

ORIGINAL ARTICLE

Modified Ashworth scale and spasm frequency score in spinal cord injury: reliability and correlation

CB Baunsgaard^{1,3}, UV Nissen^{1,3}, KB Christensen² and F Biering-Sørensen¹

Study design: Intra- and inter-rater reliability study.

Objectives: To assess intra- and inter-rater reliability of the Modified Ashworth Scale (MAS) and Spasm Frequency Score (SFS) in lower extremities in a population of spinal cord-injured persons, as well as correlations between the two scales.

Setting: Clinic for Spinal Cord Injuries, Rigshospitalet, Hornbaek, Denmark.

Methods: Thirty-one persons participated in the study and were tested four times in total with MAS and SFS by three experienced raters. Cohen's kappa (κ), simple and quadratic weighted (nominal and ordinal scale level of measurement), was used as a measure of reliability and Spearman's rank correlation coefficient for correlation between MAS and SFS.

Results: Neurological level ranged from C2 to L2 and American Spinal Injury Association impairment scale A to D. Time since injury was (mean \pm s.d.) 3.4 ± 6.5 years. Age was 48.3 ± 20.2 years. Cause of injury was traumatic in 55% and non-traumatic for 45% of the participants. Antispastic medication was used by 61%. MAS showed intra-rater $\kappa_{\text{simple}} = -0.11$ to 0.46 and $\kappa_{\text{weighted}} = -0.11$ to 0.83 . Inter-rater $\kappa_{\text{simple}} = -0.06$ to 0.32 and $\kappa_{\text{weighted}} = 0.08$ to 0.74 . SFS showed intra-rater $\kappa_{\text{weighted}} = 0.94$ and inter-rater $\kappa_{\text{weighted}} = 0.93$. Correlation between MAS and SFS showed non-significant correlation coefficients from -0.11 to 0.90 .

Conclusion: Reliability of MAS is highly affected by the weighting scheme. With a weighted- κ it was overall reliable and simple- κ overall unreliability. Repeated tests should always be performed by the same rater and in a very standardized manner. SFS was found reliable. MAS and SFS are poorly correlated, and ratings were inversely distributed and suggest that it assesses different aspects of spasticity. *Spinal Cord* (2016) **54**, 702–708; doi:10.1038/sc.2015.230; published online 9 February 2016

INTRODUCTION

Spasticity is a symptom of the upper motor-neuron syndrome and common among people with spinal cord injury (SCI). The prevalence of spasticity among people with SCI varies in different studies from 65 to 78% depending on the definition of spasticity, level and severity of the SCI, method of measurement and possibly the time post injury.^{1–3} About half of these use antispastic medication.² In one study, roughly 40% reported the spasticity as problematic.³

An often used definition of spasticity is the one by Lance:⁴ 'a motor disorder characterized by a velocity-dependent increase in the tonic stretch reflex with exaggerated tendon jerks, resulting from hyper excitability of the stretch reflex, as one component of the upper motor-neuron syndrome'. This definition has the benefit of being precise but has been challenged⁵ and does not reflect the multidimensional nature of spasticity. For the purpose of this study, we used a relatively broader definition developed by the SPASM consortium:⁶ 'Disordered sensory–motor control resulting from an upper motor-neuron lesion presenting intermittent or sustained involuntary activity of muscles'. The definition distinguishes spasticity from the passive viscoelastic changes of muscle properties such as contractures, which are also associated with SCI.³ These changes in soft tissues and joints can be very difficult to distinguish clinically from the active part of spasticity.⁷ It is also important to keep in mind that

the symptoms increased tone, clonus and spasms, and hyperreflexia can exist independently of each other and do not necessary share the exact same pathophysiology.⁸

A number of ways to measure spasticity exist: scales for manually testing, biomechanical and electrophysiological methods,⁷ as well as self-rating scales.

A measurement of spasticity that is simple and easy to administer is important both in daily clinical practice, for monitoring fluctuations and changes over time, to measure the effect of antispastic medication, as well as for research purposes.

Two of such clinical measures of spasticity are The Modified Ashworth Scale (MAS) and the self-reported Spasm Frequency Score (SFS). Each of the two scales could be viewed as indirect measures of two single constructs of spasticity. The MAS is a scale of perceived resistance (tone) against passive movement of the limb and is an adjustment of the original Ashworth Scale where the category 1+ has been added. The SFS is a scale from 1 to 4 on the number of self-reported spasms per day.⁹ The SFS is an adaption of Penn SFS (PSFS) where self-reported number of spasms the last hour is rated.¹⁰ Definitions are shown in Table 1¹¹ and Table 2.⁹

Correlation between the MAS, other clinical measures, as well as biomechanical measures of spasticity has generally been low, and the validity of the MAS has been questioned.¹² Despite the limitations of

¹Clinic for Spinal Cord Injuries, Rigshospitalet, University of Copenhagen, Hornbaek, Denmark and ²Department of Public Health, Section of Biostatistics, University of Copenhagen, Hornbaek, Denmark

³These authors contributed equally to this work.

