ORIGINAL ARTICLE

Modified Ashworth scale reliability for measurement of lower extremity spasticity among patients with SCI

BC Craven[1,2] and AR Morris[2]

[1]Department of Medicine, University of Toronto, Toronto, Ontario, Canada and [2]Toronto Rehabilitation Institute (Lyndhurst Centre), Toronto, Ontario, Canada

Study design: Observational study.

Objectives: To report the intra-rater (one rater), inter-rater (two raters) and inter-session (one subject, sessions 1–5) reliability of lower extremity modified Ashworth scale (MAS) scores among patients with chronic spinal cord injury (SCI).

Setting: Tertiary Academic Rehab Centre in Toronto, Canada.

Methods: MAS scores of 20 subjects with chronic SCI (C5–T10 AIS A–D > 12 months) were recorded for the hip abductors and adductors, knee flexors and extensors, and ankle plantar and dorsiflexors. MAS scores were assessed by two blinded raters (A and B) at the same time of day, weekly for 5 weeks using standardized test positions, a one-cycle per second metronome, with ratings recorded on the second cycle. MAS score reproducibility [intra-rater, inter-rater] were calculated using Cohen's Kappa. Intraclass correlation coefficients (ICCs) were calculated to determine inter-session (trials 1–5) reliability; Kappa values $\geqslant 0.81$ and ICC values $\geqslant 0.75$ were desired.

Results: Intra-rater reliability was fair to almost perfect ($0.2 < \kappa < 1.0$) and differed between raters. Inter-rater reliability was poor-to-moderate ($\kappa < 0.6$) for all muscle groups. Inter-session reliability for a single rater was fair-to-good ($0.4 < \text{ICC} < 0.75$) for all muscle groups.

Conclusions: MAS was not reliable as an intra-rater tool for all raters, and showed poor inter-rater and modest inter-session reliability. MAS has inadequate reliability for determining lower extremity spasticity between raters (inter-rater) or over time (inter-session). It is recommended that the rehabilitation science community seek alternative measures for quantifying spasticity.

Spinal Cord (2010) 48, 207–213; doi:10.1038/sc.2009.107; published online 29 September 2009

Keywords: reliability; spasticity; spinal cord injury; modified Ashworth scale; lower extremity

Introduction

Spasticity

The term 'spasticity' is used to indicate exaggerated involuntary muscular activity. Lance's commonly cited definition (1980) refers to spasticity as a velocity-dependent increase in muscle tone characterized by (1) exaggerated tendon jerks (hyper-reflexia); (2) increased muscle response to applied stretch, positively correlated with the lengthening rate (velocity-dependent hypertonia); it is usually attributed to the hyperexcitability of the myotatic 'stretch' reflex arc(s).[1] Decq (2003) suggested a modified definition of spasticity as a symptom of the upper motor neuron syndrome characterized by an exaggeration of the stretch reflex secondary to hyperexcitability of spinal reflexes.[2,3] In a recent review, the authors acknowledged 'the common clinical conundrum of

the ability of clinicians to easily recognize spasticity although its quantification remains elusive'.[4]

Spasticity is often categorized symptomatically as either tonic or phasic. Decq (2003)[2] suggests that tonic spasticity is the increased muscle tone resulting from the exaggerated tonic component of the stretch reflex. Decq has also defined phasic spasticity (normally characterized as increased clonus and tendon hyper-reflexia) as the exaggerated phasic component of the stretch reflex.[2]

Both positive and negative symptoms of spasticity are evident on clinical examination. Positive symptoms include involuntary movements, stiff muscles and joints, exaggerated cutaneous reflexes and contracture. Negative symptoms include paresis, loss of fine dexterity and early fatigability of voluntary movement.[5] The distinction between positive and negative symptoms is important from a functional and outcome perspective. Functionally, tonic spasticity can be painful, often interfering with activities of daily living, self-care and sleep. Phasic spasticity in patients with spinal cord injury (SCI) can lead to other secondary health

complications including: falls from the wheelchair or pressure sores. Severe spasticity can interfere with health[6,7] thereby preventing a person from returning to independent living and gainful employment.[8] Priebe et al.[9] and Hsieh et al.[10] have indicated that a single spasticity outcome measure may misrepresent the severity and influence of spasticity on patients with SCI. There is currently no clinical measure that captures or quantifies the phasic and tonic aspects of spasticity.

