previously described.¹⁰

(3) Activities of daily living self questionnaire (ADL): This questionnaire consists of a list of 25 activities that could be affected by tremor. In this questionnaire patients are asked to circle a number (from 1–4) that describes most accurately how easy or difficult it is to perform that activity. The sum of the scores for each item is

Table 5 Correlations of tremor ratings with scores obtained from hand function tests and tremor related ADL questionnaire

Tremor assessments	Finger tapping test	9 Hole peg test	Tremor (ADL questionnaire)
On posture (P2)	-0.61 (n=25/27, p<0.005)	-0.62 (n=25/27, p<0.005)	0.58 (n=25/27, p<0.005)
Spirals (dominant hand)	-0.68 (n=21/21, p<0.05)	-0.74 (n=16/21, p<0.005)	0.77 (n=20/21, p<0.005)
Spirals (non-dominant hand)	-0.70 (n=22/22, p<0.005)	-0.87 (n=16/22, p<0.005)	0.50 (n=21/22, p<0.05)
Handwriting	-0.74 (n=20/20, p<0.005)	-0.78 (n=15/20, p<0.005)	0.76 (n=19/20, p<0.005)

then converted into a percentage indicating the level of tremor related disability (the higher the score, the more disabled the patient; appendix 2)).

STATISTICAL ANALYSIS

The intrarater and interrater reliabilities of rating tremor on posture and during movement, from spiral drawings and from handwriting, were calculated using Cohen's κ coefficient and weighted κ .¹¹ This provides a measure of the degree of interobserver agreement for pairs of observers assigning individual observations subjectively to one of a range of categories and also comparisons made by a single observer at two different times.

The weighting system adopted was modified from the standard κ weighting so that:

- (1) A weighting of 1 was assigned when all raters assigned equal scores (perfect agreement).
 - (2) A weighting of 0.8 was given when the scores differed by ± 1 out of 10 among the raters.
 - (3) Any other score was given zero weight.
- Thus, only scores in perfect agreement or deviation by ± 1 between raters were permissible.

Although this weighting system is more rigorous than the conventional κ weighting (table 1)) and thus reduces the κ values obtained, we considered a variation of ± 1 out of 10 to be reasonable in the context of a clinical trial, whereas interrater differences of $> \pm 1$ would be less acceptable. Although this decision is arbitrary, it is based on the authors' previous experiences of rating other forms of tremor with this scale.

Construct validity is defined as the extent to which results obtained using a measure concur with the results predicted from the underlying theoretical model. If it is accepted that manual dexterity is affected by ataxia, then the results of a measure of ataxia should correlate with an independent measure of dexterity. If the correlation is perfect, the measure becomes redundant but this is rarely the case. If there were no correlations at all, then one of these two measures would probably be invalid.¹² The construct validity of rating severity of tremor by each of the three methods was evaluated by calculating the Spearman correlation coefficients between the mean tremor scores given by the three raters with the patients' scores on the arm function tests. These were the FTT, 9HPT, and the ADL self questionnaire. The patients' tremor scores given on the second posture (P2) were used in the analysis as this position is most similar to that used in the functional tasks such as handwriting and spiral drawing.

Results

RELIABILITY WITHIN AND BETWEEN RATERS

Rating tremor on posture and during movement

Analysis was performed on 27 patients as in three patients right arm weakness interfered with their tremor grading for this arm. The κ coefficients for the intrarater and interrater reliability of clinical grading using the 0–10 scale are shown in table 2. Overall the rating scale had substantial to almost perfect intrarater and fair to substantial interrater reliability for assessment of the postural components of tremor in patients with multiple sclerosis. The strength of agreement of the ratings of kinetic tremor were fair to substantial for both intrarater and interrater reliability. Similarly, the examiners had fair to moderate intrarater reliability in assessing dysmetria and dysidiadochokinesia as is seen in table 3.

Rating tremor from spiral drawings or handwriting

The strength of agreement for the intrarater reliability of each of the three examiners rating tremor from spirals was on the whole substantial to almost perfect (table 4). The strength of agreement for interrater reliability was fair to moderate (table 4).

As with spirals, rating tremor from handwriting samples was associated with good intrarater reliability and less interrater reliability (table 4).

CONSTRUCT VALIDITY

Spearman's correlation coefficients for the relations between each of the three tremor assessment methods and the FTT, 9HPT, and the ADL questionnaire are shown in table 5. Two patients did not return their tremor ADL questionnaires. Two patients who were videotaped did not do the FTT and 9HPT and five of the spirals/handwriting group did not do the 9HPT. Right arm postural tremor scores correlated well with right arm FTT and 9HPT scores (table 5). There was also a good correlation of postural tremor scores and patient perceived disability as quantified by the tremor ADL questionnaire. Tremor scores from spiral drawings of both dominant and non-dominant hands and dominant handwriting had a high correlation with the 9HPT (table 5). However, tremor scores from the non-dominant hand spirals correlated less with the tremor ADL, as would be expected, because most of the items on the scale are usually performed by the dominant hand.