Correspondence: Dr CB Baunsgaard, Clinic for Spinal Cord Injuries, Rigshospitalet, University of Copenhagen, Havnevej 25, Hornbaek 3100, Capital Region, Denmark. E-mail: carsten.bach.baunsgaard@regionh.dk or cbbunsgaard@sund.ku.dk

Received 30 March 2015; revised 23 October 2015; accepted 30 October 2015; published online 9 February 2016

Table 1 The Modified Ashworth Scale (MAS)¹¹

Score	Modified Ashworth Scale
0	No increase in muscle tone.
1	Slight increase in muscle tone, manifested by a catch and release or by minimal resistance at the end ROM when the affected part(s) is moved in flexion or extension.
1+ [2]	Slight increase in muscle tone, manifested by a catch, followed by minimal resistance throughout the remainder (less than half) of the ROM.
2 [3]	More marked increase in muscle tone through most of the ROM, but the affected part(s) is easily moved.
3 [4]	Considerable increase in muscle tone, passive movement is difficult.
4 [5]	Affected part(s) rigid in flexion or extension.

Abbreviation: ROM, range of motion.
In square brackets, transcription sometimes used for the Modified Ashworth Scale.

Table 2 Spasm Frequency Score⁹

Score	Spasm Frequency Score
0	No spasms
1	One or a few spasms per day
2	Between 1 and 5 spasms per day
3	5 or <10 spasms per day
4	Ten or more spasms per day or continues contraction

the MAS, it is still the most widely used scale in the clinical setting,³ and its use is recommended by the National Institute of Neurological Disorders and Stroke Common Data Element.¹³ The MAS has been tested in populations with spasticity after SCI, stroke, multiple sclerosis, traumatic brain injury, as well as cerebral palsy, showing varying results of the reliability; in general, the reliability was unsatisfactory.^{7,14,15} The mechanism of spasticity between SCI and other upper motor-neuron lesions may have a different etiology.¹⁶ A literature search found four studies^{17–20} that assessed the reliability of the MAS of lower extremities among people with SCI. The choice of statistical method differed; some used simple (unweighted) kappa other used weighted kappa, while others used correlation coefficients.⁷ There were also differences in test protocols and differences in how confounders were controlled. This makes studies difficult to compare, and in light of this we find that there still is a need to reassess the MAS for lower extremity after SCI. The need to reassess MAS and self-reported spasticity scales are also supported by the review from Hsieh *et al*.¹⁴

Self-rating scales are a way of clinically assessing another construct of spasticity. To our knowledge, test–retest reliability has been studied in only one published study of the PSFS and none of the SFS scale. The study showed identical test–retest scores for PSFS in 40% of observations.²¹ Reliability of a self-reported severity of spasticity on a numeric rating scale from 0 to 10 in a population of multiple sclerosis found the scale to be reliable.²² Correlations between self-reported spasticity and examiner-based assessments have shown varying results from significant to poor correlation.^{21,23}

The aim of this study is to test the reliability of MAS and SFS in a population of individuals with SCI in a standardized setting and test the correlation between the two scales.

MATERIALS AND METHODS

Participants

Data were collected from February 2010 to February 2011. A total of 31 participants with SCI and the presence of spasticity in the lower extremities were enrolled in the study. All participants were inpatients at the Clinic for Spinal Cord Injuries, Rigshospitalet, Hornbaek, Denmark. Inclusion criteria were SCI, presence of spasticity and good general health condition. Exclusion criteria were new or ongoing urinary tract infection and severe limitation in

range of motion (ROM) with a limited ROM of knee and hip together of > 30°, same criteria as previously used by Craven and Moris 2010.²⁰

Raters and procedure

All assessments were carried out by three physiotherapists experienced in treating people with SCI and used MAS in their daily practice. To ensure MAS assessment was performed in a standardized and an uniform manner,³ the three raters practiced the MAS assessment and interpretation prior to the study. During the data collection, the raters did not discuss the testing procedures, outcomes or other study-related issues.