Spasticity treatment

Amongst patients with chronic SCI (>1 year after injury), 65–78% report symptoms of spasticity; and 37% whom require treatment.[11,12] One of the greatest challenges facing the SCI clinician and rehabilitation team is evaluating the effectiveness of drug and/or rehabilitation interventions intended to ameliorate spasticity. The mainstays of treatment for spasticity include rehabilitation therapies such as hot or cold application, stretching, positioning and splinting to prevent contracture, in addition to oral and or injectable pharmacologic treatments and neurosurgical procedures. A recent evidence-based review of current interventions indicated that several pharmacalogical agents (baclofen, tizanidine, clonidine, cyproheptadine, gabapentine and L-threonin) and transcutaneous electrical nerve stimulation had good to excellent levels of evidence of their efficacy for reducing spasticity after SCI.[13,14] A prior systematic review of the comparative and efficacy and safety of skeletal muscle relaxants for treatment of spasticity among patients with diverse neurologic impairments revealed equivalent efficacy of baclofen and tizanidine with a higher frequency of dry mouth with tizanidine and more weakness with baclofen.[15] A recent systematic review of stretching efficacy indicated inconclusive evidence regarding its efficacy for spasticity reduction.[16] Quantitative assessment of spasticity (both relative and absolute) is vital to the detection of spasticity among individuals with SCI and determination of treatment efficacy in a clinical trial setting or effectiveness in the clinic setting. A reliable tool to quantify lower extremity spasticity is needed.

Spasticity measures

Clinical, biomechanical and electrophysiological methods are available to measure lower extremity spasticity. The reader is referred to two recent reviews discussing the merits of these various techniques.[4,17]

The modified Ashworth scale (MAS) is a six-category ordinal scale used to assess the resistance encountered during passive muscle stretching that does not require instrumentation and is quick to perform.[18] The MAS is the current standard for clinical assessment of lower extremity spasticity, and the most commonly used tool to evaluate the efficacy of pharmacologic and rehabilitation interventions for treatment of spasticity among patients with SCI. The MAS is the gold standard against which new assessment tools are evaluated.

Haas et al.[19] earlier reported the reliability of the MAS in the SCI population. A doctor and physiotherapist rated hip adductors, hip extensors, hip flexors and ankle plantar flexors using the MAS. Assessment of reliability using Kappa showed a mean of $\kappa = 0.37$ (range, $0.21 < \kappa < 0.61$). The authors concluded that the MAS was of limited use in assessing lower extremity spasticity in patients with SCI. To our knowledge, this is the only prior evaluation of MAS lower extremity reliability reported in subjects with SCI. The reliability of a measurement is broadly defined as freedom from random error. Reliability is essential to interpret a measure's results and to answer questions, 'such as whether two numeric results really (probably) differ and whether one should have high, moderate or low confidence in inferences from the measures: unreliability constrains validity'.[20] Our objective was to determine the reliability (intra-rater, inter-rater and inter-session) of the MAS as a measure of lower extremity spasticity among a representative sample of the chronic SCI population whilst trying to eliminate potential confounders.

Materials and methods

Design

A convenience sample of 20 subjects were recruited through a local poster campaign and the assistance of outpatient program staff. Eligible subjects were male or female, 18–80 years of age, with chronic SCI (C4–T10, AIS A–D, >12 months after injury), with lower extremity spasticity, intact skin, normal urine microscopy at baseline and on stable doses of oral anti-spasticity medications. Subjects were excluded if they had (1) a lower extremity fracture within 6 months of enrolment; (2) a non-union lower extremity fracture; (3) >5 symptomatic urinary tract infections (UTIs) within the last year; (4) a symptomatic UTI within 2 weeks of enrolment; (6) syringomyelia; (7) severe lower extremity contracture; (8) heterotopic ossification of the hip or knee regions; (9) bilateral total knee or hip arthroplasty; (10) ingrown toenails; (11) severe lower extremity neuropathic pain; (12) >30° of combined hip and knee flexion contracture; (13) ingrown toenails; and (14) Botox injections in the 6 months before enrolment or phenol in 12 months before enrolment. This study was approved by the research ethics board of Toronto Rehab REB # 03–083), and we certify that all applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed.

