

# Physical Therapy

Journal of the American Physical Therapy Association and



**de Fysiotherapeut**

Royal Dutch Society for Physical Therapy



## **Reliability of Measurements Obtained With the Modified Ashworth Scale in the Lower Extremities of People With Stroke**

Marjan Blackburn, Paulette van Vliet and Simon P Mockett

*PHYS THER.* 2002; 82:25-34.

The online version of this article, along with updated information and services, can be found online at: <http://ptjournal.apta.org/content/82/1/25>

### **Collections**

This article, along with others on similar topics, appears in the following collection(s):

[Hemiplegia/Paraplegia/Quadriplegia](#)

[Spasticity](#)

[Stroke \(Geriatrics\)](#)

[Stroke \(Neurology\)](#)

[Tests and Measurements](#)

### **e-Letters**

To submit an e-Letter on this article, click [here](#) or click on "Submit a response" in the right-hand menu under "Responses" in the online version of this article.

### **E-mail alerts**

Sign up [here](#) to receive free e-mail alerts

# Reliability of Measurements Obtained With the Modified Ashworth Scale in the Lower Extremities of People With Stroke

**Background and Purpose.** Abnormal muscle tone is a common motor disorder following stroke, which may require rehabilitation. The Modified Ashworth Scale is a 6-point rating scale that is used to measure muscle tone. The interrater and intrarater reliability of measurements obtained with the scale remain equivocal. The purpose of this study was to investigate the reliability of measurements obtained with the scale in the lower limb of patients with stroke. **Subjects.** Twenty patients were tested 2 weeks after their stroke, and 12 patients were tested 12 weeks after their stroke. **Methods.** Gastrocnemius, soleus, and quadriceps femoris muscles on the hemiplegic side were tested. **Results.** Interrater reliability for 2 raters was poor, with a Kendall tau-b correlation for the combined muscle group of .062 ( $P = .461$ ). For intrarater reliability, the Kendall tau-b correlation was .567 ( $P < .001$ ). The agreement within one rater occurred mostly on the grade of 0. **Discussion and Conclusion.** The Modified Ashworth Scale yielded reliable measurements in the lower limb for a single examiner, and agreement was best on the grade of 0. The reliability between examiners was not good, which may bring into question the validity of measurements obtained with the scale. [Blackburn M, van Vliet P, Mockett SP. Reliability of measurements obtained with the Modified Ashworth Scale in the lower extremities of people with stroke. *Phys Ther.* 2002;82:25–34.]

**Key Words:** Cerebrovascular disorders, Hemiplegia, Hypertonicity, Measurement, Muscle tone, Reliability.

*Marjan Blackburn, Paulette van Vliet, Simon P Mockett*

**T**his study investigated the reliability of measurements obtained by a scale for muscle tone—the Modified Ashworth Scale (MAS).<sup>1</sup> *Muscle tone* has been defined as “the resistance of muscle being passively lengthened or stretched.”<sup>2(p35)</sup> Tone results from 2 general physiological mechanisms: (1) the intrinsic properties of muscle, tendon, and connective tissue and (2) the active contraction of muscle.<sup>2</sup>

Increased resistance to passive stretch (hypertonus) is a common motor disorder.<sup>3,4</sup> Hypertonus has been assumed to be due primarily to an increased reflex response. However, Dietz et al<sup>5</sup> presented evidence that altered mechanical properties of muscle may contribute to hypertonia in patients. Further evidence is provided by a study of 24 patients where hypertonia was associated with contracture but not with reflex hyperexcitability.<sup>6</sup>

Despite recent evidence suggesting that weakness may contribute more to disability than increased reflex responses in patients with stroke,<sup>7–9</sup> management of increased muscle tone is still a major component of some rehabilitation protocols. Physical therapy texts<sup>3,4</sup> and Kidd et al<sup>10</sup> have advocated methods directed at normalizing these responses. It can be argued, therefore, that many therapists are interested in a tool that reliably measures reflex responses. Pederson<sup>11</sup> and Katz and Rymer<sup>12</sup> have reported the need for the development of scales that yield valid and reliable measurements to characterize increased reflex responses in order that the effectiveness of treatment techniques can be evaluated.

There are a number of different means of assessing muscle tone. Functional scales can contain a component for muscle tone (eg, the Fugl-Meyer Assessment<sup>13</sup>),

although the connection between function and muscle tone is not well established.<sup>14</sup> Additionally, electromyography, the Wartenburg/pendulum test,<sup>15</sup> and electrophysiological tests<sup>16</sup> are some measures used to characterize muscle tone. These measures, however, are not easily accessible and administrable by the clinicians, and other measures such as tendon jerks<sup>16</sup> and rating scales<sup>10</sup> (measures of behavior that involve an evaluation based on a checklist of criteria<sup>17</sup>) are more practical in clinical use.

