



Intra-rater and inter-rater reliability of the Penn Spasm Frequency Scale in People with chronic traumatic spinal cord injury

Patricia B. Mills^{1,2,6} · Alok P. Vakil³ · Cassidy Phillips⁴ · Lawrence Kei² · Brian K. Kwon^{1,5}

Received: 15 November 2017 / Revised: 30 December 2017 / Accepted: 8 January 2018 / Published online: 25 January 2018
© International Spinal Cord Society 2018

Abstract

Study design Psychometric study.

Objectives Evaluate the intra-rater and inter-rater reliability of the Penn Spasm Frequency Scale (PSFS) in people with chronic traumatic spinal cord injury (SCI).

Setting Community setting.

Methods Participants with chronic traumatic SCI were administered the PSFS, a self-reported spasticity measure in which the individual documents the frequency (Part 1) and severity (Part 2) of their muscle spasms. One rater administered the PSFS at baseline, 5 to 10 days, and 4 to 6 weeks after baseline. The second rater administered the PSFS within 3 days of the first rater.

Results Reliability was assessed using weighted-kappa (95% confidence interval) statistics. In 61 participants, the intra-rater reliability between 5 to 10 days and 4 to 6 weeks after baseline was 0.822 (0.709, 0.935) and 0.734 (0.586, 0.883), respectively, for PSFS Part 1. With the addition of Part 2, the intra-rater reliabilities were 0.812 (0.705, 0.919) and 0.729 (0.586, 0.872) for 5 to 10 days and 4 to 6 weeks, respectively. The PSFS inter-rater reliability within a 3-day time interval was 0.862 (0.759, 0.965) for Part 1 and 0.857 (0.762, 0.952) with the addition of Part 2.

Conclusions PSFS Part 1 intra-rater and inter-rater reliability weighted-kappa statistics are in the “almost perfect” category, with lower confidence bounds in the “substantial” range. By adding Part 2 the reliability decreases, but the kappa statistics remain “substantial,” with a lower confidence bound in the “moderate” range. The PSFS is probably a reliable outcome measure to assess self-reported spasticity after chronic traumatic SCI.

Introduction

Spasticity, a sensorimotor disorder characterized by intermittent or sustained involuntary muscle activation [1], is a common secondary complication after spinal cord injury (SCI). A large prospective study identified the prevalence of spasticity in people with chronic traumatic SCI as high as 65%, and “problematic spasticity,” as defined by interference of function and/or the need for treatment, as high as 41%, despite current management strategies [2].

Therefore, there is a need to further investigate treatment options for spasticity management in SCI. Studies require outcome measures with well-established psychometric properties. Spasticity in people with SCI typically fluctuates throughout the day and from day to day [3]. Consequently, spasticity is difficult to assess solely with objective clinical measures; the additional use of self-report measures is thought to more accurately capture the experience of spasticity in people with SCI [4, 5]. The Penn Spasm Frequency

✉ Patricia B. Mills
patricia.mills@vch.ca

¹ ICORD (International Collaboration on Repair Discoveries), University of British Columbia, Vancouver, BC, Canada

² Department of Medicine, Division of Physical Medicine and Rehabilitation, University of British Columbia, Vancouver, BC, Canada

³ University of Alberta, Edmonton, AB, Canada

⁴ Queens University, Kingston, ON, Canada

⁵ Department of Orthopaedics, Vancouver Spine Surgery Institute, University of British Columbia, Vancouver, BC, Canada

⁶ Rehabilitation Research Program, Vancouver Coastal Health Research Institute, Vancouver, BC, Canada

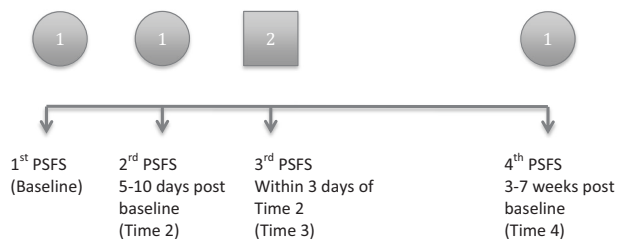


Fig. 1 Study timeline per participant. Circle with number 1 = evaluator 1. Square with number 2 = evaluator 2. Arrows = time interval for phone call administering PSFS to participant. (Colour figure online)

