Age and Ageing 2000; **29**: 223–228

Reliability of measurements of muscle tone and muscle power in stroke patients

JANINE M. GREGSON, MICHAEL J. LEATHLEY, A. PETER MOORE[1], TUDOR L. SMITH, ANIL K. SHARMA, CAROLINE L. WATKINS

Aintree Stroke Unit, Department of Medicine for the Elderly, University Hospital, Aintree, Longmoor Lane, Liverpool L9 7AL, UK
[1]The Walton Centre for Neurology and Neurosurgery, Liverpool, UK

Address correspondence to: J. M. Gregson. Fax: (+ 44) 151 529 3787. Email: cwatkins@aintreestar.u-net.com

Abstract

Objectives: to establish the reliability of the modified Ashworth scale for measuring muscle tone in a range of muscle groups (elbow, wrist, knee and ankle; flexors and extensors) and of the Medical Research Council scale for measuring muscle power in the same muscle groups and their direct antagonists.
Design: a cross-sectional study involving repeated measures by two raters. We estimated reliability using the κ statistic with quadratic weights (Kw).
Setting: an acute stroke ward, a stroke rehabilitation unit and a continuing care facility.
Subjects: people admitted to hospital with an acute stroke—35 patients, median age 73 (interquartile range 65–80), 20 men and 15 women.
Results: inter- and intra-rater agreement for the measurement of power was good to very good for all tested muscle groups (Kw = 0.84–0.96, Kw = 0.70–0.96). Inter- and intra-rater agreement for the measurement of tone in the elbow, wrist and knee flexors was good to very good (Kw = 0.73–0.96, Kw = 0.77–0.94). Inter- and intra-rater agreement for the measurement of tone in the ankle plantarflexors was moderate to good (Kw = 0.45–0.51, Kw = 0.59–0.64).
Conclusions: the Medical Research Council scale was reliable in the tested muscle groups. The modified Ashworth scale demonstrated reliability in all tested muscle groups except the ankle plantarflexors. If reliable measurement of tone at the ankle is required for a specific purpose (e.g. to measure the effect of therapeutic intervention), further work will be necessary.

Keywords: *Ashworth scale, measurement, Medical Research Council scale, muscle power, muscle tone, spasticity, stroke*

Introduction

Spasticity is defined as a velocity-dependent increase in muscle tone with hyperactive stretch reflexes [1]. It occurs in disorders of the central nervous system and can affect up to two-thirds of patients with stroke [2]. Impaired function may be a combination of spasticity and weakness in the same or antagonist muscle groups. In order to measure these factors clinical scales are used.

The accepted clinical tool used in measurement of muscle power is an ordinal scale, the Medical Research Council (MRC) scale [3]. This was initially developed to assess lower motor neurone lesions, specifically peripheral nerve injuries sustained in wartime. However, it has also been used in upper motor neurone lesions, including stroke [4], although its reliability as a clinical tool in this context remains unproven.

While there is no validated method of measuring muscle spasticity *per se*, the most widely accepted clinical scale to measure muscle tone is the modified Ashworth scale [5]. Originally, Ashworth described a five-point scale which assessed tone of 'the limbs' [6]. This was later modified to a six-point scale, and inter-rater reliability was established in the measurement of tone in the elbow flexors [5]. Subsequent studies have also shown good inter-rater agreement of the modified Ashworth scale in measuring tone in the elbow flexors [7] and in the wrist flexors [8]. However, the inclusion criteria used in previous studies make the results difficult to generalize: problems include exclusion of those with cognitive impairment [5] and exclusion of patients with 'known' spasticity [8].

The modified Ashworth scale has proved less reliable in the muscle groups of the leg [7, 9]. This affects studies which have attempted to assign a global spasticity score by combining these potentially unreliable upper and lower leg measurements [10].

Bohannon and Smith [5] established inter-rater reliability. However, intra-rater reliability is important in detecting changes in tone occurring over time. Few studies have examined this and some of those that have are of questionable methodology [10]. Similarly, applied statistical tests have not always been appropriate, in particular those which do not allow for chance agreement [11].

In determining agreement between and within raters on an ordinal scale, perfect agreement can be estimated using the κ statistic, which accounts for chance agreement [12]. κ is calculated by counting ratings of the same score as agreement and different scores as no agreement [12]. Weighted κ has been suggested as a method of taking account of the extent of disagreement: a difference of 1 point on the scale is no longer disagreement [12].

