A COMPARATIVE STUDY ON THE INTER-RATER RELIABILITY OF THE ASHWORTH SCALES IN ASSESSMENT OF SPASTICITY

N. Nakhostin-Ansari*, S. Naghdi, H. Moammeri and S. Jalaie

Department of Physiotherapy, School of Rehabilitation, Tehran University of Medical Sciences, Tehran, Iran

Abstract- Ashworth and modified Ashworth scales are the most widely used tests to assess the severity of muscle spasticity. These clinical scales offer qualitative and subjective information and consequently there are issues concerning reliability when two or more clinicians are involved in assessment of spasticity. This article presents the result of a study assessing inter-rater reliability of the original and of the modified Ashworth scales for the assessment of elbow flexor muscle spasticity in patients with hemiplegia. Thirteen patients with hemiplegia (8 men and 5 women) participated in this study. Two physiotherapists rated the muscle tone of elbow flexors according to the original and to the modified Ashworth scales. Each patient was assessed during a single session in a supine position on a bed with the arms at the sides of the body. Movements were performed three times by each assessor. No discussion of the results between the assessors occurred during the course of the study to ensure they were blind to each others' results. Kappa values for the original Ashworth and the modified Ashworth scales were 0.22 (SE 0.27, $P = 0.43$) and 0.24 (SE 0.23, $P = 0.24$), respectively. The modified Ashworth scale was slightly more reliable than was the original scale but this difference was not significant ($P > 0.05$). Inter-rater reliability of the original and of the modified Ashworth scales in the assessment of elbow flexor spasticity was poor and therefore these spasticity scales may not be valid.

Acta Medica Iranica, 44(4): 246-250; 2006

Key Words: Spasticity, Ashworth scales, reliability, assessment, hemiplegia

INTRODUCTION

Spasticity is one of the most common disabling features of the motor disorders associated with the upper motor neuron syndrome (1). It is a serious debilitating problem that creates great difficulty for both patients and clinicians (2). Traumatic brain injury, anoxic brain injury, stroke, spinal cord injury, cerebral palsy, spinal degenerative diseases such as multiple sclerosis and familial spastic paraparesis are common causes of spasticity (1, 3). Spasticity is a complex phenomenon. Lance defines spasticity as "a velocity dependent increase in tonic stretch reflex (muscle tone) with exaggerated tendon reflexes, resulting from the hyperexcitability of the stretch reflex, as one component of the upper motor neuron syndrome" (4).

The quantification of spasticity has been a difficult and challenging problem. Muscle tone tests measuring the resistance to passive movement are commonly used to quantify spasticity, although factors other than spasticity such as dystonia, rigidity and other mechanical factors may also be involved. Ashworth developed a scale that has been widely used clinically and included a five point ordinal scale of tone severity (5). In 1987, Bohannon and Smith

Received: 14 Apr. 2004, Revised: 4 Apr. 2005, Accepted: 5 July 2005

*** Corresponding Author:**
N. Nakhostin-Ansari, Department of Physiotherapy, School of Rehabilitation, Tehran University of Medical Sciences, Tehran, Iran
Tel: +98 21 77533939
Fax: +98 21 77882009
E-mail: nakhostin@sina.tums.ac.ir

modified Ashworth scale (6). They added the grade '1+' and slightly modified the definitions. 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 (table 1). Although these clinical scales are widely used in researches (7-8), there are relatively few data available on the reliability of them. There is only one study that has compared the reliability between the two scales (9). The results showed varying levels of reliability for different muscle groups and limbs. The original scale was slightly more reliable than the modified scale but this difference was not significant and it was concluded that both Ashworth scales are of limited use in the assessment of spasticity in the lower limb of patients with spinal cord injury. No study has so far compared the reliability between the two scales in hemiplegic patients.

The purpose of this study was to examine the inter-rater reliability of Ashworth scales in assessing spasticity of elbow flexors in patients with hemiplegia and to compare the reliability between the two scales.

MATERIALS AND METHODS

Thirteen subjects with hemiplegia (8 men and 5 women) attending the physiotherapy clinic of the rehabilitation faculty were recruited. Subjects were those who could follow instructions. Exclusion criteria were loss of range of motion, contraindication of passive movements and taking tone modifying drugs. We obtained informed consent from all participants.

