Spinal Cord (2000) 38, 675–679

www.nature.com/sc

Inter-rater reliability of the 1992 international standards for neurological and functional classification of incomplete spinal cord injury

M Jonsson*[,1], A Tollbäck[1], H Gonzales[2] and J Borg[3]

[1]Department of Physical Therapy, Karolinska Hospital, Stockholm, Sweden; [2]Department of Neurology, Karolinska Hospital, Stockholm, Sweden; [3]Rehabilitation Medicine, Danderyd Hospital, Stockholm, Sweden

Aims: To determine the inter-rater reliability in scoring sensory and motor function and in defining sensory and motor levels in incomplete spinal cord injury, using the revised 1992 International Standards for Neurological and Functional Classification of Spinal Cord Injury (ISCSCI-92) and to determine the effect on raters agreement of one standardising assessment.
Methods: Two physicians and two physiotherapists at the Spinal Cord Injury Unit, Karolinska Hospital, classified 23 patients according to the ISCSCI-92. Kappa values were calculated.
Results: Kappa values varied from 0 to 0.83 (poor to very good) for the pin-prick scores, from 0 to 1 for the light touch scores and from 0 to 0.89 for motor function after the standardising assessment. Kappa values for sensory and motor levels were fair to poor after the standardising assessment. The results showed improvement in degree of agreement in 35/46 dermatomes for scoring pin-prick, in 15/42 for light touch, in 14/19 segments for motor function and for three out of four sensory and motor levels.
Conclusion: This study indicates a weak inter-rater reliability for scoring incomplete SCI lesions using the ISCSCI-92.
Spinal Cord (2000) **38,** 675–679

Keywords: classification; spinal cord injury; measurement; reliability

Introduction

Reliable assessment of sensory and motor functions in spinal cord injury (SCI) patients is crucial for clinical management and research – in the acute stage for adequate treatment and prediction of outcome and at follow up examinations for detection of progression of symptoms due to, eg, posttraumatic complications.[1,2] During the last decades several systems on how to accurately classify the severity of a spinal cord lesion have been introduced.[3–6] Ota and co-workers suggested a five grading scale for determination of motor scores combined with a functional scale for the optimal classification of the SCI patient.[7] To improve accurate communication between clinicians and researchers working with SCI patients, the American Spinal Injury Association (ASIA) developed standards for neurological and functional classification of SCI published in 1982,[8] later revised according to data from evaluation studies.[9,10] The third revised version was accepted and recommended by the International Medical Society of Paraplegia (IMSOP) in 1992 and has been established as the International Standards for Neurological and Functional Classification of Spinal Cord Injury (ISCSCI-92). In 1994 Ditunno *et al* published the 'International standards booklet for neurological and functional classification of spinal cord injury', which has been widely distributed.[11] A training package including four videotapes and a reference manual upon the 1992 version have been published.[12]

The classification in use has to be valid and, further, to be reliable between and within raters to fulfil the purpose of determination of the exact sensory and motor levels during both the acute stage and the long-term follow-up examinations. The reliability is a matter of both examination and classification skills. After changes made in the major revision in 1992, reliability of the examination procedures was found to be excellent, while there were still discrepancies in the classification of injury.[13,14] The reference manual published in 1994 includes a study of reliability of the 1992 Standards. It shows very good reliability for total pinprick and motor scores and disparities determining light touch scores, sensory and motor levels.[12] El-Masry *et al* reported that both the motor scores of the Standards and the National acute spinal cord injury study (NASCIS) motor scores were valid for evaluation of the motor deficit in SCI patients.[15]

The reliability was further addressed in a study by Cohen *et al* published in 1998 where the effect of training in a conference setting on the skills of classification was tested using the ISCSCI-92 classifi-

*Correspondence: M Jonsson, Department of Physiotherapy R1:07, Karolinska Hospital, S-17176 Stockholm, Sweden

cation and the training system used for this system.[16] Preliminary data from this setting, showing classification problems for motor levels when muscles of grade 4 are present, caused a change in defining motor levels in the 1996 revision. Required strength of the muscle above a key muscle graded 3 was '5' instead of '4' or '5'.[16] Further, the study highlighted that classification of incomplete injuries is 'problematical in many areas', in contrast to classification of complete injuries.