Consenting subjects were required to attend a screening visit to discern their eligibility for inclusion. The subject's medical history, current medications, caffeine, alcohol and nicotine intake, neurologic level, ASIA impairment scale and lower extremity range-of-motion were recorded during the screening visit to ensure they met inclusion criteria. Subjects were required to maintain a stable intake of caffeine, nicotine and alcohol during the 5-week study period. Subjects were instructed to routinely take their spasticity medication at the same time of day during the study period. The subjects' self-reported caffeine, nicotine, alcohol and prescription drug intake were recorded at each visit. A review of systems was done at each visit; however, subjects' were not

queried regarding their anxiety or life circumstances. Subjects who developed an acute illness or a symptomatic UTI during the study were removed ($n = 1$). Subjects were blinded to the study objectives and MAS results during the data collection.

Raters

The MAS test was scored by four trained and experienced raters (three physical therapists and an MD). All of the study raters participated in the SCI-301, SCI-302 and SCI-300 fampridine clinical trials (www.clinicaltrials.gov) and the associated baseline tests of intra-rater reliability. Prior to conduct of this study, each rater participated in three group-training sessions and performed MAS on three pilot subjects to familiarize them with the study testing protocol. Each subject had two consistent raters at each of the five sessions—an MD and assigned therapist. Rater A assessed the MAS at each of the five sessions, 1 week apart; Rater B assessed the MAS at session 1 and 5 only. Ratings were performed at the same time of day. Raters A and B were blinded to all prior MAS session results throughout the study.

MAS assessment

All subjects were transferred onto the plinth and asked to lie unaided for 3 min before assessment to avoid measurement of the exacerbation of spasticity evoked by the transfer/the transfer mechanism. The MAS assessments were performed at the beginning of the testing session by rater A and repeated by rater B about an hour later after biomechanical measures (ramp-and-hold and knee pendulum tests, not shown), and repeat transfer onto the plinth and repetition of the pre-assessment procedures. The day of the week and time of day for assessments were maintained within a single subject during the study period to minimize time of day effects (that is subject X was assessed every Monday at 2:00 pm and subject Y every Thursday at 4:00 pm for five consecutive weeks). The MAS was used to assess both the right and left: hip abductors and adductors, knee flexors and extensors and ankle plantarflexors and dorsiflexors. Subjects were transferred onto a height-adjustable plinth and their shoes were removed. After a rest period of at least 3 min, the MAS scores were determined using standardized test positions, right–left test order and a one-cycle per second metronome. Subjects were supine on a plinth with their lower extremities fully supported during testing of hip adductors, hip abductors, ankle plantarflexors and dorsiflexors. During testing of the knee flexors, the leg was positioned at 90 degrees of hip flexion and the knee was allowed to rest in full flexion (Figure 1). During testing of the knee extensors, the lower extremity distal to the knee was suspended over the plinth edge. The distal limb was moved through the range of available knee extension, whereas the proximal thigh was secured in the start position. MAS scores were recorded on the second cycle for each muscle group using a once-cycle per second metronome to select and maintain test velocity.

Figure 1 Example of modified Ashworth scale testing on right lower extremity.

Table 1 Kappa values and strength of agreement

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

Source: Landis and Koch, 1977.