Aim

We designed this study to assess inter- and intra-rater reliability of the modified Ashworth scale when applied to the elbow, wrist and knee flexors and to the ankle plantarflexors, and of the MRC scale when applied to the above muscle groups and their direct antagonists. Inclusion criteria were to be well defined, use of the clinical tools standardized and procedures governed by written guidelines.

Methods

Subjects and sampling

We included patients admitted with an acute stroke, surviving to and still in hospital at the study start date. Subjects were on one of three wards: the acute stroke ward, the stroke rehabilitation unit and the continuing care facility. We obtained informed verbal consent. The only exclusion criterion was non-consent.

Procedures

We recorded basic demographic details comprising age and sex. Data pertaining to the stroke, comprising side of hemiplegia and time from onset of stroke to assessment, were also collected.

We assessed subjects at the specified joints (elbow, wrist, knee and ankle), first for tonal abnormality using the modified Ashworth scale, and then for power of the agonist and antagonist muscle groups using the MRC scale. We made each measurement three times after a rest period of 30 s and recorded the optimal score (i.e. lowest modified Ashworth and highest MRC score). We counterbalanced the order in which the patients and individual joints were assessed. We took a pragmatic approach to ensure order compliance in that for each patient we produced an assessment pack containing testing sheets with predefined joint assessment order. We assessed each subject at about the same time on two consecutive days using two raters (A, a research medical specialist registrar, and B, a research physiotherapist), with a 10 min rest between, and in a defined pattern (Table 1). Once an assessment had been performed by one of the raters, we removed the assessment sheet to ensure that the next rating would be performed without access to previous ratings.

Table 1. Order of patient assessment

Day 1		Day 2	
A	B	A	B
1	2	2	1
2	1	1	2
3	4	4	3
4	3	3	4
etc.	etc.	etc.	etc.

Where possible, patients were assessed seated in a Wolfson stroke chair with the relevant joint positioned by the assessor (as outlined in the Appendix). Where limitations applied, we made a record of the examining position used (e.g. lying supine), so that this position could be reproduced by the other rater when testing. The ambient room temperature was similar in all test areas ($\geq 20°C$) and we discreetly removed clothing to expose the joint and allow the muscle group the full range of movement.

A detailed description of the procedures for the measurement of tone and power can be found in the Appendix.

Data analysis

We calculated agreement between raters using the κ statistic. κ statistics with quadratic weights (Kw) [12] were used as we felt that a difference of 1 point on each of the scales would not be considered clinically significant. The results were interpreted as suggested by Brennan and Silman [13].

Results

We included 35 patients. The median age was 73 years (interquartile range 65–80). There were 20 men and 15 women. The median time since stroke was 40 days (interquartile range 19–78). Eighteen had right-sided hemiplegia and 17 had left-sided hemiplegia. The data demonstrating agreement between and within raters are summarized in Tables 2–5.

As can be seen from Table 3, only moderate inter-rater agreement for the measurement of tone in the ankle plantarflexors (Kw = 0.45–0.51) and only moderate to good intra-rater agreement for the measurement of tone in the ankle plantarflexors (Kw = 0.59–0.64) was demonstrated. As there is a relatively small range of movement at the ankle, it may be that the rater cannot distinguish between several grades. However, they may be able to determine whether there is tonal abnormality or not. Values for tone were therefore re-coded to allow comparison of normal tone with abnormal tone (i.e. 0 against 1–5 combined) as a dichotomous variable. We recalculated agreement using the unweighted κ statistic [12]. The inter-rater agreement remained poor ($\kappa = 0.17$–0.21). Similarly, intra-rater agreement was poor to moderate ($\kappa = 0.33$–0.46).

Table 2. Percentage intra-rater agreement for the modified Ashworth scale

Joint	Rater	Agreement	κ	Weighted κ^a	Interpretation
Elbow	A	72	0.53	0.77	Good
Elbow	B	62	0.39	0.83	Very good
Wrist	A	59	0.35	0.88	Very good
Wrist	B	71	0.48	0.80	Good
Knee	A	79	0.54	0.94	Very good
Knee	B	50	0.27	0.77	Good
Ankle	A	52	0.34	0.64	Good
Ankle	B	73	0.17	0.59	Moderate

[a]Quadratic weights applied.

