



The inter rater reliability of the original and of the modified Ashworth scale for the assessment of spasticity in patients with spinal cord injury

BM Haas¹, E Bergström², A Jamous² and A Bennie²

¹University of Brighton, Department of Occupational Therapy and Physiotherapy; ²National Spinal Injuries Centre, Stoke Mandeville Hospital, Aylesbury, Bucks, UK

Thirty patients with spinal cord injury (SCI) were randomly selected to participate in this study which evaluated the inter rater reliability of the original and of the modified Ashworth scale for the assessment of spasticity in the lower limbs. A doctor and a physiotherapist rated the muscle tone of hip adductors, hip extensors, hip flexors and ankle plantarflexors according to the original and to the modified Ashworth scale. The results were analyzed using a Cohen's Kappa statistical test and showed varying levels of reliability for different muscle groups and limbs. Kappa values ranged between 0.21 and 0.61 (mean 0.37). The original scale was slightly more reliable than was the modified scale. However, this difference was not significant ($P > 0.05$), and was not consistent between the two limbs and between different muscle groups. It was concluded that the Ashworth scale is of limited use in the assessment of spasticity in the lower limb of patients with SCI. Further work is required to establish a standardised speed of muscle stretching during the test, or to find more appropriate grades and descriptions of spasticity for this patient group. The effects of training of the raters in the use of the scales also warrants further investigation.

Keywords: assessment; Ashworth scale; spasticity; measurement; spinal cord injury

Introduction

A generally accepted definition of spasticity is that of a velocity dependent increase in muscle tone with exaggerated tendon reflexes.¹ The mechanisms of the phenomenon remain poorly understood. Amongst those described in the literature are intrinsic changes in the muscle itself,² increased excitability of either alpha and/or gamma motor neurons in the spinal cord,^{3–5} abolition of pre-synaptic inhibition^{6,7} and plastic adaptive changes in the central nervous system.⁸ This lack of understanding or consensus may contribute to the difficulties in measuring spasticity. The need for the objective quantification of spasticity has been widely accepted.^{9–12} Recent changes in the Health Service in the UK put further emphasis on demonstrating effectiveness of treatment. It is therefore vital to evaluate the reliability and validity of the current methods of measuring spasticity. The literature offers a wide variety of alternatives for measuring spasticity but no single one of them is in wide use.^{13–15,33} Measurement tools which offer a quantifiable method to evaluate spasticity such as the Wartenberg Pendulum test^{16–18} and isokinetic measurements¹⁹ appear too complex and expensive for use in the clinical setting.

Measurements that are simple, inexpensive and appropriate to the clinical setting include rating scales, tendon jerks, myometer measurements and functional measurements. The use of functional measurements however appears inappropriate for measuring an impairment²⁰ such as spasticity since they only make indirect reference to this physiological phenomenon and may not measure spasticity at all.² The only rating scale that has been evaluated^{11,21–23,30} or has been used in recent studies^{14,24} is the Ashworth scale. This scale is available in its original form devised by Ashworth in 1964²⁵ and was modified by Bohannon and Smith in 1987.²⁶ Both scales ask the examiner to move a limb through its full range of movement and rate the amount of resistance felt according to descriptions which are shown in Figures 1 and 2. The modified scale has been evaluated for use with hemiplegic patients^{11,22,26} and the original scale has been evaluated for use with patients suffering from Multiple Sclerosis and SCI.²³ The studies found each scale to be reliable between raters. However, doubts have been raised as to the use of appropriate statistical analysis in these studies and therefore the scales may not after all be reliable.¹³ No study has so far compared the reliability between the two scales and therefore this study has been devised to fulfil this need and to evaluate both scales for the measurement of spasticity in the lower limbs of patients with spinal cord injury.

Methodology

Patients

Thirty subjects were randomly recruited from the patient population of the National Spinal Injuries Centre. All patients in the unit during the period of the study were considered to be included as subjects. Excluded from the study were patients with loss of range of movement in the lower limb and where passive movements were contradicted. One subject had limited range of movement in the left limb and only the ratings of the right limb were included in the study. Informed consent was obtained before testing. Prior to the commencement of the study approval had been obtained from the Aylesbury Vale Local Research Ethics Committee. The final subject list comprised 24 men and six women (mean age 40.3 years, range 17–72) with a mean time from injury of 17.23 months (range 1–294). Table 1 shows the neurological level and the Frankel grades³² of the subjects.