Analysis

Descriptive statistics were used to characterize the subject's demographic and impairment characteristics. To analyze the ordinal MAS scores, Cohen's Kappa was used to determine test–retest intra-rater and inter-rater reliability. Intraclass correlation coefficients (ICCs) were used to calculate inter-session reliability of the MAS. The MAS scale was converted to discrete categorical scores (0, 1, 2, 3, 4 and 5). Linearly weighted Cohen's Kappa's were calculated in MATLAB (Version 2006A, Mathworks Inc., Natuck, MA, USA) and ICCs (two-way mixed, absolute agreement, 95% confidence interval) in SPSS (v12; SPSS, Chicago, IL, USA). Intra-rater reliability was calculated for each rater for a single session (trials a/b for sessions 1 and 5 separately) using Cohen's Kappa. Similarly, inter-rater reliability (session 1 trial a, session 5 trial a) between raters A and B using Cohen's Kappa. Inter-session reliability (1st trial, sessions 1–5, rater A only) were calculated using the ICC. The strength of agreement assigned is based Kappa interpretation guidelines from Landis and Koch (1977) (Table 1) and the strength of agreement guidelines for ICC interpretation from Fleiss (1986) (Table 2).[21,22] Kappa values $\geqslant 0.81$ and ICC values $\geqslant 0.75$ were desired.

Table 2 ICC values and strength of agreement

Value of ICC	Strength of agreement
<0.4	Poor
0.4–0.75	Fair to good
>0.75	Excellent

Abbreviation: ICC, intraclass correlation coefficient.
Source: Fleiss, 1986.

Table 3 SCI cohort demographic and impairment characteristics

Variable (units) statistic	Cohort (n = 20)
Gender n (%) male	17 (85)
Age (years) mean (min–max) ± s.d.	38.9 (16–67) ± 13.6
Weight (kg) mean (min–max) ± s.d.	77.7 (57–114) ± 13.2
Body mass index (kg m^{-2}) mean (min–max) ± s.d.	24.8 (19.7–31.5) ± 2.8
Time after SCI (years) mean (min–max) ± s.d.	8.89 (1–24) ± 8.0
Spinal cord impairment n (%)	
Para AIS A	3 (15%)
Para AIS B–D	5 (25%)
Tetra AIS A	3 (15%)
Tetra AIS B–D	9 (45%)
Cause of injury n (%)	
Trauma	19 (95%)
Non-trauma	1 (5%)
Current health (0–10) mean (min–max) ± s.d.	8.4 (6–10) ± 1.33
Lower extremity contracture n (%) >10°	
Hip	5 (25%)
Knee	4 (20%)
Ankle	3 (15%)
Spasticity medication n (%)	
Baclofen	8 (47%)
Dantrium	1 (6%)
Tizanidine	1 (6%)
Valium	2 (12%)
Neurontin	1 (6%)
Lifestyle behaviours	
Caffeine	12 (71%)
Nicotine	4 (24%)
Alcohol	8 (74%)

Abbreviation: SCI, spinal cord injury.

Table 4 SCI Cohort frequency of MAS scores for rater A and B session 1a

Muscle group	Rater	MAS category frequency distribution session 1a (right leg)					
		0	1	2	3	4	5
Hamstrings	A	3	6	2	6	3	0
	B	4	2	3	5	4	0
Hip abductors	A	16	0	0	4	0	0
	B	12	3	2	1	0	0
Hip adductor	A	2	7	3	5	3	0
	B	3	5	2	6	2	0
Dorsiflexors	A	14	3	0	1	0	0
	B	15	0	0	0	1	0
Plantarflexors	A	0	8	2	4	2	4
	B	2	4	0	3	5	3
Quadriceps	A	1	9	3	2	2	3
	B	3	5	4	2	2	2

Abbreviations: MAS, modified Ashworth scale; SCI, spinal cord injury.
Legend for Ashworth score: 0 = No increase in muscle tone (MAS 0); 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 affected part(s) moved in flexion or extension (MAS 1); 2 = 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 (MAS +1); 3 = More marked increase in muscle tone through most of the range of movement, but affected parts easily moved (MAS 2); 4 = Considerable increase in muscle tone, passive movement difficult (MAS 3); 5 = Affected part(s) rigid in flexion or extension (MAS 4).