Of these measures, we believe that rating scales are probably the most suitable for measuring muscle tone in the clinical setting. Scales with categories of “mild,” “moderate,” and “severe” increased reflex responses have been shown to result in unreliable measurements with large interrater disagreements.<sup>10</sup> The Ashworth scale is a 5-point rating scale for measuring muscle tone (whether of mechanical or neural origin), with ratings from 0 (“no increase in tone”) to 4 (“limb rigid in flexion or extension”).<sup>18</sup> It has been suggested that the Ashworth scale grade 0 could cover patients with “low tone” as well as “normal” muscle tone.<sup>19</sup> Bohannon and Smith,<sup>1</sup> in an early investigation of the reliability of measurements obtained using the Ashworth scale, found a clustering of scores at its lower end. In order to increase the sensitivity of the scale, they added an extra item to the lower end (grade 1+). Bohannon and Smith then retested the MAS for reliability on elbow flexor muscle tone in 30 patients with intracranial lesions and found that 2 raters agreed on 86.7% of the ratings. The Kendall tau correlation between the ratings was .847 ( $P < .001$ ).<sup>1</sup> The reliability of measurements obtained with the Ashworth scale and the MAS have been evaluated in a variety of patient groups.<sup>1,20–23</sup> In this article,

M Blackburn, PT, MCSP, is Lecturer, Division of Physiotherapy Education, School of Community Health Sciences, University of Nottingham, Clinical Sciences Building, Hucknall Road, Nottingham, United Kingdom NG5 1PB (Marjan.Blackburn@Nottingham.ac.uk). She was Research Physiotherapist, Division of Stroke Medicine, School of Community Health Sciences, University of Nottingham, at the time of the study.

P van Vliet, PhD, is Post Doctoral Research Physiotherapist, Centre for Vascular Research, School of Medical and Surgical Science, Division of Stroke Medicine, University of Nottingham.

SP Mockett, MPhil, MCSP, is Lecturer and Deputy Head, Division of Physiotherapy Education, School of Community Health Sciences, University of Nottingham.

All authors provided concept/research design, writing, data analysis, project management, fund procurement, institutional liaisons, and consultation (including review of manuscript before submission). Ms Blackburn provided data collection and clerical support. Ms Blackburn and Dr van Vliet provided subjects and facilities/equipment.

Ethical approval was provided by the Nottingham City Hospital.

This research was supported by the Hospital Saving Association.

The main findings were presented at the Physiotherapy Research Conference; Leeds, West Yorkshire, England; April 15, 1999.

*This article was submitted December 8, 2000, and was accepted July 11, 2001.*

we discuss those studies on patients with stroke, head injury, and multiple sclerosis.

Nuyens et al<sup>23</sup> tested interrater reliability of using the Ashworth scale in patients with multiple sclerosis. The patients were tested bilaterally, and Kendall tau coefficients of .55 or lower were found for interrater reliability of measurements obtained for the adductor and medial (internal) rotator muscles of the hip. Reliability of these measurements was reflected by Kendall tau coefficients of .70 and .77 for the soleus muscle, .67 and .72 for the gastrocnemius muscle, and .86 and .71 for the psoas major muscle. Reliability for the quadriceps femoris muscles was reflected by Kendall tau coefficients of .63 ( $P=.0001$ ) and .36 ( $P=.0297$ ).

In patients with stroke or head injury, measurements obtained with the MAS have been shown to have good interrater reliability for the elbow (Kendall tau=.847,  $P<.001$ )<sup>1</sup> and wrist (Kendall tau=.847,  $P<.001$ ),<sup>20</sup> but poorer reliability for the lower limb.<sup>23</sup> Sloan et al,<sup>22</sup> using the MAS, found a poor correlation (a Spearman coefficient of  $r=.37$ ,  $P<.01$ ) between 2 physicians acting as testers in their study in the lower limb for knee flexion without reinforcement (subjects were asked to clench their teeth when reinforcement was requested). Between the second physician and a physical therapist, coefficients of .32 with reinforcements and .26 without reinforcements were nonsignificant.

Pandyan et al<sup>24</sup> suggested that the lower reliability that they found for the lower limb, compared with the upper limb, can be attributed to difficulties the examiners may have had in perceiving “reflex-mediated stiffness” when moving the heavier shank and foot segments or that it may be explained by the differences in the mass of the limb segments. Pandyan et al<sup>24</sup> reported that testing procedures have not always been described in detail. Therefore, the reasons for reduced reliability across studies are difficult to determine. Better standardization of procedures might lead to improved reliability. Nuyens et al<sup>23</sup> also contended that reliability might be improved with additional standardization.