Scale (PSFS) is a measure of self-assessed muscle spasm frequency and severity commonly applied in studies assessing spasticity in the SCI population [5]. The PSFS is the shortest self-report questionnaire available (<5 min) and therefore has the least burden for administration. The construct validity of the PSFS has been previously studied with comparison to other measures including the Ashworth Scale, Spinal Cord Assessment Tool for Spinal Reflexes, and the SCI Spasticity Evaluation Tool [6, 7]. However, the intra-rater and inter-rater reliability of the PSFS has not been reported in the literature, thereby limiting its current use in research. The PSFS can be used to characterize the state of spasticity in an individual at any given time and to measure a treatment response to intervention. As oral medications for spasticity take approximately 1 week to exert their initial effect [8], and intramuscular botulinum toxin for spasticity management typically has peak action between 4 and 6 weeks after injection [9], the reliability of the PSFS needs to be established in these time intervals if it is to be used as an instrument to measure response to these anti-spasmodic interventions.

The objective of this study was to evaluate the intra-rater and inter-rater reliability of the PSFS in individuals with chronic traumatic SCI within 5–10 days and 4–6 weeks after baseline administration for intra-rater reliability and within 3 days of administration for inter-rater reliability.

Methods

Participants

Potential participants were identified through two recruitment methods. The first was through the local Rick Hansen Spinal Cord Injury Registry (RHSCIR), a national registry that collects prospective information on individuals with traumatic SCI. Patients contacted with this approach were those who sustained an acute traumatic SCI, were admitted to the acute care (Vancouver General Hospital) and rehabilitation (GF Strong Rehab Centre) hospitals in Vancouver, British Columbia, participated in RHSCIR, and indicated the presence of spasticity on community discharge and

community follow-up questionnaires. The second recruitment method was identification of previous SCI inpatients of the GF Strong Rehab Centre using the hospital medical record system, so as to allow inclusion of local SCI patients who were not enrolled in RHSCIR to participate in the study. Inclusion criteria included ≥ 1 year duration of traumatic SCI, age 18–65 years, currently experiencing spasticity on history (defined by the presence of intermittent or sustained involuntary muscle activation below the level of SCI), history of stable spasticity over the past 2 weeks (e.g., consistent medication routine, absence of conditions known to affect spasticity such as urinary tract infection), and no anticipated change in spasticity treatment during study enrollment. Participants were excluded if they were not English speaking or could not provide informed consent secondary to cognitive disorders such as concomitant traumatic brain injury. Potential participants were contacted through phone calls for eligibility screening, enrollment, collection of demographic data, and administration of the PSFS. Demographic data included gender, age, time since injury, and neurological level of injury and injury severity as per the International Standards for the Neurological Classification of SCI that utilizes the American Spinal Injury Association Scale (AIS) [10]. Ethical approval for the study was obtained from the university institutional review board. Written informed consent was obtained from all participants.

Raters

The raters were individuals who were either completing or had received an undergraduate level of education (APV, CP, LK). All raters participated in a group 1 h training session on use of the PSFS led by the senior author (PBM).