The mean age of the subjects was 46.5 years (SD = 19.7, range = 13-70). The patients had hemiplegia due to cerebrovascular accident (9 subjects), tumor (3 subjects) and head trauma (1 subject). The mean time since injury was 22 months (SD = 35.6). Seven patients had right hemiplegia and six patients had left hemiplegia.

Two female physiotherapists provided all original Ashworth scale (OAS) and modified Ashworth scale (MAS) scores for all subjects. Each patient was assessed by two physiotherapists during a single session. The testing area was quiet and screened from other patients and therapists.

Patients were positioned in supine position on a bed. Patients' heads were in midline with arms at the sides of the body. Testing commenced five minutes after the subjects had been positioned. Elbow flexors of the involved side were assessed. Patient's elbow was extended from a position of maximal possible flexion to maximal possible extension over a duration of about one second (by counting one thousand one).

Patients were asked to be relaxed during the procedure. Each assessor was allowed to perform three (9) such sequential extension but they were allowed to score the tone based on first stretch. The same procedure was then repeated by the second assessor after a period of five minutes. The order of assessment between the two assessors was randomized.

Table 1. Definitions of the Ashworth and modified Ashworth scales

Score	Ashworth Scale (Ashworth, 1964)	Modified Ashworth Scale (Bohannon and Smith, 1987)
0	No increase in tone	No increase in muscle tone
1	Slight increase in tone giving a catch when the limb was moved in flexion or extension	Slight increase in muscle tone, manifested by a catch and release or by minimal 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 catch, followed by minimal resistance throughout the remainder (less than half) of the ROM (range of motion)
2	More marked increase in tone but limb easily flexed	More marked increase in muscle tone through most of the ROM, but affected part(s) easily moved
3	Considerable increase in tone – passive movement difficult	Considerable increase in muscle tone, passive movement difficult
4	Limb rigid in flexion or extension	Affected part(s) rigid in flexion or extension

No discussion of the results between the assessors occurred during the test sessions and duration of the study to ensure they were blinded to each others' results. Both therapists completed a result sheet immediately after the procedure which was then collected by the first author of the paper. He was also blind to the raters' results during the period of study. One of the physiotherapists was experienced in the field of physiotherapy of patients with hemiplegia but the other one was less experienced in neurologic physiotherapy. Neither had used the OAS and MAS scales routinely before conducting this study and no formal training in the use of these scales had been offered to them. However, they had the descriptions of the ratings criteria. This situation was thought to best reflect the clinical situation.

The level of agreement between the assessors was analyzed using the Cohen's kappa test. The Kappa values indicate the measure of agreement corrected for chance. The Kappa values were interpreted as suggested by Brennan and Silman (10) (Table 2). Percentage agreements and Kendall's τ coefficients were also calculated. The Chi Square test was used to compare the Kappa values. Statistical calculations were performed with the software package SPSS for windows, version 9.0.

RESULTS

Table 3 illustrates the agreement between the raters for the OAS and MAS. For the OAS, two raters agreed on 8 patients (n = 13) and for the MAS, two raters agreed on 7 patients (n = 13). For the OAS the Kappa value was 0.22 (SE 0.27) or poor (P = 0.43) and percentage agreement was 61.6% (Table 4). For the MAS, the Kappa value was 0.24 (SE 0.23) or poor (P = 0.24) and percentage agreement was 53.9% (Table 4). The Kendall's tau correlation value for the MAS (0.48, P = 0.02) was better than the OAS (0.22, P = 0.42).