Discussion

To our knowledge there is only one other published study examining the reliability of a clinical scale for scoring tremor in multiple sclerosis. Hooper *et al* developed a modified version of Fahn's tremor rating scale to accommodate goal directed tremor and studied its reliability in patients with multiple sclerosis.^{13 14} However, in their study, Spearman's correlation coefficient analysis and Friedman's analyses of variance were carried out to determine reliability and the validity of the scale was not assessed.

Ataxia scales can be used for rating tremor in multiple sclerosis but tend to lack precision and have not been validated for measuring tremor.^{15 16} Thus a valid and reliable way of measuring multiple sclerosis tremor would be useful. Consequently, we examined the validity and comparative reliability of the 0–10 tremor severity scale, which had already been validated in other tremulous conditions.⁷ This 10 point scale was chosen to reduce the percentage error caused by a single unit discrepancy between the raters. The smaller the number of gradations on the scale, the less sensitive it is and the greater the error caused by a 1 unit difference between raters' scores. However a 0–10 scale also suffers from "bottom" and "ceiling" effects and it is more difficult for the raters to "place" a score on an expanded scale although the scale that we used has additional verbal cues (mild, moderate, severe) to help accurate scoring.⁷

All three raters found this 0–10 scale to be rapid and easy to use. The scale was found to be a valid way of assessing tremor in these patients. The most reliable test was rating tremor on posture (P2) which had good intrater and interrater reliabilities. All the patients irrespective of tremor severity or accompanying deficits, could be rated on posture, giving this method an advantage over handwriting and spiral analysis, which required patients to have a certain amount of dexterity.

The limitations of rating multiple sclerosis tremors from spirals and handwriting samples were disclosed in this study. Both of these methods have a low ceiling effect in that some patients with incapacitating tremors were unable to draw spirals or write although their tremors could be rated from simple postures and movements. Secondly, the interrater reliability of scoring tremor from spirals and handwriting was lower than that when tremor was scored on posture (handwriting was worse) (see table 6). This gradient in the interrater reliabilities reflects increasing disagreement among the three raters

as the complexity of movement and perhaps ataxia in the three tasks increased.

The 0–10 scale was found to be least reliable for rating kinetic tremor. Two possible explanations for this finding are: firstly, that multiple sclerosis tremors can vary throughout a movement making it difficult for the examiner to know what magnitude is representative, a problem previously encountered in a study of essential and dystonic tremor.⁷ Secondly, multiple sclerosis tremor is confounded by other ataxic deficits that manifest during movement which blur the phenotypic features of tremor making grading difficult and perhaps also more subjective.

The difficulty of distinguishing between tremor and other ataxic deficits is also shown by the problems involved in scoring dysmetria and dysdiadochokinesia in these tremulous patients; the 0–4 ataxia scale used in this study produced a fair to moderate intrater and interrater reliability, although this may in part have been a product of the scale's design.

The results obtained from the 9HPT and the FTT correlated well with postural tremor scores and thus may provide useful objective methods for assessing arm dexterity in tremulous patients with multiple sclerosis. This is particularly useful for clinical trials evaluating upper limb function as well as the efficacy of tremor treatments in patients with multiple sclerosis. Furthermore, the results indicated that spiral drawing and handwriting samples correlated highly with patients' perceived tremor disability and hence, if used by the same rater, provide valid and reliable measures of tremor in these patients.

In summary, all three ways of rating tremor in patients with multiple sclerosis are valid. The most reliable method was shown to be scoring tremor on posture (P2). The same assessor should rate spirals and handwriting samples, if these methods are deployed with patients with multiple sclerosis. Kinetic tremor and associated ataxic deficits are more difficult to score reliably. On the other hand, the 9HPT and FTT provide results that correlate well with patients' tremor induced disability, but are also susceptible to other impairments of upper limb function.

We are grateful to Mrs. Caroline Dore; senior statistician, Imperial School of Medicine, London, for her statistical advice. We are grateful to the authors of the ataxia scale (appendix 1) who could not be identified by a search of the literature. We also thank SEARCH for providing a grant that enabled us to perform this work.

Appendix 1: dysmetria and dysdiadochokinesia scale

Dysdiadochokinesia:

- 0 No problem
- 1 Mild but detectable clumsiness and slowing of pronation-supination rate
- 2 Moderate clumsiness and slowing of pronation-supination rate
- 3 Severe clumsiness and slowing of pronation-supination rate
- 4 Unable to perform repetitive sequential movements

Dysmetria:

- 0 No impairment
- 1 Mild dysmetria but reaches the target
- 2 Moderate dysmetria, reaches target after several attempts
- 3 Severe dysmetria, short of target after many attempts
- 4 Cannot use hands

Table 6 Summary of the averaged weighted κ results; showing substantial intrater reliability (except for kinetic tremor) and lower interrater reliabilities for all the methods

Rating tremor method	Intrater reliability (mean)	Interrater reliability (mean)
On posture P1	0.79 (substantial)	0.49 (moderate)
On posture P2	0.70 (substantial)	0.61 (substantial)
During movement M	0.57 (moderate)	0.52 (moderate)
Spirals	0.72 (substantial)	0.47 (moderate)
Handwriting	0.74 (substantial)	0.36 (fair)

Appendix 2: tremor activities of daily living questionnaire

(please read carefully. For each item circle the number which best describes how easy or difficult it is for you to perform the activity)

- 1 Able to do the activity without difficulty
2 Able to do the activity with little effort
3 Able to do the activity with a lot of effort
4 Cannot do the activity by yourself

How well are you able to.....