In total, four subjects were smokers, eight subjects consumed alcohol and 12 consumed caffeine on a regular basis. One subject reported excess alcohol intake at baseline (24 per day) and abstention during the study, which likely influenced their spasticity assessment. Of the three lifestyle behaviours, caffeine intake was the most variable across sessions (mean 1 cup per day, range 1–3 cups per day), with all 10 caffeine consumers reporting an increase or decrease of one serving per day at one of the five visits. Baclofen and Valium were the most frequently prescribed spasticity medications; with stable doses used throughout. Plantar flexion contractures were the most common followed by knee flexion contractures.

All subjects completed the 5-week intervention; some sessions and data collection points were missed because of transportation/timing issues, subject illness or discomfort on the test date. To maximize statistical power in the absence of one or more data points, reliability statistics of different sample sizes (n) were calculated and compared from the 20 included subjects. Reliability statistics were calculated on n = 14–17 subjects for intra-rater reliability, n = 16–17 for inter-rater reliability and n = 17 for inter-session reliability. The calculated statistics (linearly weighted Cohen's Kappa and ICCs) are shown for intra-rater reliability (Table 5), inter-rater reliability (Table 6) and inter-session reliability (Table 7).

The intra-rater reliability for lower extremity MAS was substantial to high (0.6 < κ < 1.0) for Rater A for three of six muscle groups. Those muscle groups with poor reliability were the knee extensor, ankle plantarflexor and dorsiflexor muscle groups. For rater B reliability varied between

Results

A total of 28 consenting subjects attended the screening visit; 8 subjects were excluded because of knee contracture (n = 1), hip contracture (n = 1), UTI criteria (n = 4) and pain (n = 4). A total of 20 subjects were eligible for inclusion. The subject's baseline demographic and impairment characteristics, spasticity medication profiles and intake of caffeine, nicotine and alcohol are presented (Table 3). Table 3 illustrates that we were able to recruit a sample with diverse impairments and spasticity; whereas Table 4 is a snapshot of frequency distribution of MAS scores during the study indicating the spectrum of lower extremity MAS scores.

Table 5 Intra-rater reliability of lower extremity MAS scores

| Muscle group | Rater A | | | | | | Rater B | | | | | |
| | Intra-session: trial 1a/1b | | | Intra-session: trial 5a/5b | | | Intra-session: trial 1a/1b | | | Intra-session: trial 5a/5b | | |
	κ	P	n	κ	P	n	κ	P	n	κ	P	n
Ankle dorsiflexors [R]	0.622	0.009	17	0.632	0.046	14	0.363	0.199	17	1.000	0.000	16
Ankle dorsiflexors [L]	0.851	0.000	17	0.759	0.009	14	0.469	0.065	17	1.000	0.000	16
Ankle plantarflexors [R]	0.913	0.000	17	0.652	0.000	14	0.788	0.000	17	0.780	0.000	17
Ankle plantarflexors [L]	0.612	0.000	17	0.745	0.000	14	0.496	0.000	17	0.632	0.000	17
Knee quadriceps [R]	0.190	0.196	17	0.527	0.000	14	0.288	0.036	17	0.419	0.002	17
Knee quadriceps [L]	0.478	0.000	17	0.700	0.000	14	0.270	0.154	17	0.480	0.000	17
Knee hamstrings [R]	0.773	0.000	17	0.468	0.001	14	0.406	0.002	17	0.853	0.000	17
Knee hamstrings [L]	0.555	0.000	17	0.816	0.000	14	0.314	0.016	17	0.608	0.000	17
Hip abductors [R]	0.843	0.000	17	0.571	0.003	14	0.320	0.116	17	0.575	0.006	17
Hip abductors [L]	0.843	0.000	17	0.317	0.100	14	0.404	0.029	17	0.764	0.001	17
Hip adductors [R]	0.485	0.001	17	0.451	0.002	14	0.615	0.000	17	0.692	0.000	17
Hip adductors [L]	0.575	0.000	17	0.404	0.023	14	0.545	0.000	17	0.692	0.000	17

Abbreviation: MAS, modified Ashworth scale.