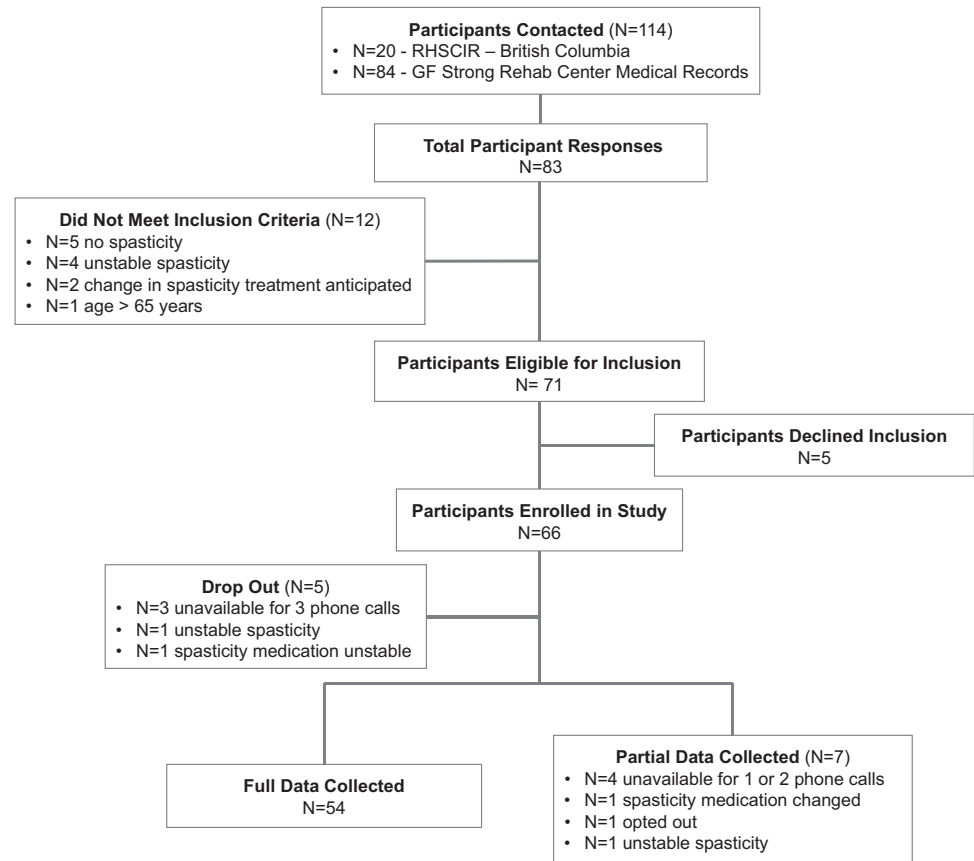
Outcome measure

The PSFS is a self-report measure composed of two parts (see Appendix). For Part 1, participants are asked to rate their spasm frequency during the past 7 days on a 5-level scale ranging from 0 = no spasms to 4 = spasms occurring more than 10 times per hour. If the participant indicates no spasms in Part 1, then they do not proceed to Part 2. Part 2 of the PSFS is a three-level scale assessing the severity of spasms. The PSFS was administered over the phone, so as to minimize participant burden with use of this outcome measure in potential future studies.

Intra-rater and inter-rater reliability

To assess intra-rater reliability, one evaluator administered the PSFS Part 1 and Part 2 over the phone at the time of enrollment into the study (baseline, Time 1), within

Fig. 2 Study participant flowchart. (Colour figure online)



5–10 days after baseline (Time 2), and within 3–6 weeks after baseline (Time 4) (see Fig. 1). To assess inter-rater reliability, a second evaluator administered the PSFS over the phone within 3 days of the first evaluator administering the PSFS for at Time 2 (Time 3) (see Fig. 1).

All evaluations after baseline were scheduled for within 3 h of the time at which the baseline assessment was taken (e.g., if the baseline assessment was performed at 12 noon, follow-up assessments had to occur between 9 a.m. and 15:00 p.m.). Participants were first asked if they had experienced medical instability over the previous week (e.g., urinary tract infection). If medically unstable over the previous week, participants were contacted every day within that time interval until they were medically stable; if medical instability persisted for that time interval (i.e., 5–10 days and 4–6 weeks after baseline administration), the information was entered as “missing data due to medical instability.” If medical stability was confirmed on history, then the PSFS was administered.

Statistics

Spasm frequency was developed as a 5-level scale (see Appendix). For each level above “no spasm,” the spasm severity was rated on a three-level scale. Combining these

produced a 13-level spasm frequency-severity scale, with levels: 0, 1.1, 1.2, 1.3, 2.1, 2.2, 2.3, 3.1, 3.2, 3.3, 4.1, 4.2, and 4.3.