The purpose of our study was to examine the intrarater and interrater reliability of measurements obtained with the MAS in the lower limb of patients with stroke using a standardized procedure.

## Method

### Subjects

Subjects were recruited consecutively from a Health Care of the Elderly ward over a period of 3 months. A Health Care of the Elderly ward is a hospital ward for people over the age of 65 years with medical or rehabil-

itative needs. The first 36 people who met the following inclusion criteria were recruited into the study: having a diagnosis of stroke, residing within 25 km of the ward, being referred for physical therapy, and being willing and able to give informed consent.

The exclusion criteria were having musculoskeletal conditions that prevented the test procedure from being carried out and not having the ability to understand simple instructions. Other researchers have either included<sup>1</sup> or excluded<sup>20,23</sup> subjects with normal tone. In our study, there was no a priori exclusion of subjects with normal muscle tone, because the purpose was to assess reliability of measurements of muscle tone, which includes determining whether normal muscle tone is present.

The subjects had a mean age of 76.1 years ( $SD=7.89$ , range=62–90). Seventeen subjects were male, and 19 subjects were female. Eighteen subjects had right hemiplegia, and 18 subjects had left hemiplegia. As defined by the Bamford Scale,<sup>25</sup> 1 subject had a total anterior circulation infarct, 23 subjects had a partial anterior circulation infarct, 6 subjects had a lacunar circulation infarct, 3 subjects had a posterior occipital circulation infarct, and 3 subjects were unclassified.

### Procedure

Twenty subjects were tested 2 weeks following their stroke, and 20 subjects were tested after 12 weeks. Four of the subjects tested at 2 weeks were also tested at 12 weeks, due to limited availability of subjects. Thus, the total number of subjects in the trial was 36. Two testing times were included to increase the likelihood of recruiting subjects with both high and low muscle tone and not to assess reliability between 2 occasions 10 weeks apart. Subjects recruited at 12 weeks post-stroke were expected to have higher levels of muscle tone.

A researcher (MB) and an independent examiner examined each subject. Both testers were physical therapists, each with more than 10 years of experience in handling the limbs of people with stroke. As in previous studies,<sup>19,20,23</sup> no extensive training with regard to the procedure was done. Bohannon and Smith,<sup>1</sup> Lee et al,<sup>21</sup> and Sloan et al<sup>22</sup> used a period of training. Although this training, theoretically, might have improved the reliability of the measurements, we wanted the procedure used in this study to resemble how the scale would be used in clinical practice. After consultation with a group of local hospital and community-based physical therapists, we came to the conclusion that usual clinical practice could accommodate written guidelines, but was unlikely to involve extensive training.

**Table 1.**

Test Position for Patient and Examiner and Test Movement

Muscle	Patient	Examiner
Soleus	Side lying, with the hips in 45° of flexion and the knees in 45° of flexion. Head and trunk are in a straight line.	In front of the patient, examiner places one hand proximal to the ankle joint to stabilize the lower leg and one hand under the foot, with the thumb on the lateral aspect of the calcaneus, the fingers on the medial aspect of the calcaneus, and the palm on the plantar surface of the foot. The patient is asked to relax as much as possible. The ankle is moved from maximum plantar flexion to maximum dorsiflexion.
Gastrocnemius	Side lying, with the hips in 45° of and the knee in maximum extension. Head and trunk are in a straight line.	In front of the patient, examiner places one hand on the knee to stabilize the leg and one hand under the foot, as above. The patient is asked to relax as much as possible. The ankle is moved from maximum plantar flexion to maximum dorsiflexion.
Quadriceps femoris	Side lying, with the hips and knees in maximum extension. Head and trunk are in a straight line. A pillow can be used behind the hips, if necessary, to stabilize the patient.	Behind the patient, examiner places one hand just proximal to the knee, on the lateral surface of the thigh, to stabilize the femur and one hand just proximal to the ankle. The knee is moved from maximum extension to maximum flexion.

The level of muscle tone can be affected by the emotional status (fear, anxiety, or apprehension) and general unwellness of the subject, the environment, the temperature, and fatigue.<sup>26</sup> Therefore, for the interrater reliability component of the study, tests were performed 1 hour apart. A 1-hour interval was considered long enough to allow for the testing effect of the first test to disappear and short enough to prevent major changes in the subject or environment that may affect muscle tone. The order of testing by the 2 examiners was reversed for half of the subjects to control for order effects, such as familiarity with the examiner and the testing procedure. For the intrarater reliability component of the study, the researcher (MB) repeated the test 1 week later. Although muscle tone may change over this period, it may also change from day to day, and there is no evidence to suggest that the change will be consistently in one direction or the other. We believed that a week was sufficient time to prevent exact recall of the initial grade by the examiner.