Table 2. Interpretation of Kappa values as suggested by Brennan and Silman (1992)

Strength of agreement	Kappa Statistic
Poor	< 0.21
Fair	0.21-0.40
Moderate	0.41-0.60
Good	0.61-0.80
Very good	0.81-1.00

DISCUSSION

In this study the reliability between the two commonly used scales, OAS and MAS were compared. The results showed that inter-rater reliability of the original and the modified scales to test elbow flexor muscle spasticity was poor. The percentage agreement when using the OAS and the MAS was 61.6% and 53.9% and this might be characterized as "low" or "poor" (11). Hass *et al.* found fair reliability for the assessment of spasticity in the lower limbs (9). Our findings confirm the limited reliability of the Ashworth scales even when measuring spasticity in the upper limb. Therefore, it might be concluded that the limb, upper or lower, may not be a factor in agreement between assessors, despite the fact that Bohannon and Smith and Sloan *et al.* found the opposite (12). Experience and

Table 3. Agreement between two raters using the Ashworth scale for grading spasticity (n=13)

	Original Ashworth scale				
		Rater 2			
Rater 1	0	1	2	3	4
0					
1		**3**	2		
2		3	**5**		
3					
4					

	Modified Ashworth scale					
			Rater 2			
Rater 1	0	1	+1	2	3	4
0						
1		**2**	1			
1+		2	**4**	1		
2			2	**1**		
3						
4						

*Bold numbers show agreements between two raters.

Table 4. Comparison of results for original and modified Ashworth scales

Scales	Agreement (%)	Kendall's tau coefficient	*P*	Kappa	*P*	rating
OAS	61.6	0.22	0.42	0.22	0.43	Poor
MAS	53.9	0.48	0.02	0.24	0.24	Poor

Abbreviations: OAS, original Ashworth scale; MAS, modified Ashworth scale.

interactions and ample training in the use of the scale may have contributed to the high agreement between Bohannon and Smith (6). Bodin and Morris also concluded that the MAS was a reliable measure of wrist flexor spasticity when used by two trained assessors (13). In our study, no formal training was offered to the raters and they had no practice and interactions before beginning study and during the duration of study. This situation might have had an effect on low agreement between assessors. However, Pandyan *et al.* based on the raw data values provided by Bohannon and Smith (6) and Bodin and Morris (13) calculated the Cohen's κ values for these studies as 0.826 (95% CI 0.664-0.988) and 0.745 (95% CI 0.615-0.855) and concluded although the κ value was high, the standard error of the κ would suggest that the MAS was moderately reliable for classifying the resistance to passive movement at the elbow and the wrist flexors (7). Therefore, the MAS has no optimal reliability even at the elbow. In our study, MAS was slightly more reliable than was the OAS. This finding is not in agreement with Hass *et al.* who found that original scale was slightly more reliable than the modified scale (9).

Skold *et al.* concluded that the experienced person should perform the Ashworth gradings as in their study (14). In our study, both raters were women, one experienced but the other one less experienced. The less experienced therapist, at the end of the study reported that towards the end of the study she could differentiate the scores better because of better perception of spasticity and increasing experience. Our sample size was small and with increasing subjects, firm conclusion could be drawn. In this study, most agreement between raters occurred on score 2 of the original (38.5%) and score 1+ of the modified (30.8%) scales. This finding was not in agreement with findings of Hass *et al.* that discussed 1+ score had made low agreement between the raters (9). The most agreement with higher grades points to this fact that the reliability may be greater in a more 'spastic' patients group (9). In previous studies, repeated cycles of passive stretching prior to grading spasticity were used. In different studies, different cycles of stretching were used (6, 9, 11). Viscoelastic contributions to the resistance to passive movement are likely to decrease with repeated cycles of stretching (7, 8, 15). This repeated stretching cause variability in tone (16) that affect raters decision making about the true degree of spasticity. This factor may contribute to low agreement between the raters. It is therefore essential that repeated movements are kept to a minimum (7, 8). We believe for grading spasticity, clinicians rely on the first stretching as it is shown that even one stretch decreases spasticity (17).

Spasticity levels are influenced by a host of factors (18). Therefore, it is possible that some subjects' spasticity levels really changed between the raters' testing. Therefore, this factor probably contributed to the disagreement. Finally, we tested patients in supine position with the arms at the sides of the body. We were of the view that arm position will affect the degree of spasticity and patients' relaxation. Shoulder was not positioned in abduction. Shoulder pain is extremely common in patients with hemiplegia. Of these patients, 70 to 80% experience shoulder discomfort (19). As a result, we selected a comfortable position to control the factors influencing the amount of spasticity. In this position, no pain was reported at shoulder or elbow.