Table 6 Inter-rater reliability of lower extremity MAS scores

| Muscle group | Rater A and B | | | | | |
| | Trial 1a | | | Trial 5a | | |
	κ	P	n	κ	P	n
Ankle dorsiflexors [R]	0.320	0.072	17	0.458	0.022	16
Ankle dorsiflexors [L]	0.419	0.021	17	0.247	0.229	16
Ankle plantarflexors [R]	0.370	0.007	17	0.177	0.140	17
Ankle plantarflexors [L]	0.416	0.001	17	0.221	0.051	17
Knee quadriceps [R]	0.089	0.497	17	0.352	0.005	17
Knee quadriceps [L]	-0.013	0.936	17	0.146	0.332	17
Knee hamstrings [R]	0.267	0.049	17	0.335	0.007	17
Knee hamstrings [L]	0.335	0.013	17	0.204	0.165	17
Hip abductors [R]	0.261	0.216	17	0.255	0.171	17
Hip abductors [L]	0.346	0.100	17	0.242	0.268	17
Hip adductors [R]	0.220	0.118	17	0.401	0.003	17
Hip adductors [L]	0.220	0.118	17	0.401	0.003	17

Abbreviation: MAS, modified Ashworth scale.

Table 7 Inter-session reliability (ICC) of lower extremity MAS scores

| Muscle group | Rater A | | | Rater B | | |
| | Inter-session: trial 1a/5a | | | Inter-session: trial 1a/5a | | |
	κ	P	n	κ	P	n
Ankle dorsiflexors [R]	0.486	0.021	18	0.233	0.323	14
Ankle dorsiflexors [L]	0.400	0.023	18	0.222	0.262	14
Ankle plantarflexors [R]	0.185	0.143	18	0.246	0.059	15
Ankle plantarflexors [L]	0.323	0.004	18	0.005	0.972	15
Knee quadriceps [R]	0.169	0.153	18	0.088	0.537	15
Knee quadriceps [L]	0.316	0.011	18	−0.071	0.652	15
Knee hamstrings [R]	0.038	0.769	18	0.250	0.061	15
Knee hamstrings [L]	0.300	0.022	18	0.073	0.641	15
Hip abductors [R]	0.408	0.047	18	0.612	0.002	15
Hip abductors [L]	0.348	0.084	18	0.516	0.051	15
Hip adductors [R]	0.274	0.033	18	0.483	0.001	15
Hip adductors [L]	0.175	0.273	18	0.483	0.001	15

Abbreviations: ICC, intraclass correlation coefficient; MAS, modified Ashworth scale.

sessions—in session 1 was poor-to-fair ($\kappa < 0.4$), except for ankle plantarflexors and contrarily were substantial-to-high ($0.6 < \kappa < 1.0$) for session 5 for all except knee quadriceps and hamstrings.

Inter-rater reliability was poor-to-moderate ($\kappa < 0.6$) for all muscle groups except the hip adductors. The agreement between the two raters was inconsistent across muscle groups and sessions (1a and 5a) and much lower than the desired Kappa ($\kappa > 0.81$).

Inter-session reliability was fair-to-good ($0.4 < ICC < 0.75$) for all muscle groups. The ICC for the knee flexors, knee extensors and adductors were lower ($0.4 < ICC < 0.4$) than the ankle plantarflexors ($ICC = 0.75$). These results indicate that lower extremity MAS score reliability is much lower than the clinically desired value ($ICC > 0.75$) for all but one muscle group.

Discussion

When one tries to assess the reliability of a clinical measure, consideration must be given to multiple sources of variability: subject, rater and that inherent in the tool. In this study, we tried to eliminate subject variability by having them maintain stable caffeine, nicotine and alcohol consumption and timing of prescription medication during the study period. We also made every attempt to collect data from the same subjects at the same time of day each week. We were unable to control for environmental impacts on spasticity specifically air temperature during the winter months in Canada. We tried to ameliorate the effects of subject's transfer, arousal and anxiety levels by having them rest for 3 min before data collection. A similar analogy would be the choice to measure resting blood pressure among ambulatory patients after quiet sitting as opposed to immediately after walking up the stairs. These procedures were implemented under the assumption that the clinical phenomena of spasticity are relatively stable over time, which may not be true.