Table 3. Inter-rater agreement for the modified Ashworth scale

Joint	Day	Agreement	κ	Weighted κ^a	Interpretation
Elbow	1	59	0.34	0.77	Good
	2	78	0.67	0.96	Very good
Wrist	1	66	0.43	0.84	Very good
	2	71	0.51	0.89	Very good
Knee	1	63	0.36	0.79	Good
	2	53	0.21	0.73	Good
Ankle	1	38	0.06	0.51	Moderate
	2	38	0.09	0.45	Moderate

[a]Quadratic weights applied.

Table 4. Intra-rater agreement for the Medical Research Council scale

Muscle group		Rater	Agreement	κ	Weighted κ^a	Interpretation
Elbow	Flexors	A	61	0.53	0.96	Very good
		B	44	0.32	0.81	Very good
	Extensors	A	55	0.47	0.93	Very good
		B	53	0.43	0.86	Very good
Wrist	Flexors	A	55	0.46	0.95	Very good
		B	56	0.47	0.96	Very good
	Extensors	A	61	0.47	0.92	Very good
		B	53	0.41	0.84	Very good
Knee	Flexors	A	45	0.36	0.96	Very good
		B	37	0.28	0.89	Very good
	Extensors	A	56	0.47	0.92	Very good
		B	53	0.45	0.90	Very good
Ankle	Dorsiflexors	A	42	0.26	0.90	Very good
		B	40	0.26	0.70	Good
	Plantarflexors	A	50	0.42	0.90	Very good
		B	57	0.45	0.91	Very good

[a]Quadratic weights applied.

Table 5. Percentage inter-rater agreement for the Medical Research Council scale

Muscle group		Day	Agreement	κ	Weighted κ[a]	Interpretation
Elbow	Flexors	1	26	0.14	0.85	Very good
		2	51	0.42	0.87	Very good
	Extensors	1	41	0.30	0.92	Very good
		2	47	0.36	0.94	Very good
Wrist	Flexors	1	55	0.45	0.94	Very good
		2	47	0.35	0.92	Very good
	Extensors	1	61	0.49	0.96	Very good
		2	53	0.39	0.89	Very good
Knee	Flexors	1	37	0.28	0.85	Very good
		2	30	0.20	0.89	Very good
	Extensors	1	50	0.40	0.95	Very good
		2	47	0.38	0.91	Very good
Ankle	Dorsiflexors	1	52	0.20	0.85	Very good
		2	52	0.38	0.89	Very good
	Plantarflexors	1	38	0.29	0.91	Very good
		2	39	0.29	0.84	Very good

[a]Quadratic weights applied.

Discussion

These results show that the inter-rater agreement for the measurement of power was very good for all muscle groups (Kw = 0.84–0.96). Similarly, intra-rater agreement for the measurement of power was good to very good for all muscle groups (Kw = 0.70–0.96). Although these measures have often been used clinically, the question of reliability has been assumed but not addressed. This issue is important when assessment of power is to be used as an outcome measure following intervention, both clinically and for research. As far as we are aware, this is the first study to demonstrate the reliability of the MRC scale. Our results suggest that it is valid to use this tool clinically or in research. However, it is important to remember that standardized guidelines were used when making these measurements. We believe that development and/or use of standardized guidelines are critical to the reliability of the scale.

Previous work on the modified Ashworth scale lacked well defined exclusion criteria, and excluded those patients with cognitive impairment [5] and 'known' spasticity [8]. Our study improved on previous research by including all patients who were able to give consent. Patients were not excluded if they were deemed to have normal muscle tone and/or power. Inter-rater agreement for the measurement of tone in the elbow, wrist and knee flexors was good to very good (Kw = 0.73–0.96). These results agree with those of Sloan *et al.* [7]. In addition to inter-rater reliability, we have extended the work of Bohannon and Smith [5] by measuring intra-rater agreement. Intra-rater agreement for the measurement of tone in the elbow, wrist and knee flexors was good to very good (Kw = 0.77–0.94).