The subjects tested at 2 weeks were tested on the ward. At 12 weeks, subjects were tested either on the ward or in the place to which they were discharged.

The muscles tested were the gastrocnemius, soleus, and the quadriceps femoris group on the hemiplegic side. They were chosen as lower-limb muscle groups important for rehabilitation and because they are said to be among the most common muscles affected by increased muscle tone.<sup>3,6</sup> Previous investigations of using the Ashworth scale to measure the calf muscles either have taken the approach of grouping the soleus and gastrocnemius muscles together as the plantar flexors<sup>27</sup> or have tested these muscles separately,<sup>23</sup> but the investigators did not explain their reasons for each approach. It is

common practice when testing the extensibility of these 2 muscles<sup>28</sup> to test them separately because of their different anatomical attachments. The soleus muscle crosses only one joint (talocrural), whereas the gastrocnemius muscle crosses 2 joints (talocrural and knee). The changes in passive muscle stiffness that are associated with disuse<sup>29</sup> and that can occur following a stroke are likely to affect each of these muscles differently because of their different attachments and depending on their pattern of use after stroke. Because the resistance felt when testing muscle tone derives partly from passive muscle properties, we reasoned that we should attempt to test the soleus and gastrocnemius muscles separately. At the time of conducting this study, we could find no evidence for or against the idea that the reflex response may be different for the 2 muscles. The positions described in Table 1 were based on testing positions for extensibility of the 2 muscles,<sup>28</sup> but evidence indicating that they can be differentiated during any form of testing is lacking.

In order to ensure that conditions were similar for testing by the 2 examiners and by the same examiner at different times, a standardized procedure was developed by our group of local hospital and community-based physical therapists. By using the standardized procedure and testing at the same time of the day, the test conditions at weeks 2 and 2 were as similar as possible. Insofar as it was possible for us to determine, there was no change in the subjects' general health, medication, or emotional status between test periods.

The standardized procedure was as follows. Each subject was put in a resting position for 5 minutes, with socks and shoes removed. The handling and positioning of the subject's limbs by the tester are described in Table 1.



**Table 2.**  
Modified Ashworth Scale

Grade	Description
0	No increase in muscle tone
1	Slight increase in muscle tone, manifested by a catch or by minimal resistance at the end of the range of motion (ROM) 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
2	More marked increase in muscle tone through most of the ROM, but affected part(s) easily moved
3	Considerable increase in muscle tone, passive movement difficult
4	Affected part(s) rigid in flexion or extension
9	Unable to test

Each test movement was performed over a duration of about 1 second (by counting “one thousand one”), as described by Bohannon and Smith.<sup>1</sup> The movement was repeated 3 times because once may not be sufficient for a rater to attribute a score.<sup>23</sup> After performing the 3 test movements, the tester graded the resistance felt, with a single score, according to the MAS<sup>1</sup> as described in Table 2.

A separate recording sheet was used for each subject so that the test results would not influence subsequent test results. The observers were unaware of each other's results.

### Data Analysis

Scores for each subject from both testers were assembled in tables of agreement for each muscle and for the combined muscle groups. The scores obtained during the 2- and 12-week tests were pooled for analysis. The numbers of assignments and percentage of agreement for each muscle were calculated.

Because the MAS is an ordinal level measure of resistance to passive movement,<sup>24</sup> reliability was tested statistically with the Kendall tau-b. This test allowed for comparison of our results with those of other studies.<sup>1,20,27</sup> The Kendall tau-b statistic is a nonprobabilistic estimate. Kappa coefficients were not calculated because the prevalence within each of the categories differed greatly and Altman<sup>30</sup> suggested that it is an inappropriate statistic to use under such circumstances.

Because of the high frequency of scores of 0 obtained, the data were skewed. Therefore, a further analysis was performed, excluding all of the test movements on which both examiners (for interrater reliability) or one examiner on both occasions (for intrarater reliability) obtained scores of 0, to more closely examine the other categories of the scale. Statistical calculations were per-

formed with the software package SPSS for Windows, release 6.0.\*

### Results

The MAS scores were distributed across the entire scale, with grades ranging from 0 to 4, but subjects were most often assigned a score of 0, and there were few subjects with a moderate to severe increased reflex response (grades 3 and 4) (Tabs. 3 and 4).

Most agreement occurred for scores of 0—60% within one rater and 40.8% between raters, for an overall percentage of agreement of 50.4% (121 out of 240 possible agreements) (Tabs. 5 and 6). Agreement of scores for grade 1 was 46.15% within one rater and 9.75% between raters. Agreement of scores for grade 1+ was 0% within one rater and 7.69% between raters. Agreement of scores for grade 2 was 12.5% within one rater and 0% between raters (Tabs. 3–6).