To keep testing conditions as constant as possible, all assessments were carried out at the same time in the morning, before participants got out of bed, and participants only wore the clothes they had been sleeping in. This was done to keep activity level before testing at an absolute minimum. Participants who did not have an indwelling catheter went to the bathroom before testing.

The participants were moved directly from bed to bench in supine position to avoid any unnecessary activity. Testing was always performed for the right side first in the order of hip flexors, hip extensors, knee flexors, knee extensors, ankle plantar flexors and ankle dorsi flexors. Then, the rater moved around the bench and performed the same procedure for the left side.

On the day of MAS assessment, participants were screened for symptoms of urinary tract infection²⁴ and asked whether they felt ok, and pain was documented if present. According to the protocol, the rater was instructed to postpone testing to another day if the participant did not feel ok. This was not necessary for any of the tests however. Room temperature was measured, and the participant was asked to rate the number of spasms the day before according to the SFS.

During each movement raters counted ‘one second’ to move the joint as smooth and constant as possible, to imitate the way MAS was performed in the original Bohannon study.¹¹ It was decided not to use a metronome, as we wanted the test procedure to be similar to the daily clinical setting.

In case the movement triggered clonus, making MAS assessment impossible, it was reported as clonus and a missing value for the MAS analysis.

Information on medication was collected from medical records. No changes in regular medication were made between the two days, and medicine was administered at the same time of the day.

All tests were performed four times on each participant. Two ratings were performed at the first test day (Tuesday) and two on the second test day (Thursday) within the same week. Two different raters each performed one test on the first day for inter-rater evaluations. On the second day, one of the raters from the first test day would perform a retest for intra-rater evaluations and a third rater performed the last test also for inter-rater evaluations. The raters followed a rolling schedule to minimize possible inter-rater bias, by distributing the tests in a more random manner among the raters.

Statistical analysis

Statistical analysis was performed by SAS software package version 9.4 (SAS Institute Inc., Cary, NC, USA) and IBM SPSS Statistics version 22 (IBM Corp. Released 2013. Armonk, NY, USA). Kappa (κ) statistics was chosen as a reliability measure of MAS and SFS, as this is recommended as the most appropriate measure.^{7,12} The original Ashworth Scale (AS) and MAS are designed with the level of measurement as ordinal scales, but it has been

questioned whether the MAS can be treated as such, but rather should be considered a nominal scale.¹² Whether the scale is considered ordinal or nominal is important for the choice of statistics. Both simple and weighted kappa values for MAS were calculated for comparison for reasons discussed further below. The SFS were considered an ordinal scale. When using weighted kappa, different weighting schemes can be chosen.²⁵ The quadratic weights were chosen, as quadratic weights yield a kappa coefficient that is equivalent to the ICC under certain conditions.^{26,27}

Intra-rater reliability (κ) was performed by the same rater on different days. Inter-rater reliability was calculated between the two ratings carried out on the same day by different raters. Data from day one and two were merged for the inter-rater analysis. The same procedure was used for reliability testing of SFS.

Interpretation of strength of agreement of the κ -value is based on Landis and Koch:²⁸ Poor $\kappa < 0$; Slight $\kappa = 0-0.20$; Fair $\kappa = 0.21-0.40$; Moderate $\kappa = 0.41-0.60$; Substantial $\kappa = 0.61-0.80$; and Almost perfect $\kappa = 0.81-1.00$.

Difference between first and second rating on the day of MAS is reported by median and interquartile range of the two corresponding ratings. The same procedure is used to test for systematic difference between raters.

For correlation analysis between MAS scores and SFS self-reported frequencies of spasms, Spearman's rank correlation coefficient was used with Bonferroni correction for multiple comparisons.

Statement of ethics

We certify that all applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed during the course of this research. All participants received written as well as oral information about the study before giving their written consent. One participant was below 18 years of age, and both parents gave written consent. The study followed the guidelines of the Helsinki Declaration.

RESULTS

Cohort data

Thirty-four participants with SCI were recruited for the study. Three participants were excluded because of urinary tract infection appearing during the study period, leaving 31 participants for analyses. Study population characteristics are described in Table 3 and reported according to recommendations.²⁹

Room temperature varied between 20 and 23 °C over the data collection period. Room temperature did not vary across test days for each participant.

The distribution of the location of MAS scores is shown in Figure 1. This figure shows a majority of the value zero, in particular for knee extensors, ankle dorsiflexors and hip flexors. Distribution of MAS and SFS scores is shown in Figures 2 and 3, respectively. The total number of ratings from all four assessments is reported.