Attempts to reduce variability in the measurement procedure included standardized test positions, right–left test order and the one-cycle per second metronome with measures performed on the second cycle. Although this is an unconventional means of administering the MAS, this methodology allowed us to ensure a consistent velocity for assessment across subjects and raters as it was impractical to apply the velocity constraint on the first cycle. Despite this standardization of the measurement procedure, our results are similar to Haas et al.[19] who compared the test–retest reliability of the Ashworth and MAS among subjects with SCI and lower extremity spasticity. In Haas' study,[19] the need to establish a standardized speed of muscle stretching during the test was identified. In addition, they identified that the effects of rater training was not well evaluated. Our study addressed some of these criticisms by using standardized test positions, use of a one-cycle per second metronome and extensive rater training before study initiation.

Despite our attempts to reduce subject variability and standardize the testing method,[23,24] our results are consistent with those of Allison et al.[23] and Blackburn et al.[24] Allison et al.[23] used the MAS to assess plantarflexor spasticity on 30 individuals with traumatic brain injury and found average intra-rater reliability ($\rho = 0.55$ and 0.74; $\kappa = 0.29$ and 0.69; $\tau{-}b = 0.48$ and 0.67) and average inter-rater reliability ($\rho = 0.73$; $\kappa = 0.4$; $\tau{-}b = 0.65$).[23] The poor reliability of the MAS when applied to the ankle was speculated as due to a short lever arm of the ankle making it more difficult to determine the resistance during movement. Blackburn et al.[24] used the MAS to assess lower-limb muscle spasticity of 20 patients 2 weeks after stroke, and repeated the test on 12 of the patients at 12 weeks after stroke. Inter-rater reliability of two raters was poor ($\tau{-}b = 0.062$ for combined muscle group), and intra-rater reliability was satisfactory ($\tau{-}b = 0.567$).[24]

Because of our study design and small sample size, we could not isolate whether the lack of agreement is due to variability in the subject's spasticity, the raters or the MAS tool. Despite having fair intra-rater reliability and poor inter-rater reliability, an inappropriate level of confidence has been placed in the MAS as a lower extremity spasticity measurement tool. We concur with the opinion of Haas et al.[19] that, '…the MAS is of limited use in the assessment of spasticity in the lower limb of patients with SCI' due to its inadequate reliability. Future research efforts should focus on identifying alternative tool(s) for quantification of lower extremity spasticity. Ideally, a new and improved tool would quantify both the tonic and phasic symptoms.

Conclusions

The reliability of the MAS for assessing lower extremity spasticity from a group of 20 subjects with diverse SCI impairment and severity of spasticity are reported. This commonly used clinical measure showed fair intra-rater reliability and poor inter-rater and inter-session reliability. The poor inter-rater and inter-session reliability of the MAS for lower extremity spasticity limit its validity and our future

ability to detect clinically meaningful changes in lower extremity spasticity (beyond standard error). It is inappropriate for SCI clinicians and researchers to continue to use assessment tools that are not psychometrically sound.[20] Perpetual use of an inadequate tool because of its familiarity is unacceptable, given the advancements in rehabilitation science and the earlier enunciated measurements standards for interdisciplinary rehab.[25] We recommended that the rehabilitation science community abandon the MAS. Several authors have purported the merit of a test battery approach, similar to that of the Spinal Cord Assessment Tool for Spastic Reflexes,[26] for the refinement of current measures or development of new spasticity measures.[3,4,9,10] These new or refined measures will also need to undergo formal assessment of their reliability and validity before broad dissemination. The author's seek to engage the clinical and research community in resolving this ongoing measurement dilemma.