1	Cut food with a knife and fork	1	2	3	4
2	Use a spoon to drink soup	1	2	3	4
3	Hold a cup of tea	1	2	3	4
4	Pour milk from a cup or carton	1	2	3	4
5	Wash and dry dishes	1	2	3	4
6	Brush your teeth	1	2	3	4
7	Use a handkerchief to blow your nose	1	2	3	4
8	Have a bath	1	2	3	4
9	Use the lavatory	1	2	3	4
10	Wash your face and hands	1	2	3	4
11	Tie up your shoe laces	1	2	3	4
12	Do up buttons	1	2	3	4
13	Do up a zip	1	2	3	4
14	Write a letter	1	2	3	4
15	Put a letter in an envelope	1	2	3	4
16	Hold and read a newspaper	1	2	3	4
17	Dial a telephone	1	2	3	4
18	Make yourself understood on the phone	1	2	3	4
19	Watch the television	1	2	3	4
20	Pick up your change in a shop	1	2	3	4
21	Insert an electric plug into a socket	1	2	3	4
22	Unlock your front door with the key	1	2	3	4
23	Walk up and down the stairs	1	2	3	4
24	Get up out of an armchair	1	2	3	4
25	Carry a full shopping bag	1	2	3	4

- 1 Alusi SH, Glickman S, Worthington J, et al. A study of tremor in multiple sclerosis. *Proceedings. J Neurol Neurosurg Psychiatry* 1999;**66**:264.

- 2 Haddow LJ, Mumford C, Whittle IR. Stereotactic treatment of tremor due to multiple sclerosis. *Neurosurgery Quarterly* 1997;**7**:23-34.
- 3 Alusi SH, Glickman S, Aziz TZ, et al. Tremor in multiple sclerosis. *J Neurol Neurosurg Psychiatry* 1999;**66**:131-4.
- 4 Speelman JD, Van Manen J. Stereotactic thalamotomy for the relief of intention tremor in multiple sclerosis. *J Neurol Neurosurg Psychiatry* 1984;**47**:596-9.
- 5 Koller WC. Pharmacologic trials in the treatment of cerebellar tremor. *Arch Neurol* 1984;**41**:280-1.
- 6 Sechi GP, Zuddas M, Piredda M, et al. Treatment of cerebellar tremors with carbamazepine: a controlled trial with long term follow up. *Neurology* 1989;**39**:1113-15.
- 7 Bain PG, Findley LJ, Atchison P, et al. Assessing tremor severity. *J Neurol Neurosurg Psychiatry* 1993;**56**:868-73.
- 8 Bain PG, Findley LJ. Assessing tremor severity. In: Bain PG, Findley LJ, eds. *Standards in neurology, series A: Assessment, diagnosis and evaluation*. London: Smith-Gordon, 1993.
- 9 Mathiowetz V, Weber K, Kashman N, et al. Adult norms for the nine-hole peg test of finger dexterity. *The Occupational Therapy Journal of Research* 1985;**5**:24-37.
- 10 Worthington J, De Souza LH. A simple measurement of speed of index finger movement. *Clinical Rehabilitation* 1989;**3**:117-23.
- 11 Francis DA, Bain PG, Swan AV, et al. An assessment of disability rating scales used in multiple sclerosis. *Arch Neurol* 1991;**48**:299-301.
- 12 Wade DT. Measurements in neurological rehabilitation. In: Wade DT, ed. *Measurements in neurological rehabilitation*. Oxford: Oxford University Press, 1992:37.
- 13 Fahn S, Tolosa E, Marin C. Clinical rating scale for tremor. In: Jankovic E, Tolosa E, eds. *Parkinson's disease and movement disorders*. Baltimore: Urban and Schwarzenberg, 1988: 225-34.
- 14 Hooper J, Taylor R, Pentland B, et al. Rater reliability of Fahn's tremor rating scale in patients with multiple sclerosis. *Arch Phys Med Rehabil* 1998;**79**:1076-9.
- 15 Notermans NC, van Dijk GW, van der Graaf Y, et al. *J Neurol Neurosurg Psychiatry* 1994;**57**:22-6.
- 16 Trouillas P, Takayanagi T, Hallet M, et al. International cooperative ataxia scale for pharmacological assessment of the cerebellar syndrome. *J Neurol Sci* 1997;**145**:205-11.