Difference between first and second rating showed a median of the difference (interquartile range) for the first day: 0 (−0.25;0) and the second day 0 (0;1). Similar analyses addressing the levels of the individual raters suggested that one rater had a tendency to score higher than the other rater on the same day and that one rater had a tendency to score lower than the other rater on the same day (results not shown).

MAS reliability

Intra-rater reliability ranged for the weighted kappa from 'substantial' to 'almost perfect', except for ankle dorsiflexors, which were 'poor' to 'slight'. Simple kappa ranged from 'poor' to 'moderate' for all muscle groups.

Inter-rater reliability ranged for the weighted kappa from 'fair' to 'substantial', except ankle dorsiflexors, which had 'poor' to 'slight' reliability. Simple kappa ranged from 'poor' to 'fair'.

Table 4 shows the results of intra- and inter-rater reliability of the 12 assessed muscle groups.

Table 3 Study population characteristics

	N = 31
Age, mean \pm s.d. (median; range) (years)	48.3 \pm 20.2 (53; 15–88)
Gender, Number (%)	
Male	20 (65%)
Female	11 (35%)
Time since injury, mean \pm s.d. (median; range) (years)	3.4 \pm 6.5 (0.4; 27)
0–12 month since injury	65%
> 12 month since injury	35%
Cause of injury, number (%)	
Traumatic	17 (55%)
Nontraumatic	14 (45%)
Neurological level and severity of injury, number (%)	
C1–C4 AIS A, B, C	3 (10%)
C5–C8 AIS A, B, C	6 (19%)
T1–S5 AIS A, B, C	9 (29%)
All AIS D	13 (42%)
Medication at the time of the study	
Antispastic	61%
Analgetic	18%

Abbreviation: AIS, American Spinal Injury Association Impairment Scale.

Clonus, resulting in a missing value, was reported for plantar flexors in 24% of the cases for intra-rater ratings and 29% for inter-rater ratings, calculated as an average of left and right side. It was evenly distributed on right and left side. A few instances of clonus were reported for dorsal flexors but none for the remaining muscle groups.

SFS reliability

Intra-rater reliability of SFS showed weighted $\kappa = 0.94$ (95% CI 0.87–1) and simple $\kappa = 0.80$ (95% CI 0.62–0.98). Inter-rater reliability showed weighted $\kappa = 0.93$ (95% CI 0.89–0.98), and simple kappa $\kappa = 0.74$ (95% CI 0.60–0.88). Interpretation of the weighted kappa for both groups was 'almost perfect' and with simple kappa for both groups 'substantial'. The corresponding crude agreement to the kappa values was approximately 80% for both groups.

Correlation between MAS and SFS

Spearman's rank correlation rho was calculated between observed spasticity (MAS) and self-reported spasms (SFS). The correlation coefficients ranged from a negative to a weak correlation (Table 5). Significance levels were not met ($P < 0.004$) after Bonferroni correction for multiple comparisons.

DISCUSSION

The objective of the study was to test reliability of the MAS and SFS in lower extremities in a population of SCI individuals and to test how the two scales correlate, as advised by previous reviews.^{7,14}

Reliability of the MAS is greatly affected by the choice of using weighted or simple kappa coefficients. Overall with the weighted kappa values, both intra- and inter-rater reliability was satisfactory, except for ankle dorsiflexors, where the poorest reliability also has

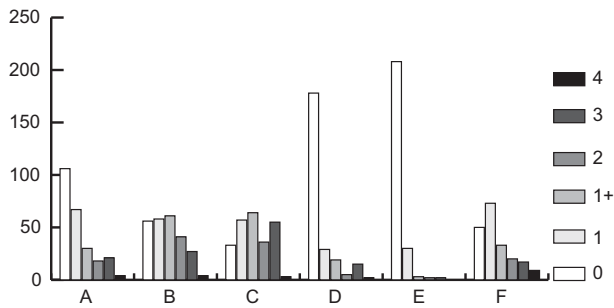


Figure 1 Distribution of MAS values based on location. The figure shows the distribution of MAS values (cf. Table 1) based on location of the muscle group. The number counts on the vertical axis show the total counts of all ratings performed. Right and left side is added together. A, hip flexors; B, hip extensors; C, knee flexors; D, knee extensors; E, ankle dorsiflexors; F, ankle plantar flexors. MAS, Modified Ashworth Scale.