As with previous work which has shown poor reliability in measuring tone in the leg [7, 9], the current study demonstrated only moderate inter-rater agreement for the measurement of tone in the ankle plantarflexors (Kw = 0.45–0.51). Intra-rater agreement for the measurement of tone in the ankle plantarflexors was only moderate to good (Kw = 0.59–0.64). This has important implications for clinical and research intervention studies. If a patient is measured, has an intervention and is then re-measured, it is possible that there will be a difference between pre- and post-intervention scores purely due to the unreliability of the measurement (as opposed to a real effect of the intervention).

This adds further weight to the criticism of studies which have attempted to assign a global spasticity score by combining upper and lower leg measurements [10]. The error of adding together scores from individual joints to give a global score is exacerbated by using items that have been shown to be unreliable.

Even at the level of reducing polychotomous variables to dichotomous variables, measurement of tone at the ankle remained unreliable—i.e. even when values for tone were re-coded to allow comparison of normal tone with abnormal tone at the ankle.

We concede that intra-rater agreement may be overestimated as, in theory, it is possible that a rater could remember their previous rating. However, the number of subjects assessed between the first and second examination of the same patient militates against raters remembering previous ratings. The results suggest that this was not the case as, should raters be able to remember their previous ratings, intra-rater agreement would have been improved across all measures of tone and power. Furthermore, it is likely that this improvement in intra-rater agreement would have been at the expense of inter-rater agreement.

In this study we have shown that it is possible to measure power and tone reliably at the elbow, wrist and knee. However, we used standardized guidelines. Other studies, which have used these measures in a non-standardized way, may have assumed reliability where none existed. Interventions may thus have been inadequately evaluated and wrongly thought to have or lack efficacy. This may have been because of artefact, observer bias or confounding. The results of future intervention studies will therefore be robust only if strict guidelines are applied. However, this cannot be said of measurement of tone at the ankle plantarflexors. The ability to measure tone accurately and reliably in the ankle is of clinical relevance as abnormal tone in this muscle group may have a marked impact on walking.

Conclusion

The measurement of tone using the modified Ashworth scale and of power using the MRC scale demonstrates inter- and intra-rater reliability at the elbow, wrist and knee. Measurement of power at the ankle was reliable, but that of tone was not. If reliable measurement of tone at the ankle was required for a specific purpose (e.g. to measure the effect of therapeutic intervention) further work would be necessary.

Key points

- The Medical Research Council scale is reliable for measuring muscle power.
- The modified Ashworth scale is reliable for measuring muscle tone in the elbow, wrist and knee flexors, but not in the ankle plantarflexors.
- If reliable measurement of tone at the ankle was required for a specific purpose (e.g. to measure the effect of therapeutic intervention) further work would be necessary.

References

1. Lance JW. Control of muscle tone, reflexes and movement: Robert Wartenberg Lecture. Neurology 1976; 30: 1303–13.

2. Wallesch C-W, Maes E, Lecomte P *et al.* Feasibility study on pharmacoeconomics of botulinum toxin A (Botox) in spasticity following stroke. Abstracts of the 3rd European Botulinum Toxin Symposium,1997; 1(4).

3. Medical Research Council/Guarantors of Brain. Aids to the Examination of the Peripheral Nervous System. London: Baillière Tindall, 1986.

4. Demeurisse G, Demol O, Robaye E. Motor evaluation in vascular hemiplegia. Eur Neurol 1980; 19: 382–9.

5. Bohannon RW, Smith MB. Inter-rater reliability of a modified Ashworth scale of muscle spasticity. Phys Ther 1987; 67: 206–7.

6. Ashworth B. Preliminary trial of carisoprodol on minimal to moderate spasticity in multiple sclerosis. Practitioner 1964; 192: 540–2.

7. Sloan RL, Sinclair E, Thompson S *et al.* Inter-rater reliability of the modified Ashworth scale for spasticity in hemiplegic patients. Int J Rehab Res 1992; 15: 158–61.

8. Bodin PG, Morris ME. Inter-rater reliability of the modified Ashworth scale for wrist flexor spasticity following stroke. Proceedings of the 11th Congress of the World Federation of Physical Therapy, London, 1991; (II)505–7.

9. Haas BM, Bergstrom E, Jamous A *et al.* 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; 34: 560–4.

10. Lee K-C, Carson C, Kinnin E *et al.* The Ashworth scale: a reliable and reproducible method of measuring spasticity. J Neuro Rehab 1989; 3: 205–9.

11. Haas BM. Measuring spasticity: a survey of current practice among health care professionals. Br J Ther Rehab 1994; 1: 90–5.