Procedures

Each patient was assessed by a physiotherapist and a doctor during a single session with the patient in a supine position on a bed or plinth. Both assessors

were experienced in the field of SCI rehabilitation. No formal training in the use of the spasticity scales had been offered to them, however they had the descriptions of the rating criteria shown in Figures 1 and 2 available at the time of testing. Testing commenced 5 min after the subjects had been positioned. Muscle tone of the hip adductors was assessed by placing one hand underneath the leg close to the knee, the other hand supported the limb close to the ankle. The patient's straight limb was then moved into full abduction within 1 s. Muscle tone of the hip flexors was assessed with the same hand position and the limb was moved from full flexion to neutral within 1 s. Hip extensor muscle tone was assessed with the previous handholds and the limb was moved from a neutral position into full flexion within 1 s. In assessing plantar flexor muscle tone the assessors placed one hand under the ball of the foot while the other hand stabilised the limb around the ankle joint. The patient's ankle was then moved into full dorsiflexion within 1 s. Patients were asked to relax during the procedures. All movements were performed three times per movement per assessor. Following the performance of the procedure by the first assessor he/she then rated the patient's muscle tone with both spasticity scales. The same procedure

Grade	Description
0	No increase in muscle tone
1	slight increase in muscle tone giving a catch when the limb is moved
2	more marked increase in tone but limb is easily moved
3	considerable increase in tone-passive movement difficult
4	limb rigid in flexion or extension (abduction/adduction)

Figure 1 Original Ashworth scale

Grade	Description
0	no increase in muscle tone
1	slight increase in muscle tone, manifested by a catch and release or by minimum resistance at the end of the range of motion when the affected part(s) is moved in flexion or extension
1+	slight increase in muscle tone, manifested by a slight catch, followed by minimal resistance throughout the remainder (less than half) of the range of movement
2	more marked increase in muscle tone through most of the range of movement, but affected part(s) easily moved
3	considerable increase in muscle tone, passive movement difficult
4	affected part(s) rigid in flexion or extension

Figure 2 Modified Ashworth scale

Table 1

	Neurological level	Frankel grade
1	C7	A
2	T10/11	A
3	C7	B
4	C2	E
5	C7	A
6	C5	D
7	T8	D
8	T6	A
9	C4/5	A
10	L1	C
11	T9	A
12	T4	A
13	C5/6	B
14	C4	D
15	C4	D
16	C5	D
17	T10	D
18	C6	B
19	T4	A
20	T10	A
21	C3	A
22	C6	A
23	C4	A
24	C5	A
25	C6	A
26	C4	A
27	T5	A
28	T7	A
29	C4	C
30	C3	A

was then repeated by the second assessor after a period of 5 min between tests. The order of assessment between the two assessors was randomised and no discussion of the results between the assessors occurred during the duration of the study to ensure they were blind to each others results. Both assessors completed a result sheet immediately after the procedure which was then collected by an independent investigator.

Data analysis

The level of agreement between the assessors was analyzed using the Cohen's Kappa test.^{27,28} The Kappa test is a measure of agreement for variables which can be categorised. Differences between the two spasticity scales were analyzed with an analysis of variance (ANOVA).

Results

An overview of the results for the original scale is given in Table 2 and for the modified scale in Table 3. The percentage agreement indicates how often the raters agreed in their ratings of a particular movement. The Kappa values indicate the measure of agreement corrected for chance. A value of one would indicate total agreement between the raters whereas a value of 0 would mean no agreement. The 'agreement rating' describe the strength of agreement for the Kappa values. These descriptions should only be viewed as qualitative benchmarks and are somewhat arbitrary.²⁹ The order of the descriptions ranges from 'almost

perfect', 'substantial', 'moderate', 'fair', 'slight' to 'poor'.

The *P* values in Tables 2 and 3 indicate that the Kappa values were all significant, most of them highly significant. This shows that the agreements between the assessors did not occur by chance. The means of the Kappa values for the two scales and the two limbs are shown in Table 4. They indicate that the original scale had a slightly higher level of agreement (mean 0.41) than the modified scale (mean 0.34). However, this difference was not significant ($P=0.31$). The right limb (mean 0.40) showed a slightly higher level of agreement than the left limb (mean 0.34) for both scales but again the difference was not significant ($P=0.46$).