References

1 Lance JW. The control of muscle tone, reflexes, and movement: Robert Wartenberg Lecture. *Neurology* 1980; **30**: 1303.
2 Decq P. Physiopathologie de la spasticite. *Neuro-Chirurgie* 2003: **49** (2–3 Pt 2): 163.
3 Adams MM, Hicks AL. Spasticity after spinal cord injury. *Spinal Cord* 2005; **43**: 577.
4 Biering-Sorensen F, Nielsen JB, Klinge K. Spasticity-assessment: a review. *Spinal Cord* 2006; **44**: 708.
5 Young RR. Spasticity: a review. *Neurology* 1994; **44**(11 Suppl 9): S12.
6 Anderson K, Aito S, Atkins M, Biering-Sørensen F, Charlifue S, Curt A et al. Functional recovery measures for spinal cord injury: an evidence-based review for clinical practice and research. *J Spinal Cord Med* 2008; **31**: 133.
7 Walter JS, Sacks J, Othman R, Rankin AZ, Nemchausky B, Chintam R et al. A database of self-reported secondary medical problems among VA spinal cord injury patients: its role in clinical care and management. *J Rehabil Res Dev* 2002; **39**: 53.
8 Canadian Paraplegic Association. *Spinal Cord Injury Workforce Participation National Survey*. Canadian Paraplegic Association: Ottawa, Ontario, Canada, 1997.
9 Priebe MM, Sherwood AM, Thornby JI, Kharas NF, Markowski J. Clinical assessment of spasticity in spinal cord injury: a multi-dimensional problem. *Arch Phys Med Rehabil* 1996; **77**: 713.
10 Hsieh JT, Wolfe DL, Miller WC, Curt A. Spasticity outcome measures in spinal cord injury: psychometric properties and clinical utility. *Spinal Cord* 2008; **46**: 86.
11 Skold C, Levi R, Seiger A. Spasticity after traumatic spinal cord injury: nature, severity, and location. *Arch Phys Med Rehabil* 1999; **80**: 1548.
12 Maynard FM, Karunas RS, Waring III WP. Epidemiology of spasticity following traumatic spinal cord injury. *Arch Phys Med Rehabil* 1990; **71**: 566.

13 Hsieh JTC, Wolfe DL, Connolly S, Townson AF, Curt A, Blackmer J *et al.* Spasticity after spinal cord injury: an evidence-based review of current interventions. *Top Spinal Cord Rehabil* 2007; **13**: 81.

14 Sackett DL, Stauss SE, Richardson WS. *Evidence-Based Medicine: How to Practice and Teach EBM.* Churchill-Livingstone: Philadelphia, PA, 2000.

15 Chou R, Peterson K, Helfand M. Comparative efficacy and safety of skeletal muscle relaxants for spasticity and musculoskeletal conditions: a systematic review. *J Pain Symptom Manage* 2004; **28**: 140.

16 Bovend'Eerdt TJ, Newman M, Barker K, Dawes H, Minelli C, Wade DT *et al.* The effects of stretching in spasticity: a systematic review. *Arch Phys Med Rehabil* 2008; **89**: 1395.

17 Wood DE, Burridge JH, van Wijck FM, McFadden C, Hitchcock RA, Pandyan AD *et al.* Biomechanical approaches applied to the lower and upper limb for the measurement of spasticity: a systematic review of the literature. *Disabil Rehabil* 2005; **27**: 19.

18 Bohannon RW, Smith MB. Interrater reliability of a modified Ashworth scale of muscle spasticity. *Phys Ther* 1987; **67**: 206.

19 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 [see comment]. *Spinal Cord* 1996; **34**: 560.

20 Johnston MV, Graves DE. Towards guidelines for evaluation of measures: an introduction with application to spinal cord injury. *J Spinal Cord Med* 2008; **31**: 13.

21 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159.

22 Fleiss JL. *Design and Analysis of Clinical Experiments.* John Wiley & Sons: Toronto, 1986.

23 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; **19**: 67.

24 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.

25 Johnston MV, Keith RA, Hinderer SR. Measurement standards for interdisciplinary medical rehabilitation. *Arch Phys Med Rehabil* 1992; **73**(12-S): S3.

26 Benz EN, Hornby TG, Bode RK, Scheidt RA, Schmit BD. A physiologically based clinical measure for spastic reflexes in spinal cord injury. *Arch Phys Med Rehabil* 2005; **86**: 52.