Percentages of agreement and Kendall tau-b values are presented in Table 7 (including all test movements) and Table 8 (excluding percentages of agreement for scores of 0) for each muscle group and for the combined muscle groups for one tester and between testers.

Table 7 shows that there was poor interrater reliability (percentage of agreement of 45%) for all muscle groups tested. However, intrarater reliability was significant to at least the  $P=.001$  level for each muscle group (ranging from tau-b=.443 for the gastrocnemius muscle to tau-b=.660 for the quadriceps femoris muscle) as well as for all muscles combined (tau-b=.566).

When percentages of agreement for scores of 0 were excluded (Tab. 8) for the interrater reliability test, 71 cases from 120 remained, and there was considerably less agreement between examiners. When percentages of agreement for scores of 0 were excluded (Tab. 8) for the intrarater reliability test, 48 cases from 120 remained, and there was again considerably less agreement between examiners than when percentages of agreement for scores of 0 were included.

### Discussion

The results indicate that measurements of muscle tone in the lower limbs of people with stroke with the MAS have what we would consider acceptable intrarater reliability (Kendall tau-b=.567, percentage of agreement=73.3%) but poor interrater reliability.

Because people with stroke might be expected to develop increased tone gradually and mechanical muscle properties will change over time, we had expected

\* SPSS Inc, 444 N Michigan Ave, Chicago, IL 60611.

**Table 3.**Agreement in Modified Ashworth Scale (MAS) Scores Between Tester 1 and Tester 2 for Each of the Three Muscle Groups<sup>a</sup>

Gastrocnemius Muscle		MAS Score—Tester 1						
MAS Score—Tester 2	0	1	1+	2	3	4	9	No. of Assignments
0	15	5	2		1			23
1	6	2		2		2		12
1+		1						1
2	1	2						3
3	1							1
4								0
9								0
Total	23	10	2	2	1	2	0	40

Soleus Muscle		MAS Score—Tester 1						
MAS Score—Tester 2	0	1	1+	2	3	4	9	No. of Assignments
0	18	7	2	3		1		31
1	5	2						7
1+								0
2	1							1
3	1							1
4								0
9								0
Total	25	9	2	3	0	1	0	40

Quadriceps Femoris Muscle		MAS Score—Tester 1						
MAS Score—Tester 3	0	1	1+	2	3	4	9	No. of Assignments
0	16	8	6	2	2			34
1			2		1			3
1+			1					1
2	1							1
3								0
4								0
9					1			1
Total	17	8	9	2	4	0	0	40

<sup>a</sup> On one occasion, a score of 9 (unable to test) was given, as the movement was unable to be rated due to pain.

the inclusion of subjects at 12 weeks would have allowed the whole range of points on the scale to be represented. However, this was not the case, as a substantial number of assignments were scored as 0.

Because most agreement was found for assigned scores of 0 ( $n=121$ ), it appears that the MAS, when used with this standardized procedure, can yield reliable measurements to establish whether normal or low muscle tone is present. The remaining 119 assigned grades represent a substantial number of opportunities of assessing agreement on the higher grades of the scale. Of these, only 21 were in agreement, 17 of which were at grades 1 and 1+. There were very few scores in grades 3 and 4, which has been found in other studies and therefore seems typical of people with stroke.<sup>1,20,27</sup>

The poor agreement on grades 1, 1+, and 2 in our study is comparable to the results of studies by Bohannon and Smith<sup>1</sup> and Bodin and Morris,<sup>20</sup> who also found noticeably poor agreement on the 1+ grade and slight disagreement on grade 2. In a review by Pandyan et al,<sup>24</sup> it was noted that much of the reduction of reliability of measurements obtained with the MAS appears to center on the disagreements at the lower end of the scale (ie, between the grades of 1 and 1+). Pandyan et al<sup>24</sup> suggested that the lower reliability observed when using the MAS, as compared with using the Ashworth scale, could be attributed to the extra level of classification, which has increased the probability of error. The descriptors for grades 1 and 1+ are different mainly in terms of the range of movement over which resistance is felt (see testing criteria in Tab. 2). Pandyan et al<sup>24</sup>

**Table 4.**  
Agreement in Modified Ashworth Scale (MAS) Scores By Tester 1 for Each of the Three Muscle Groups<sup>a</sup>

Gastrocnemius Muscle		MAS Score—Tester 1						
MAS Score—Tester 1	0	1	1+	2	3	4	9	No. of Assignments
0	16	5	1					22
1	3	5	2	1				11
1+	2	1						3
2		2		1				3
3					1			1
4								0
9								0
Total	21	13	3	2	1	0	0	40