The mean values for both scales also showed a difference (non significant with $P=0.13$) in agreement for the individual muscle groups assessed: the Kappa values for adductors were 0.47, for extensors 0.34, for flexors 0.41 and for plantarflexors 0.26. The mean of all Kappa values was 0.37 which is described as fair.

Table 4 Means of Kappa values and % agreements

	Kappa	P	% agreement	P
Original	0.41	0.31	57.2	0.15
Modified	0.34		49.54	
Right	0.40	0.46	55.4	0.45
Left	0.34		51.3	

Table 2 Results for original Ashworth scale: Percentage agreement, Kappa values, *P* value and rating

Muscle group	Leg	% Agreement	Kappa value	P value	Agreement rating
Adductors	right	70.0	0.61	<0.0001	substantial
	left	48.3	0.32	<0.0001	fair
Extensors	right	60.0	0.44	<0.0001	moderate
	left	48.3	0.28	<0.0001	fair
Flexors	right	63.3	0.49	<0.0001	moderate
	left	72.4	0.61	<0.0001	substantial
Plantarflexors	right	43.3	0.21	<0.005	fair
	left	51.7	0.31	<0.0001	fair

Table 3 Results for modified Ashworth scale: Percentage agreement, Kappa values, *P* value and rating

Muscle group	Leg	% Agreement	Kappa value	P value	Agreement rating
Adductors	right	70.0	0.62	<0.0001	substantial
	left	48.3	0.34	<0.0001	fair
Extensors	right	50.0	0.34	<0.0001	fair
	left	48.3	0.32	<0.0001	fair
Flexors	right	46.7	0.28	<0.0001	fair
	left	44.8	0.28	<0.0001	fair
Plantarflexors	right	40.0	0.20	<0.005	slight
	left	48.3	0.29	<0.0001	fair

Table 5 shows the numbers of assignments (allocations to a particular category) and the numbers of agreements in the respective categories for the two scales. It can be seen from the table that the subjects were most often allocated to the category '0' and that there were few subjects with spasticity so severe that it could be described as rigid (category '4'). The numbers of agreements also indicate that the assessors agreed more often on subjects in the '0' category and few agreements occurred in the '1' and '1+' categories of the modified scale.

Discussion

The results of this study showed that the inter rater reliability of the Ashworth scale varied between the original and the modified scale, between different muscle groups and different limbs. The mean for all Kappa values was 0.37 and described as being only fair. From a clinical point of view the scales should therefore only be used with extreme caution. Results of this study can confirm the limited reliability when measuring spasticity in the lower limb which has been reported by others.^{22,30} A direct comparison with these studies is only possible in a limited way since this study used a different statistical test. The Kappa coefficient used in this study was believed to be the most appropriate test since it evaluates the level of agreement between raters which assign a subject to a certain category. The statistical tests used in the previously mentioned studies have either used a Kendall's tau coefficient or a Spearman's rank correlation coefficient. Both of these tests assume a rank order between subjects.²⁸ For the purpose of this study it was necessary to rate the scores of the subject's muscle tone against predetermined categories. This then allowed for the evaluation of the ratings of the two testers. By using the percentage agreements between raters a comparison is possible with the study by Bohannen and Smith.²⁶ They agreed in 86.7% of their ratings and concluded a high inter rater reliability. This study showed the percentage agreement for various muscle groups to range from 40.0% (Plantarflexors/modified scale/right leg) to 72.4% (Flexors/original scale/left leg). The high percentage of agreement in the Bohannen and Smith study may have produced a high Kappa value but unfortunately they did not calculate it. Bohannen and Smith also had ample training in the use of the scale which may have contributed to the high inter rater

reliability. No training was offered to the two raters in this study as this was thought to best reflect the clinical situation. Future work is needed to evaluate the effect of training on the reliability of the scale.

The original scale showed slightly better results than the modified scale. Table 5 shows that few agreements in the modified scale occurred in the grades '1' and '1+' and that the introduction of a new grade may have contributed to this lower level of agreement.