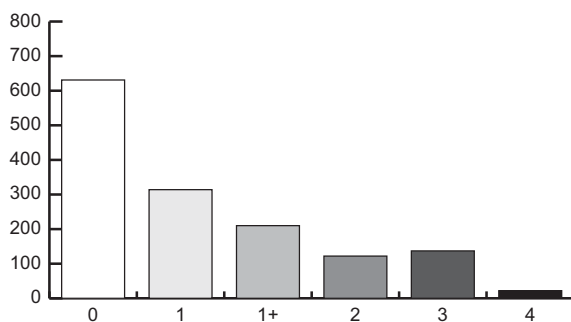


Figure 2 Distribution of MAS values. The figure shows the distribution of all MAS values (cf. Table 1). MAS, Modified Ashworth Scale.

been found in previous studies.¹⁷ With simple kappa, neither intra- nor inter-rater reliability was satisfactory. Intra-rater reliability is higher for test-retest by the same rater than inter-rater reliability irrespective to weighting or not. Results showed that some raters had a tendency to systematically score higher than others on the MAS. This could partly explain some of the disagreement between raters.

The standardization made in the study, in order to minimize confounding factors, could give rise to a higher reliability than may be seen in the clinical setting, where the same standardization would not always be possible. Examples of this are the activity prior to examination, positioning of patient during examination, angle of joints during measurements where the inertia could affect hip and knee differently, etc. It is thus important for repeated measures that the measures are taken in a very standardized way. Repeated measures of MAS should when possible be performed by the same rater, as reliability is higher and the measure could be performed in a more uniform manner. This is in line with the recommendation of previous studies.⁷

The prerequisite for using a weighted kappa is that the scale can be considered an ordinal level of measure. Following this a higher MAS score should reflect a higher degree of spasticity. The ordinal nature of MAS has been questioned for a number of reasons.¹² First, it has been argued that adding the 1+ category introduced ambiguity in the scale, as the ordinality of the 1 and 1+ category is questionable and thus making it a nominal scale based on subjective criteria. Second, lack of biomechanical references to 'catch' and 'release' in the definition of MAS and third, the MAS description 'minimal resistance to passive movement at the end of movement' is hard to distinguish from changes due to passive viscoelastic changes often seen after

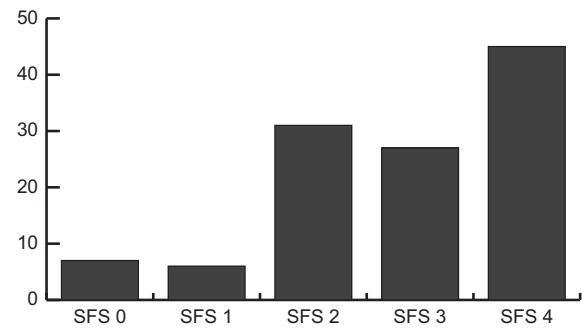


Figure 3 Distribution of SFS values. The figure shows the distribution of all SFS values (cf. Table 2). SFS, Spasm Frequency Score.

neurological injury. These last two arguments are also indirectly supported by the poor correlation between MAS and biomechanical measures documented in previous studies.^{30,31}

Most of the reported MAS values were 0, 1 and 1+ (~80%, Figure 2). By regarding MAS a nominal scale, only exact agreement will be captured as agreement between raters, as simple kappa is blind to off-diagonal association, when more than two categories exist.²⁵ Overlooking the validity issues of the 1+ category of MAS and seeing isolated from a reproducibility point of view, one could argue that a difference between two raters of one rank category is a sign of better agreement compared with a difference of two or more. This partial agreement would be captured by the weighting scheme and could argue for including the weighted kappa in the evaluation alongside with the simple kappa. It also suggests that the poor reliability seen in the simple kappa is due to problems when distinguishing between values at the lower end of the scale. The fact that it is difficult to distinguish between spasticity and changes in viscoelastic properties and that a certain degree of contractures was present, would very likely be an explanation for the poor reliability.

SFS has not previously been tested for reliability, beside one study showing a crude agreement of 40% of the PSFS, which is a scale similar to the SFS. The present study showed almost perfect reliability when data are collected both by the same or different raters. As it is self-reported number of spasms one could expect a smaller difference between intra- and inter-raters, compared with a rater-based scale. Our results support this assumption.