Soleus Muscle		MAS Score—Tester 1						
MAS Score—Tester 1	0	1	1+	2	3	4	9	No. of Assignments
0	25	4	1					30
1	2	5		1				8
1+	0							0
2		1						1
3					1			1
4								0
9								0
Total	27	10	1	1	1	0	0	40

Quadriceps Femoris Muscle		MAS Score—Tester 1						
MAS Score—Tester 1	0	1	1+	2	3	4	9	No. of Assignments
0	31	1		1				33
1	2	2						4
1+							1	1
2		1	0					1
3								0
4								0
9							1	1
Total	33	4	0	1	0	0	2	40

<sup>a</sup> On 2 occasions, a score of 9 (unable to test) was given, as the movements were unable to be rated due to pain.

suggested that the resting limb posture before stretch should be standardized to ensure that the limb is moved throughout the same range during each testing occasion. Our procedure specified the start position; however, the disagreement on grades 1 and 1+ remained.

Both testers in our study were physical therapists who were experienced in handling the limbs of people with stroke and were familiar with, though did not regularly use, the MAS. In order to make testing conditions similar to the conditions of clinical practice, the therapists did not undergo extensive training in the use of the scale. Other researchers<sup>1,27</sup> ensured that their testers had extensive practice prior to their studies, and they consequently found good interrater reliability (Bohannon and Smith<sup>1</sup> (Kendall tau=.847 at  $P<.001$  in the study by Bohannon and Smith<sup>1</sup> and Kendall tau=.647 at  $P<.05$

in the study by Allison et al<sup>27</sup>). This finding indicates that training prior to testing can provide greater reliability and that a standardized procedure in itself is not enough.

The standardized procedure included written instructions regarding a 5-minute rest period prior to testing and specified positions for both subject and tester in addition to the guidelines used by Bohannon and Smith.<sup>1</sup> Although this procedure was sufficient to ensure good agreement within one examiner on scores obtained for the grade of 0, it was insufficient to ensure acceptable intrarater reliability on other grades or interrater reliability on any grade. Other reasons for poor interrater reliability could be the number of movements performed to establish the grade, the method of scoring, and extraneous factors.



**Table 5.**

Agreement in Modified Ashworth Scale (MAS) Scores Between Tester 1 and Tester 2 for the Combined Muscle Groups

MAS Score—Tester 1								No. of Assignments
MAS Score—Tester 2	0	1	1+	2	3	4	9	
0	49	20	10	5	3	1		88
1	11	4	2	2	1	2		22
1+		1	1					2
2	3	2						5
3	2							2
4								0
9					1			1
Total	65	27	13	7	5	3	0	120

**Table 6.**

Agreement in Modified Ashworth Scale (MAS) Scores by Tester 1 for the Combined Muscle Groups

MAS Score—Tester 1								No. of Assignments
MAS Score—Tester 1	0	1	1+	2	3	4	9	
0	72	10	2	1				85
1	7	12	2	2				23
1+	2	1		1			1	4
2		4		1				5
3					2			2
4								0
9							1	1
Total	81	27	4	4	2	0	2	120

**Table 7.**

Interrater and Intrarater Reliability for Modified Ashworth Scale Scores

Muscle	Percentage of Agreement	Kendall tau-b	Significance (1-Tailed)
Interrater reliability			
Quadriceps femoris	42.5	.289	.066
Gastrocnemius	42.5	.158	.216
Soleus	50.0	.197	.103
All	45.0	.062	.461 (approximate)
Intrarater reliability			
Quadriceps femoris	85.0	.660	.010
Gastrocnemius	57.5	.443	.001
Soleus	77.5	.585	.001
All	73.3	.567	.001 (approximate)

The examiners in our study commented that testing each muscle 3 times, as in the study by Bohannon and Smith,<sup>1</sup> was not always enough to establish the appropriate grade. However, more testing may alter the muscle tone. Pandyan et al<sup>24</sup> recommended that repeated movements should be kept to a minimum. Variations existed in this part of the procedure in previous studies. Lee et al<sup>21</sup> measured each muscle group 5 times, and Sloan et al<sup>22</sup> measured each movement 4 times.

Variations also existed in the scoring methods used in previous studies. Lee et al<sup>21</sup> recorded the lowest score. Sloan et al<sup>22</sup> gave individual scores for each movement. Lee et al<sup>21</sup> and Nuyens et al<sup>23</sup> summed individual muscle scores, but Pandyan et al<sup>24</sup> discouraged this practice, arguing that it will mask any unreliability arising with the use of individual muscle scores.