Kappa values for the right leg were slightly better than that for the left leg. One of the raters always assessed the patient's limbs from the right side only, but for this study was made to test each limb from the corresponding side. This unfamiliar position may have caused the lower level of agreement on the left side. This study also showed differences in reliability between different muscle groups. It was poorest for plantarflexors, followed by extensors and flexors and was best for adductors. This does not confirm the results of another study³⁰ which showed that their best results were for plantarflexors. The patterns were also inconsistent for different scales but no reasons for these inconsistencies could be found. The speed of moving the limbs through the range was set for all movements at 1 s. This was timed by counting 'one thousand and one' during the stretch. This standardisation was seen as necessary because spasticity is velocity dependent¹ but further investigation may determine if additional standardisation (eg using a metronome) could help to achieve better results. The speed of movement may also have to vary between different muscle groups.

Table 5 shows that the subjects were graded '0' more often than any other grade and that most agreements between the raters also occurred within the '0' grade. The reliability of the scales may be even less in a more 'spastic' patient group since there were fewer agreements in those patients who actually had an increase in muscle tone.

Grade '0' is described as no increase in muscle tone and covers both subjects with 'low tone' as well as 'normal muscle tone'. The ratings cannot reflect this clinically significant difference and therefore the scales seem not ideally suited to this patient group. The original Ashworth scale was devised for research into the effects of carisopradol on muscle tone in patients with multiple sclerosis.²⁵ There are some differences in the characteristics of spasticity with different aetiologies despite the fact that it may be the same physiological phenomenon in all of them.³¹ Further work may help to establish more appropriate grades for this patient group but this may not necessarily produce a more reliable scale.

Conclusions

The results of this study have shown that the Ashworth scale has only limited inter rater reliability for measuring spasticity in the lower limbs in patients with SCI. Reliability was achieved with some muscle

Table 5 No. of assignments and agreements

<i>Original scale</i>	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>			<i>Total</i>
Assignments	152	115	129	68	8			472
Agreements	62	27	29	17	1			136
<i>Modified scale</i>	<i>0</i>	<i>1</i>	<i>1+</i>	<i>2</i>	<i>3</i>	<i>4</i>		
Assignments	149	55	77	117	66	8		472
Agreements	61	6	6	26	17	1		117

groups but not consistently for both limbs or for both scales. No firm conclusions could be drawn from this study as to the reasons behind this inconsistency. The original Ashworth scale was slightly more reliable than the modified version and doubts have been raised as to the appropriateness of the present grades and their

descriptions for this patient group. Ample training in the use of the scale may improve the reliability but this has yet to be confirmed by further research. The usefulness of the scale which has been demonstrated in other studies could not be confirmed by this investigation.