12. Cohen J. A co-efficient of agreement for nominal scales. Edu Psychol Measurement 1960; 20: 37–46.

13. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. Br Med J 1992; 304: 1491–4.

Received: 20 November 1998; accepted in revised form 24 June 1999

Appendix. Procedures for the measurement of tone and power

1. **Elbow flexors and extensors**
   a. Tone of the elbow flexors: the shoulder is in mid-rotation, the forearm is in mid-pronation and the hand is the functional position. The proximal upper limb rests horizontally on a variable-height table. Passive movement is achieved by the rater gripping the lateral aspect of the distal forearm just proximal to the wrist and applying a constant extensor rotational force about the elbow. The range of movement is from full flexion to full extension over about 1 s (the rater counts 'one thousand and one').
   b. Power of the elbow flexors: the patient is positioned as for 1a, with the elbow fully extended. The rater grips the lateral aspect of the distal limb just proximal to the wrist and, having first explained the required action, instructs the subject to 'pull' for about 1 s.
   c. Power of the elbow extensors: the patient is positioned as for 1a, with the elbow fully flexed. The rater grips the medial aspect of the distal limb just proximal to the wrist and, having first explained the required action, instructs the subject to 'push' for about 1 s.
2. **Wrist flexors and extensors**
   a. Tone of the wrist flexors: the patient is positioned as for 1a, with the distal limb held vertical. Passive movement is achieved by the rater grasping the hand just proximal to the metocarpophalangeal joints and applying a constant extensor rotational force about the wrist. Movement is from full flexion to full extension over 1 s.
   b. Power of the wrist flexors: the subject is positioned as for 1a, with the distal limb held vertical. The wrist is in full extension. The rater places a hand on the palmar aspect of the subject's hand and, having first explained the required action, instructs the subject to 'pull' for about 1 s.
   c. Power of the wrist extensors: the subject is positioned as for 1a, with the distal limb held vertical. The wrist is in full flexion. The rater places a hand on the dorsal aspect of the subject's hand just proximal to the metocarpophalangeal joints and, having first explained the required action, instructs the subject to 'push' for about 1 s.
3. **Knee and hip flexors and extensors**
   a. Tone of the knee flexors/hip extensors: the subject is in a seated position. The distal leg is suspended vertically with the foot off the floor. The trunk is stabilized by means of a padded strap drawn comfortably tight around the pelvis and the hip is stabilized by means of a padded strap drawn comfortably tight across the proximal legs, half-way between the knees and the hips. Passive movement is achieved by the rater grasping the posterior aspect of the distal leg just proximal to the ankle and applying a constant extensor rotational force about the knee. Movement is from 90° flexion to full extension over 1 s.
   b. Power of the knee flexors/hip extensors: the patient is positioned as for 3a and the rater places a hand on the posterior aspect of the distal leg just proximal to the ankle. Having first explained the required action, the rater instructs the subject to 'pull' for about 1 s.
   c. Power of the knee extensors/hip flexors: the patient is positioned as for 3a and the rater places a hand on the anterior aspect of the distal leg just proximal to the ankle. Having first explained the required action, the rater instructs the subject to 'push' for about 1 s.
4. **Ankle plantarflexors and dorsiflexors**
   a. Tone of the ankle plantarflexors: the patient is positioned as for 3a, with the rater stabilizing the leg by grasping it just proximal to the ankle. Passive movement is achieved by the rater grasping the foot just proximal to the metotarsophalangeal joints and applying a constant plantarflexor rotational force about the ankle. Movement is from full plantarflexion to full dorsiflexion over 1 s.
   b. Power of the ankle plantarflexors: the patient is positioned as for 3a, with the rater stabilizing the leg by grasping it just proximal to the ankle. The ankle is in full dorsiflexion. The rater places a hand against the sole of the foot just proximal to the metotarsophalangeal joints and, having explained the required action, instructs the subject to 'push' for about 1 s.
   c. Power of the ankle dorsiflexors: the patient is positioned as for 3a, with the rater stabilizing the leg by grasping it just proximal to the ankle. The ankle is in full plantarflexion. The rater places a hand against the dorsum of the foot just proximal to the metotarsophalangeal joints and, having explained the required action, instructs the subject to 'pull' for about 1 s.