Both intra-rater and inter-rater reliability were assessed using weighted-kappa statistics for ordinal data. It was determined a priori that a sample size of 50 participants would provide at least 90% power to rule out intra-rater and inter-rater reliability of slight or worse ($\kappa < 0.2$) if the true reliability is substantial ($\kappa > 0.6$), for both the spasm frequency and frequency-severity scales. Intra-evaluator reliability was measured separately for each pair of available visits to assess whether there was any change with proximity (i.e., comparing scores at Times 1 vs. 2, and 1 vs. 4). Data for all participants with at least two assessments were utilized for analysis (whether for intra-rater or inter-rater analysis). The kappas were weighted with ordinal category scores (1, 2, etc.), which was decided a priori.

Finally, available PSFS scores (spasm frequency and frequency-severity) at each cycle from participants who were unable or unwilling to provide scores at all four time points were evaluated and compared (via exact chi-square test) to the scores from those who did provide data at all four time points, to assess whether data were missing at random (i.e., non-informative).

Table 1 Participant characteristics at baseline

Age (years)	44.1 ± 12.3
Time since injury (years)	7.5 ± 5.5
Sex	
Male	17
Female	49
Level of injury and AIS	
C1–C4 AIS A, B, C	15
C5–C8 AIS A, B, C	22
T1–S1 AIS A, B, C	17
AIS D any level	12
Total participants	66

AIS Abbreviated Injury Scale

All statistical analyses were performed using SAS v9.4 (SAS Institute, Cary, NC, USA). Results unless otherwise noted are reported as mean ± SD.

Results

Participants

Figure 2 illustrates the study participant flowchart. Of the 114 participants contacted, $N = 61$ completed the study with data from at least two assessments available for analysis. Participant characteristics at baseline are described in Table 1. Age ranged from 19 to 65 years, median of 35.5 years. Time since injury ranged from 1.1 to 30.3 years, median of 7 years. SCI neurological level of injury and AIS grade are reported according to recommended standards [11]. The number of participants in each PSFS level (PSFS Part 1, levels 0–4) and severity (PSFS Part 2, severity 1–3) is reported for the first PSFS administered at baseline (Table 2). The participant with no spasms on PSFS at baseline (PSFS = 0) had a recent history of spasticity despite not experiencing spasms in the previous week therefore remained included in the study. At baseline, 32 (49%) participants reported that over the previous week, they experienced mild spasms that were induced by stimulation, 17 (27%) reported infrequent full spasms occurring less than once per hour, 12 (18%) reported spasms occurring more than once per hour, and 4 (6%) reported spasms occurring more than 10 times per hour. Of all participants, 23 (35%) reported their spasms as moderate in severity, and 13 (20%) reported their spasms as severe.

Missing data

At baseline (Time 1), all participants completed the PSFS (Time 1, $n = 66$ completed). At Time 2, 5 participants did

Table 2 PSFS participant distribution at baseline assessment

PSFS level (Part 1), Severity (Part 2)	<i>N</i>
0	1
1, 1	22
1, 2	8
1, 3	2
2, 1	5
2, 2	9
2, 3	3
3, 1	2
3, 2	4
3, 3	6
4, 1	0
4, 2	2
4, 3	2
Total participants	66

PSFS Penn Spasm Frequency Scale

not complete the PSFS (Time 2, $n = 61$ completed); at Time 3, 8 participants did not complete the PSFS (Time 3, $n = 58$ completed); and at Time 4, 11 participants did not complete the PSFS (Time 4, $n = 55$ completed). See Fig. 2 for reasons for non-completion. Results were analyzed to determine whether data was missing at random, first comparing each of the spasm frequency variables among those with complete data vs. those with at least one missing in data point. We then did the same for spasm frequency-severity variables. Comparisons were made with exact chi-square test. None were statistically significant at $\alpha = 0.05$, which supports the hypothesis that data were missing at random (i.e., non-informative missing data). Therefore, all available participant data were included in the analysis.