The distribution of the SFS ratings shows the majority of ratings to be in the highest category 4 followed by the category 3 and 2. Only a minority of the answers were in the category 1 or 0, meaning a few or none spasms per day. One of the inclusion criteria was the presence of spasticity, so the low number of the value zero is not surprising, but the skew toward the high values could also be interpreted as a sign of low sensitivity of the scale as previously suggested.⁷ It should also be noted that the questions in SFS are not location specific. The patients understanding of what is meant by spasms could also vary, as it is not defined in the description of the scale and could be an explanation to the skew toward the higher SFS values. The distribution of MAS was inversely skewed compared with SFS. Correlation analysis between MAS ratings and SFS shows a weak, non-significant correlation in 10 of 12 ratings. The remaining 2 of 12 showed a negative, non-significant correlation. The inverse skews and poor correlation indicate that MAS and SFS examine two different aspects of spasticity. A previous study by Priebe *et al.*²¹ supports the finding of poor correlation between self-reported severity and the number of spasms and clinical examination scales (modified PSFS). The fact that it has

Table 4 Modified Ashworth Scale intra- and inter-rater reliability

Location	n	Missing values	κ -quadratic weight (95% confidence limits)	Interpretation	κ -simple (unweighted) (95% confidence limits)	Interpretation
<i>Intra-rater</i>						
Hip, right flexors	31	0	0.81 (0.66–0.96)	Almost perfect	0.38 (0.18–0.59)	Fair
Hip, left flexors	31	0	0.75 (0.56–0.93)	Substantial	0.32 (0.08–0.56)	Fair
Hip, right extensors	31	0	0.79 (0.66–0.91)	Substantial	0.32 (0.11–0.53)	Fair
Hip, left extensors	31	0	0.70 (0.56–0.84)	Substantial	0.10 (–0.09–0.29)	Slight
Knee, right flexors	31	0	0.73 (0.55–0.91)	Substantial	0.39 (0.18–0.60)	Fair
Knee, left flexors	31	0	0.83 (0.72–0.93)	Almost perfect	0.34 (0.12–0.57)	Fair
Knee, right extensors	31	0	0.79 (0.55–1.00)	Substantial	0.30 (–0.04–0.63)	Fair
Knee, left extensors	31	0	0.71 (0.47–0.95)	Substantial	0.43 (0.17–0.69)	Moderate
Ankle, right dorsiflexors	30	1	0.04 (–0.34–0.42)	Slight	0.04 (–0.34–0.42)	Slight
Ankle, left dorsiflexors	30	1	–0.11 (–0.23–0.01)	Poor	–0.11 (–0.23–0.01)	Poor
Ankle, right plantarflexors	23	8*	0.61 (0.32–0.89)	Substantial	0.46 (0.21–0.71)	Moderate
Ankle, left plantarflexors	24	7*	0.75 (0.49–1.00)	Substantial	0.46 (0.20–0.72)	Moderate
<i>Inter-rater</i>						
Hip, right flexors	61	1	0.54 (0.33–0.76)	Moderate	0.17 (0.01–0.34)	Slight
Hip, left flexors	61	1	0.52 (0.35–0.69)	Moderate	0.21 (0.06–0.36)	Fair
Hip, right extensors	61	1	0.74 (0.63–0.86)	Substantial	0.23 (0.08–0.39)	Fair
Hip, left extensors	62	1	0.54 (0.39–0.68)	Moderate	–0.03 (–0.16–0.10)	Poor
Knee, right flexors	62	0	0.67 (0.53–0.81)	Substantial	0.31 (0.15–0.46)	Fair
Knee, left flexors	62	0	0.55 (0.36–0.74)	Moderate	0.17 (0.02–0.32)	Slight
Knee, right extensors	62	0	0.59 (0.37–0.80)	Moderate	0.23 (0.05–0.41)	Fair
Knee, left extensors	62	0	0.72 (0.55–0.89)	Substantial	0.32 (0.13–0.50)	Fair
Ankle, right dorsiflexors	60	2	0.07 (–0.18–0.33)	Slight	0.02 (–0.19–0.23)	Slight
Ankle, left dorsiflexors	61	1	0.08 (–0.17–0.34)	Slight	–0.06 (–0.20–0.07)	Poor
Ankle, right plantarflexors	44	18*	0.25 (0.00–0.50)	Fair	0.28 (0.10–0.46)	Fair
Ankle, left plantarflexors	44	18*	0.47 (0.18–0.76)	Moderate	0.29 (0.10–0.48)	Fair

Asterisks (*) show missing values due to clonus.

Table 5 Correlations between MAS and SFS

Location of spasticity (muscle/group)	Spearman's rho	P-value
Hip, right flexors	0.9	0.34
Hip, left flexors	0.13	0.15
Hip, right extensors	0.21	0.02
Hip, left extensors	0.08	0.38
Knee, right flexors	0.21	0.03
Knee, left flexors	0.03	0.72
Knee, right extensors	0.01	0.97
Knee, left extensors	-0.13	0.16
Ankle, right dorsiflexors	-0.04	0.65
Ankle, left dorsiflexors	0.03	0.77
Ankle, right plantarflexors	0.18	0.09
Ankle, left plantarflexors	0.18	0.09

Abbreviations: MAS, Modified Ashworth Scale; SFS, Spasm Frequency Scale.
Significance level α , after Bonferroni correction: $P < 0.004$

previously been shown that 60% of people with self-reported spasticity after SCI elicit spasticity from movement³ stresses the importance of adding other measure of spasticity in addition to MAS to capture more aspects of spasticity.