In conclusion, this study further showed that the reliability of the OAS and MAS was poor. The Modified Ashworth was slightly more reliable than the original scale; the difference was not statistically significant. Experience and training may improve the agreement between the raters, however, the poor reliability of the Ashworth scales questions the validity of the measurements.

Conflict of interests

The authors declare that they have no competing interests.

REFERENCES

1. Sehgal N, McGuire JR. Beyond Ashworth. Electrophysiologic quantification of spasticity. Phys Med Rehabil Clin N Am. 1998 Nov; 9(4):949-79, ix.
2. Haley SM, Inacio CA. Evaluation of spasticity and its effect on motor function. In: Glenn MB, Whyte J, editors. The practical management of spasticity in children and adults. Philadelphia: Lea & Febiger; 1990. p. 70-96.
3. Meythaler JM. Concept of spastic hypertonia. Phys Med Rehabil Clin N Am. 2001 Nov; 12(4):725-732, v.
4. Lance JW. Pathophysiology of spasticity and clinical experience with baclofen In: Lance JW, Feldman RG, Koella WP, editors. Spasticity: disordered motor control. Chicago: Year book; 1980. p. 185-204.
5. Ashworth B. Preliminary trial of carisoprodol in multiple sclerosis. Practitioner 1964; 192: 540-542.
6. Bohannon RW, Smith MB. Interrater reliability of a modified Ashworth scale of muscle spasticity. Phys Ther. 1987 Feb; 67(2):206-207.
7. Pandyan AD, Johnson GR, Price CI, Curless RH, Barnes MP, Rodgers H. A review of the properties and limitations of the Ashworth and modified Ashworth Scales as measures of spasticity. Clin Rehabil. 1999 Oct; 13(5):373-83.
8. Johnson GR. Measurement of spasticity In: Barnes MP, Johnson GR, editors. Upper motor neuron syndrome and spasticity. Cambridge: Cambridge University Press; 2001. p. 79-95.
9. Haas BM, Bergstrom E, Jamous A, Bennie A. The inter rater reliability of the original and of the modified Ashworth scale for the assessment of spasticity in patients with spinal cord injury. Spinal Cord. 1996 Sep; 34(9):560-564.
10. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ. 1992 Jun 6; 304(6840):1491-1494.
11. Allison SC, Abraham LD, Petersen CL. Reliability of the Modified Ashworth Scale in the assessment of plantarflexor muscle spasticity in patients with traumatic brain injury. Int J Rehabil Res. 1996 Mar; 19(1):67-78.
12. Sloan RL, Sinclair E, Thompson J, Taylor S, Pentland B. Inter-rater reliability of the modified Ashworth Scale for spasticity in hemiplegic patients. Int J Rehabil Res. 1992; 15(2):158-161.
13. Bodin PG, Morris ME. Inter rater reliability of the modified Ashworth scale for Wrist flexors spasticity following stroke. World federation of physiotherapy. 11th congress; 1991. p. 505-507.
14. Skold C, Harms-Ringdahl K, Hultling C, Levi R, Seiger A. Simultaneous Ashworth measurements and electromyographic recordings in tetraplegic patients. Arch Phys Med Rehabil. 1998 Aug; 79(8):959-65.
15. Pandyan AD, Price CI, Rodgers H, Barnes MP, Johnson GR. Biomechanical examination of a commonly used measure of spasticity. Clin Biomech (Bristol, Avon). 2001 Dec; 16(10):859-865.
16. Gregson JM, Leathley M, Moore AP, Sharma AK, Smith TL, Watkins CL. Reliability of the Tone Assessment Scale and the modified Ashworth scale as clinical tools for assessing poststroke spasticity. Arch Phys Med Rehabil. 1999 Sep; 80(9):1013-1016.
17. Vattanasilp W, Ada L, Crosbie J. Contribution of thixotropy, spasticity, and contracture to ankle stiffness after stroke. J Neurol Neurosurg Psychiatry. 2000 Jul; 69(1):34-39.
18. Katz RT. Management of spasticity. Am J Phys Med Rehabil. 1988 Jun; 67(3):108-116.
19. Martin ST, Kessler M. Neurologic intervention for physical therapist assistants. Philadelphia: W. B. Saunders company; 2000. p. 97.