Intra-rater reliability

Intra-rater reliability within 5–10 days after baseline and within 4–6 weeks after baseline was assessed with a weighted kappa comparing the time points “Time 1” vs. “Time 2,” and “Time 1” vs. “Time 4,” respectively. See Table 3 for results for analysis of the PSFS Part 1 (5-level spasm frequency) and with the addition of the PSFS Part 2 (13-level spasm frequency-severity combination).

Inter-rater reliability

Inter-rater reliability was assessed with a weighted kappa comparing the time points “Time 2” and “Time 3.” See Table 3 for results of analysis for the PSFS Part 1 and with the addition of the PSFS Part 2. Analysis of these time points for intra-rater and inter-rater reliability were determined a priori.

Table 3 Inter-rater and intra-rater reliability for Part 1 and combined Part 1 and 2 of the Penn Spasm Frequency Scale

Intra-rater reliability	Time 1 vs. Time 2	Time 1 vs. Time 4
5-level spasm frequency	0.822 (0.709, 0.935)	0.734 (0.586, 0.883)
13-level spasm frequency-severity combination	0.812 (0.705, 0.919)	0.729 (0.586, 0.872)
Inter-rater reliability	Time 2 vs. Time 3	
5-level spasm frequency	0.862 (0.759, 0.965)	
13-level spasm frequency-severity combination	0.857 (0.762, 0.952)	

Note: Values are in weighted kappa (95% CI)

CI confidence interval

Discussion

This is the first study to establish intra-rater and inter-rater reliability values for the PSFS in people with chronic traumatic SCI. What is additionally unique in this study is that there was analysis for both Part 1 and Part 2 of the PSFS, thus allowing researchers to utilize the entire PSFS for future studies. Previous studies have primarily limited use of the PSFS to Part 1 [12, 13], which means that the severity of the spasms affecting the individual, an important aspect of the spasticity experience, is not captured.

There are no universally agreed upon cut-points for kappa classification. A general classification that is well accepted is kappa = 0, poor; 0.01–0.20, slight, 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–0.99, almost perfect [14]. Using this classification, almost all our weighted kappas for both intra-rater and inter-rater reliability with analysis of PSFS Part 1 alone (spasm frequency) or with inclusion of PSFS Part 2 (spasm frequency-severity combination) of the PSFS are within the “almost perfect” category, with lower confidence bounds in the “substantial” range. The exception was the secondary intra-rater kappa comparing Time 1 vs. Time 4, which was lower, but still “substantial,” with a lower confidence bound in the “moderate” range.

Spasticity following SCI can change over time either in the short term as a result of medical instability or over the long term due to natural history. A large prospective cohort analysis of spasticity in individuals with traumatic SCI over 5 years post injury demonstrated that the natural history of problematic spasticity, as defined by spasticity requiring treatment, is that it tends to stabilize after approximately 1 year post injury [2]. Therefore, these results are applicable only to the chronic (>1 year post injury) traumatic SCI population.

Study limitations

The presence of spasticity as an inclusion criteria was determined on history and not by physical exam; however,

spasticity in people with SCI fluctuates within the day and from day to day, therefore objective observation of the absence of spasticity at one time point does not exclude the presence of spasticity in that individual. Also, the assessment of intra-rater reliability was based upon assessments conducted at different times, and so this reliability calculation assumes that the frequency and severity of the spasticity remained largely unchanged over the 6 weeks past the initial baseline assessment. In the end, the very high kappa statistic would suggest that this was indeed the case. The PSFS was administered over the phone, which could be considered a strength of the study as it demonstrates reliability using a methodology with lower participant burden, but results may vary if the PSFS is administered in person. There was a high attrition rate, possibly in part related to the telephone contact of the participants. This study reflects findings in individuals with SCI that have received their post-SCI rehabilitation at GF Strong Rehab Center, and may not be generalizable to individuals with SCI who have undergone rehabilitation in other centers.

Conclusion

Results of this study demonstrate that the PSFS has adequate intra-rater and inter-rater reliability in the time window that is applicable to the assessment of treatment response to oral medications (5–10 days post baseline) and intramuscular botulinum toxin (4–6 weeks post baseline) in people with chronic traumatic SCI. Further research is needed to determine additional psychometric variables of the PSFS including: validity, minimal detectable change, and minimally clinically important difference.