As movements of the limb can affect MAS,¹² we tested for the difference between the first and second rating to see whether the second rating would be lower than the first rating. The median of the difference was zero on both days, with interquartile range on the first day suggesting a lowering of the value between first and second, whereas the opposite was the case on second test day. Hence, in conclusion it does not seem to be the case that the movements systematically affected the MAS measure.

The study was designed to reflect the use of MAS in the clinical setting, and the choice of study population was chosen to reflect this. A certain degree of contractures in the SCI population is very common, and because of this a certain degree of contractures for the participants was accepted. It was chosen not to use a metronome, but as we do not have documentation on the actual speed of the movement this could be a confounder to distinguish the active part of spasticity from passive changes in viscoelastic properties.

Clonus was present in roughly a quarter of the plantar flexor measurements and were excluded from the κ -analysis, which might have affected the reliability. The skews in MAS and SFS have probably also affected the kappa scores. The rolling schedule of the raters led to uneven distribution of the number of ratings each rater performed in total, which also could be a confounder.

Regular medicine of the participant was administered at the same time of the day, but the time when the participant actually took the medicine was not documented. Time of awakening in the morning was documented as part of ensuring that the tests were the first morning activity for the participants, and they were asked whether they felt ok. The total amount or lack of sleep was not documented however.

CONCLUSION

The MAS is found to have acceptable reliability when partial agreement is included (weighted kappa), under the assumption of ordinality of the scale, but there are poor intra- and inter-rater reliability for exact agreement (simple kappa). Repeated measures should always be performed by the same rater when possible, as the intra-rater reliability is higher and the measurements are to be taken in a very standardized way for repeated measures. Clonus and limited

range of motion, which are highly prevalent in the SCI population, complicate assessment. Together with well-known difficulties in distinguishing spasticity from passive viscoelastic changes, this is limitations to the scale. The standardization in the setting when MAS was measured for this study could have given rise to a higher reliability than what would be seen in the clinical setting where the same standardization not always is possible. The SFS is highly reliable both for intra- as well as inter-raters, which are a benefit in the clinical setting, but the scale might have low sensitivity. Self-reported spasms on SFS and clinical examination with the MAS correlate poor to one another and show inverse skews. Our conclusion is that SFS and MAS assess different constructs of spasticity.

DATA ARCHIVING

There were no data to deposit.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the three physiotherapists Inge Bjerregaard, Mette Skov Henriksen and Alec Ellenson, who performed the MAS tests and the SFS data collection. We also thank the Bevica-, Tryk- and Augustinus foundation for funding the study.