We recognize that muscle tone can fluctuate due to extraneous factors such as anxiety, depression, fatigue, the ambient temperature, the presence of concurrent urinary tract infections or

constipation, and the use of drugs.<sup>26</sup> These extraneous factors might have been another reason for our poor interrater reliability, but this explanation seems unlikely because the tests were done only 1 hour apart. The unstable nature of the reflex response is problematic when considering the reliability of measurements obtained with procedures designed to measure this response. In the design of a research study such as ours, several strategies could be incorporated to improve the

**Table 8.**  
Interrater and Intrarater Reliability for Modified Ashworth Scale Scores

Muscle	No. of Assignments	Percentage of Agreement	Kendall tau-b	Significance (1-Tailed)
Interrater reliability <sup>a</sup>				
Quadriceps femoris	24	4.1	.159	.461
Gastrocnemius	25	8.0	-.286	.010
Soleus	22	9.0	-.714	.000
All	71	7.0	-.326	.001
Intrarater reliability <sup>b</sup>				
Quadriceps femoris	9	33.3	.246	.390
Gastrocnemius	24	29.1	.042	.821
Soleus	15	40.0	.064	.808
All	48	33.3	.085	.533

<sup>a</sup> Excluding scores of 0 when both raters had scores of 0.  
<sup>b</sup> Excluding scores of 0 when one rater had scores of 0 on both occasions.

stability of the measure. One idea is to ensure that there is agreement between raters at the start of the study by training, but we have stated why we did not choose this strategy. The stability of the population studied could be improved, for example, by selecting a group of subjects with chronic stroke. This method of subject selection, however, would limit the relevance of the results to clinical practice, as the scale is also used for people with acute stroke. In addition, the number of raters could be increased. This method would have the advantage of increasing the number of comparisons between raters. The number of measurements made could also be increased by increasing the number of subjects. The number of subjects in our study was greater than in other studies,<sup>1,20,23,27</sup> and in all of these studies, there were only 2 raters. We would recommend increasing both number of subjects and the number of raters in future, similar studies.

Of the muscles tested, the reliability of scores obtained for the quadriceps femoris muscle was highest, both for interrater and intrarater reliability. No other study has tested this muscle group using the MAS in subjects with stroke, so it is not possible to compare this result with the results of other studies.

The 2 separate testing positions for the soleus and gastrocnemius muscles used in our study resulted in different reliability values. This finding suggests that altered mechanical properties of muscle should be given more consideration in testing muscle tone and that 1- and 2-joint muscles performing the same joint action should be tested separately, as has been done in the our study and in the study by Nuyens et al.<sup>23</sup>

Our study, in common with others, could not adequately test the reliability of scores obtained using the higher grades of the MAS due to the infrequency of the

occurrence of grades 3 and 4. Further investigation of this aspect is needed by deliberate selection of subjects with moderate to severe increased reflex responses in a future study.

### Conclusion

The study showed that measurements obtained with the MAS, when a standardized procedure is used in the lower limbs of people with stroke, have acceptable intrarater reliability on the grade of 0. However, a standardized procedure, in itself, was not enough to ensure adequate interrater reliability on the grade of 0. Further investigation is needed to determine reliability of measurements on the remainder of the scale. In the absence of interrater agreement, the validity of the measurements must also be questioned.

Comparison of the results of our study with the results of other studies suggests that reliability of measurements obtained using the MAS is greatly improved by extensive training of test users. Training may be required of all test users in a clinical setting to ensure adequate reliability between testers. Further research into the reliability of measurements obtained with the MAS for the lower limb in people with stroke is indicated to assess whether combining a standardized procedure with extensive training can improve interrater reliability for this patient group. In addition, studies with greater numbers of examiners are needed.

### References

- Bohannon RW, Smith MB. Interrater reliability of a modified Ashworth scale of muscle spasticity. *Phys Ther.* 1987;67:206–207.
- Gordon J. Disorders of motor control. In: Ada L, Canning C, eds. *Key Issues in Neurological Physiotherapy*. London, England: Heinemann; 1990:35.
- Bobath B. *Adult Hemiplegia: Evaluation and Treatment*. 3rd ed. London, England: Heinemann; 1990.
- Davies PM. *Steps to Follow*. Berlin, Germany: Springer-Verlag; 1985: 24–42.
- Dietz V, Quintern J, Berger W. Electrophysiological studies of gait in spasticity and rigidity: evidence that altered mechanical properties of muscle contribute to hypertonia. *Brain.* 1981;104:431–449.
- O'Dwyer NJ, Ada L, Neilson PD. Spasticity and muscle contracture following stroke. *Brain.* 1996;119:1737–1749.
- Gowland C, de Bruin H, Basmajian JV, et al. Agonist and antagonist activity during voluntary upper-limb movement in patients with stroke. *Phys Ther.* 1992;72:624–633.
- Fellows SJ, Kaus C, Thilmann AF. Voluntary movement at the elbow in spastic hemiparesis. *Ann Neurol.* 1994;36:397–407.
- Carr JH, Shepherd RB, Ada L. Spasticity: research findings and implications for intervention. *Physiotherapy.* 1995;81:421–429.