Acknowledgements We thank the following individuals for their contributions: Farhana Dossa, MSc, Carlen K Fung, RN, and Alvin Ip, MD, for assistance with participant recruitment and data collection; Allan Aludino of the Vancouver Spine Research Program for database management and data abstraction; Eric C Sayre, PhD, for statistical analysis.

Funding Dr. Kwon is the Canada Research Chair in Spinal Cord Injury and received financial support from the VGH & UBC Hospital Foundation during conduction of this research.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Pandyan A, Gregoric M, Barnes M, Wood D, Van Wijck F, Burrige J. et al. Spasticity: clinical perceptions, neurological realities and meaningful measurement. *Disabil Rehabil.* 2005;27:2–6. <https://doi.org/10.1080/09638280400014576>.
- Holtz KA, Lipson R, Noonan VK, Kwon BK, Mills PB. Prevalence and effect of problematic spasticity following traumatic spinal cord injury. *Arch Phys Med Rehabil.* 2016. <https://doi.org/10.1016/j.apmr.2016.09.124>.
- Phadke CP, Balasubramanian CK, Ismail F, Boulias C. Revisiting physiologic and psychologic triggers that increase spasticity. *Am J Phys Med Rehabil.* 2013;92:357–69. <http://www.ncbi.nlm.nih.gov/pubmed/23620900>. Accessed 22 July 2014.
- Lechner H, Frotzler A, Eser P. Relationship between self- and clinically rated spasticity in spinal cord injury. *Arch Phys Med Rehabil.* 2006;87:15–9. <https://doi.org/10.1016/j.apmr.2005.07.312>.
- 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;77:713–6. <http://www.ncbi.nlm.nih.gov/pubmed/8670001>. Accessed 22 July 2014.
- Adams MM, Ginis KAM, Hicks AL. The spinal cord injury spasticity evaluation tool: development and evaluation. *Arch Phys Med Rehabil.* 2007;88:1185–92. <https://doi.org/10.1016/j.apmr.2007.06.012>.
- Hsieh JTC, Wolfe DL, Miller WC, Curt A. Spasticity outcome measures in spinal cord injury: psychometric properties and clinical utility. *Spinal Cord.* 2008;46:86–95. <https://doi.org/10.1038/sj.sc.3102125>.
- Kita M, Goodkin DE. Drugs used to treat spasticity. *Drugs.* 2000;59:487–95.
- Hamjian JA, Walker FO. Serial neurophysiological studies of intramuscular botulinum-A toxin in humans. *Muscle Nerve.* 1994;17:1385–92. <https://doi.org/10.1016/j.apmr.2007.06.012>.
- Kirshblum SC, Waring W, Biering-Sorensen F, Burns SP, Johansen M, Schmidt-Read M. et al. Reference for the 2011 revision of the International Standards for Neurological Classification of Spinal Cord Injury. *J Spinal Cord Med.* 2011;34:547–54. <https://doi.org/10.1179/107902611X13186000420242>.
- Biering-Sørensen F, DeVivo MJ, Charlifue S, Chen Y, New PW, Noonan V. et al. International Spinal Cord Injury Core Data Set (version 2.0)—including standardization of reporting. *Spinal Cord.* 2017;55:759–64. <https://doi.org/10.1038/sc.2017.59>.
- Guillaume D, Van Havenbergh A, Vloeberghs M, Vidal J, Roeste G. A clinical study of intrathecal baclofen using a programmable pump for intractable spasticity. *Arch Phys Med Rehabil.* 2005;86:2165–71. <https://doi.org/10.1016/j.apmr.2005.05.018>.
- Hornby TG, Rymer WZ, Benz EN, Schmit BD. Windup of flexion reflexes in chronic human spinal cord injury: a marker for neuronal plateau potentials?. *J Neurophysiol.* 2003;89:416–26. <https://doi.org/10.1152/jn.00979.2001>.
- Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med.* 2005;37:360–3. <http://www.ncbi.nlm.nih.gov/pubmed/15883903>. Accessed 8 Sept 2017