- Hitzig SL, Tonack M, Campbell KA, McGillivray CF, Boschen KA, Richards K *et al*. Secondary health complications in an aging Canadian spinal cord injury sample. *Am J Phys Med Rehabil Assoc Acad Physiatr* 2008; **87**: 545–555.
- Maynard FM, Karunas RS, Waring WP. Epidemiology of spasticity following traumatic spinal cord injury. *Arch Phys Med Rehabil* 1990; **71**: 566–569.
- Sköld C, Levi R, Seiger A. Spasticity after traumatic spinal cord injury: nature, severity, and location. *Arch Phys Med Rehabil* 1999; **80**: 1548–1557.
- Lance J. *Symposium synopsis*, In: Young RR, Koeila WE (eds). *Spasticity: Disordered Motor Control*. Yearbook Medical Publishers: Chicago, IL, USA. 1980, 485–494.
- Malhotra S, Pandyan AD, Day CR, Jones PW, Hermens H. Spasticity, an impairment that is poorly defined and poorly measured. *Clin Rehabil* 2009; **23**: 651–658.
- Pandyan AD, Gregoric M, Barnes MP, Wood D, Van Wijck F, Burridge J *et al*. Spasticity: clinical perceptions, neurological realities and meaningful measurement. *Disabil Rehabil* 2005; **27**: 2–6.
- Biering-Sørensen F, Nielsen JB, Klinge K. Spasticity-assessment: a review. *Spinal Cord* 2006; **44**: 708–722.
- Nielsen JB, Crone C, Hultborn H. The spinal pathophysiology of spasticity—from a basic science point of view. *Acta Physiol Oxf Engl* 2007; **189**: 171–180.
- Snow BJ, Tsui JKC, Bhatt MH, Varelas M, Hashimoto SA, Calne DB. Treatment of spasticity with botulinum toxin: A double-blind study. *Ann Neurol* 1990; **28**: 512–515.
- Penn RD, Savoy SM, Corcos D, Latash M, Gottlieb G, Parke B *et al*. Intrathecal baclofen for severe spinal spasticity. *N Engl J Med* 1989; **320**: 1517–1521.
- Bohannon RW, Smith MB. Interrater reliability of a modified Ashworth scale of muscle spasticity. *Phys Ther* 1987; **67**: 206–207.
- Pandyan AD, Johnson GR, Price CI, Curless RH, Barnes MP, Rodgers H. A review of the properties and limitations of the Ashworth and modified Ashworth Scales as measures of spasticity. *Clin Rehabil* 1999; **13**: 373–383.
- NIH common data elements. Available at http://www.commondataelements.ninds.nih.gov/SCI.aspx#tab=Data_Standards. Accessed 16 December 2015.
- Hsieh JTC, Wolfe DL, Miller WC, Curt A. SCIRE Research Team. Spasticity outcome measures in spinal cord injury: psychometric properties and clinical utility. *Spinal Cord* 2008; **46**: 86–95.
- Bhimani RH, Anderson LC, Henly SJ, Stoddard SA. Clinical measurement of limb spasticity in adults: state of the science. *J Neurosci Nurs* 2011; **43**: 104–115.
- Noth J. Trends in the pathophysiology and pharmacotherapy of spasticity. *J Neurol* 1991; **238**: 131–139.
- Haas BM, Bergström E, Jamous A, Bennie A. The inter rater reliability of the original and of the modified Ashworth scale for the assessment of spasticity in patients with spinal cord injury. *Spinal Cord* 1996; **34**: 560–564.
- Smith AW, Jamshidi M, Lo SK. Clinical measurement of muscle tone using a velocity-corrected modified Ashworth scale. *Am J Phys Med Rehabil* 2002; **81**: 202–206.
- Tederko P, Krasuski M, Czech J, Dargiel A, Garwacka-Jodzis I, Wojciechowska A. Reliability of clinical spasticity measurements in patients with cervical spinal cord injury. *Ortop Traumatol Rehabil* 2007; **9**: 467–483.
- Craven BC, Morris AR. Modified Ashworth scale reliability for measurement of lower extremity spasticity among patients with SCI. *Spinal Cord* 2010; **48**: 207–213.

- 21 Priebe MM, Sherwood AM, Thornby JI, Kharas NF, Markowski J. Clinical assessment of spasticity in spinal cord injury: a multidimensional problem. *Arch Phys Med Rehabil* 1996; **77**: 713–716.
- 22 Farrar JT, Troxel AB, Stott C, Duncombe P, Jensen MP. Validity, reliability, and clinical importance of change in a 0–10 numeric rating scale measure of spasticity: a post hoc analysis of a randomized, double-blind, placebo-controlled trial. *Clin Ther* 2008; **30**: 974–985.
- 23 Lechner HE, Frotzler A, Eser P. Relationship between self- and clinically rated spasticity in spinal cord injury. *Arch Phys Med Rehabil* 2006; **87**: 15–19.
- 24 Kheder A, Nair KPS. Spasticity: pathophysiology, evaluation and management. *Pract Neurol* 2012; **12**: 289–298.
- 25 Graham P, Jackson R. The analysis of ordinal agreement data: beyond weighted kappa. *J Clin Epidemiol* 1993; **46**: 1055–1062.
- 26 Brenner H, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology* 1996; **7**: 199–202.
- 27 Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973; **33**: 613–619.
- 28 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.
- 29 DeVivo MJ, Biering-Sørensen F, New P, Chen Y. International Spinal Cord Injury Data Set. Standardization of data analysis and reporting of results from the international spinal cord injury core data set. *Spinal Cord* 2011; **49**: 596–599.
- 30 Pandyan AD, Price CIM, Barnes MP, Johnson GR. A biomechanical investigation into the validity of the modified Ashworth Scale as a measure of elbow spasticity. *Clin Rehabil* 2003; **17**: 290–293.
- 31 Pandyan AD, Price CIM, Rodgers H, Barnes MP, Johnson GR. Biomechanical examination of a commonly used measure of spasticity. *Clin Biomech* 2001; **16**: 859–865.