- 10 Kidd G, Lawes N, Musa I. *Understanding Neuromuscular Plasticity: A Basis for Clinical Rehabilitation*. London, England: Edward Arnold; 1992.
- 11 Pederson E. Clinical aspect of spasticity and measurement of spasticity. In: Thomas C, ed. *Spasticity: Mechanism, Measurement, and Management*. 1969;752:36–54. American lecture series.
- 12 Katz RT, Rymer WZ. Spastic hypertonia: mechanisms and measurement. *Arch Phys Med Rehabil*. 1989;70:144–155.
- 13 Brunnström S. Motor testing procedures in hemiplegia based on sequential recovery stages. *Phys Ther*. 1966;46:357–375.
- 14 Hinderer SR, Gupta S. Functional outcome measures to assess interventions for spasticity. *Arch Phys Med Rehabil*. 1996;77:1083–1089.
- 15 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 MS. *Clinical Rehabilitation*. 1992;6:41–48.
- 16 Delwaide PJ, Martinelli P, Crenna P. Clinical neurophysiological measurement of the spinal reflex activity. In: Feldman RG, Young RR, Koella WP, eds. *Spasticity: Disordered Motor Control*. Chicago, Ill: Year Book Medical Publishers; 1980:345–371.
- 17 Thomas JR, Nelson JK. *Research Methods in Physical Activity*. 3rd ed. Champaign, Ill: Human Kinetics Publishers Inc; 1996:240.
- 18 Ashworth B. Preliminary trial of carisprodol in multiple sclerosis. *Practitioner*. 1964;192:540–543.
- 19 Haas BM, Bergstrom E, Jamous A, Bennie A. The interrater reliability of the original and the modified Ashworth scale for the assessment of spasticity in patients with spinal cord injury. *Spinal Cord*. 1996;34:560–564.
- 20 Bodin PG, Morris ME. Interrater reliability of the modified Ashworth scale for wrist flexor spasticity following stroke. In: Proceedings of the 11th Congress of the World Confederation for Physical Therapy. 1991:505–507.
- 21 Lee K, Carson L, Kinnin E, Patterson V. The Ashworth scale: a reliable and reproducible method of measuring spasticity. *Journal Neurological Rehabilitation*. 1989;3:205–209.
- 22 Sloan RL, Sinclair E, Thompson J, et al. Interrater reliability of the modified Ashworth Scale for spasticity in hemiplegic patients. *Int J Rehabil Res*. 1992;15:158–161.
- 23 Nuyens G, De Weerd W, Ketalaer P, et al. Interrater reliability of the Ashworth scale in multiple sclerosis. *Clinical Rehabilitation*. 1994;8:286–292.
- 24 Pandyan AD, Johnson GR, Price CIM, et al. A review of the properties and limitations of the Ashworth and the modified Ashworth scales as measures of spasticity. *Clinical Rehabilitation*. 1999;13:373–383.
- 25 Bamford J, Sandercock P, Dennis M, et al. Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet*. 1991;337:1521–1526.
- 26 DeSouza LH, Musa IM. The measurement and assessment of spasticity. *Clinical Rehabilitation*. 1987;1:89–96.
- 27 Allison SC, Abraham LD, Peterson 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;19:67–78.
- 28 Cole JH, Furness AL, Twomey LT. *Muscles in Action: Approach to Manual Muscle Testing*. Melbourne, Victoria, Australia: Churchill Livingstone; 1988.
- 29 Goldspink G, Williams P. Muscle fibre and connective tissue changes associated with use and disuse. In: Ada L, Canning C, eds. *Key Issues in Neurological Physiotherapy*. London, England: Heinemann; 1990.
- 30 Altman DG. *Practical Statistics for Medical Research*. London, England: Chapman & Hall; 1991.

# Physical Therapy

Journal of the American Physical Therapy Association and



## Reliability of Measurements Obtained With the Modified Ashworth Scale in the Lower Extremities of People With Stroke

Marjan Blackburn, Paulette van Vliet and Simon P Mockett

*PHYS THER.* 2002; 82:25-34.

---

### References

This article cites 17 articles, 6 of which you can access for free at:

<http://ptjournal.apta.org/content/82/1/25#BIBL>

### Cited by

This article has been cited by 20 HighWire-hosted articles:

<http://ptjournal.apta.org/content/82/1/25#otherarticles>

### Subscription Information

<http://ptjournal.apta.org/subscriptions/>

### Permissions and Reprints

<http://ptjournal.apta.org/site/misc/terms.xhtml>

### Information for Authors

<http://ptjournal.apta.org/site/misc/ifora.xhtml>

---