References

- 1 Lance JW. Pathophysiology of spasticity and clinical experience with baclofen. In: Feldman RG, Young RR and Koella WP (ed). *Spasticity: Disordered Motor Control*. Year Book Medical Publishers, Chicago: 1980, pp 185–220.
- 2 Katz RT, Rymer WZ. Spastic hypertonia: mechanisms and measurement. *Arch Phys Med Rehabil* 1989; **70**: 144–155.
- 3 Pierrot-Deseilligny E. Pathophysiology of spasticity. *Triangle* 1985; **22**: 165–174.
- 4 Dimitrijevic MR. Spasticity. In: Swash M and Kennard C (eds). *Scientific Basis of Clinical Neurology*. Churchill Livingstone, Edinburgh, 1985.
- 5 Rushworth G. Some pathophysiological aspects of spasticity and the search for rational and successful therapy. *Int Rehab Med* 1980; **2**: 23–26.
- 6 Lance JW, Burke D. Mechanisms of spasticity. *Arch Phys Med Rehabil* 1974; **55**: 332–337.
- 7 Burke D. Mechanisms underlying the tendon jerk and H-reflex. In: Delwaide PJ and Yound RR (eds). *Clinical Neurophysiology in Spasticity*. Elsevier, Amsterdam, 1985: chapter 5.
- 8 Kidd G, Lawes N, Musa I. *Understanding neuromuscular plasticity*. Edward Arnold, London: 1992, p 99.
- 9 Pederson E. Clinical aspects of spasticity and measurement of spasticity. In: Thomas C (ed). *Spasticity: Mechanisms, Measurement and Management*. American Lecture Series: 1969; **752**: 36–54.
- 10 Jones EW et al. Comments on the design of an instrument to measure spasticity in the arm – SAM. *Engineering in Medicine* 1982; **11**: 47–50.
- 11 Bodin PG, Morris ME. Interrater reliability of the modified Ashworth scale for wrist flexor spasticity following stroke. *Proceedings Book II, World Federation of Physical Therapy, 11th Congress, London, 1991*, pp 505–507.
- 12 Leslie GC, Muir C, Part NJ, Roberts RC. A comparison of the assessment of spasticity by the Wartenberg pendulum test and the Ashworth grading scale in patients with multiple sclerosis. *Clinical Rehabilitation* 1992; **6**: 41–48.
- 13 Haas BM. Measuring spasticity: A survey of current practice among health care professionals. *British Journal of Therapy and Rehabilitation* 1994; **1**: 90–95.
- 14 Katz RT, Rovai GP, Brait C, Rymer Z. Objective quantification of spastic hypertonia: Correlation with clinical findings. *Archives of Physical Medicine and Rehabilitation* 1992; **73**: 339–347.
- 15 Chan A, Hugos C, Morrison S, Theriot K. Balance and spasticity: What we know and what we believe. *J Neuro Rehab* 1994; **8**: 119–130.
- 16 Bajd T, Vodovnik L. Pendulum testing for spasticity. *Journal of Biomedical Engineering* 1994; **6**: 9–16.
- 17 Bajd T and Bowman B. Testing and modelling of spasticity. *Journal of Biomedical Engineering* 1982; **4**: 91–96.
- 18 Bohannon RW. Variability and reliability of the pendulum test for spasticity using a Cybex II isokinetic dynamometer. *Physical Therapy* 1987; **67**: 659–661.
- 19 Firoozbakhsh KK, Kunkel CF, Scemin AME, Moneim MS. Isokinetic dynamometric technique for spasticity assessment. *Am J Phys Med Rehabil* 1993; **72**: 379–385.
- 20 WHO. *International Classification of Impairments, Disabilities and Handicaps*. World Health Organisation, Geneva: 1980, p 47.
- 21 Wade DT. *Measurement in Neurological Rehabilitation*. Oxford University Press: Oxford, 1992, p 17.
- 22 Sloan RL et al. Inter-rater reliability of the modified Ashworth scale for spasticity in hemiplegic patients. *International Journal of Rehabilitation Research* 1992; **15**: 158–161.
- 23 Lee KC, Carson L, Kinnin E, Patterson V. The Ashworth scale: a reliable and reproducible method of measuring spasticity. *J Neur Rehab* 1989; **3**: 205–209.
- 24 Brar SP et al. Evaluation of treatment protocols on minimal to moderate spasticity in Multiple Sclerosis. *Arch Phys Med Rehabil* 1991; **72**: 186–189.
- 25 Ashworth B. Preliminary trial of carisoprodol in Multiple Sclerosis. *Practitioner* 1964; **192**: 540–542.
- 26 Bohannon RW, Smith MB. Interrater reliability of a modified Ashworth scale of muscle spasticity. *Physical Therapy* 1987; **67**: 206–207.
- 27 Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960; **20**: 37–46.
- 28 Siegel S, Castellan NJ. *Nonparametric Statistics for the Behavioural Sciences*. McGraw-Hill, New York, 1988, pp 285–290.
- 29 Fleiss JL. The measurement of inter rater agreement. In: *Statistical Methods for Rates and Proportions*. John Wiley, New York, 1981, pp 147–150.
- 30 Nuyens G et al. Inter-rater reliability of the Ashworth scale in multiple sclerosis. *Clinical Rehabilitation* 1994; **8**: 286–292.
- 31 Noth J. Trends in the pharmacophysiology and pharmacotherapy of spasticity. *J Neurol* 1991; **238**: 131–139.
- 32 Frankel HL et al. The value of postural reduction in the initial management of closed injuries to the spine with paraplegia and tetraplegia. *Paraplegia* 1969; **7**: 179–192.
- 33 Haas BM, Crow JL. Towards a clinical measurement of spasticity. *Physiotherapy* 1995; **81**: 474–480.