Compared to the US and Australia, Sweden has a relatively low annual incidence of SCI, with 13 traumatically injured persons/million inhabitants and about the same number of non-traumatic spinal cord disorders treated in spinal units. The majority of the SCI patients are treated in SCI units in either of the six university hospitals using a standardised protocol for assessment of SCI patients since January 1997.[17] The protocol includes the ISCSCI-92.

In a significant proportion of the traumatic SCI patients admitted to SCI units in Sweden the injury is incomplete, which is in contrast to, eg, the North American Model SCI Centres.[18] According to the Swedish National Board of Health and Welfare registry data for 1997 and 1998, 107 of 201 traumatic SCI patients treated in SCI units were classified as having incomplete lesions. In Stockholm approximately 50 SCI patients, with traumatically or benign spinal cord disorders are treated each year. The SCI unit at the Karolinska Hospital, Stockholm uses written guidelines for the treatment of SCI patients.[19] The first assessment, in the emergency setting, is performed by the neurosurgeon in charge, continued assessments are then performed by either a SCI unit physiotherapist or physician.

The aim of the study

The aim of this study was to determine the inter-rater reliability of ISCSCI-92 regarding sensory and motor scores and sensory and motor levels among mainly incomplete SCI patients. An additional aim was to see whether reliability improved after one standardising assessment.

Methods

Raters

Four team members at our SCI unit using ISCSCI-92, two physicians and two physiotherapists, took part in the study. Three had more than 10 years of experience from SCI patients and had used the Standards for several years. One physician had 1 year of experience from SCI patients and the 1992 Standards. All were updated on the 1996 revision of the Standards.

Patients

Twenty-three patients, eight women and 15 men, in-patients for the time of the study, were classified; 12 with a cervical lesion, six with a thoracic lesion and five with a lumbar lesion. Seventeen were traumatic and six non-traumatic. Three were classified as complete and 20 as incomplete lesions according to the ordinary protocol, to which the raters were blinded. Five of six non-traumatic lesions were classified as incomplete.

Inclusion criteria for the patients to be classified were spinal cord injury verified by magnetic resonance tomography (MRT), more than 1 week from onset of the SCI and ability to adequately communicate. Malignant or other progressive lesion was an exclusion criteria. All patients gave their informed consent to participate in the study.

Ethic consideration

As the classification is a regular examination of the SCI patients this study was considered by the Ethic committee at the Karolinska Hospital as a clinical quality assurance work and therefore did not need their approval.

Procedure

The classification of each patient was performed by all the raters on the same day, from 1 to 14 (mean 7.8) weeks after the injury date. The patients were numbered 1:1 to 1:12 for the group before the standardising assessment and 2:1 to 2:12 for the group after the standardising assessment, the physicians Ph 1 and Ph 2 and the physiotherapists PT 1 and PT 2. A time schedule was set up for the raters and the patients. The raters classified the patients according to the schedule in the same rotating order so that the first rater was the first to classify patient no. 1, the fourth to classify patient no. 2, the third to classify patient no. 3, the second to classify patient no. 4 and again the first to classify patient no. 5 and so on. This was done to minimise influence of systematic errors. The patients were instructed not to discuss the results from the previous raters during the ongoing study.

Data to be filled into the protocol were: sensory scores including light touch and pin prick, total sensory scores, motor score for each segment, total motor score, anal sensation, sensory and motor levels for right and left sides. The second rater, Ph 2, was the only one to examine voluntary anal contraction due to ethic consideration. Criteria not included in the study were completeness of the injury, zone of partial preservation, ASIA impairment scale, clinical syndromes and FIM. The protocols were to be handed over to first the author after each completed classification.

On the day before the first classifications took place the first author gave a 1-h information session to the four raters about the purpose of the study and gave a brief review of the reference manual for the ISCSCI-92. Further the raters were given the opportunity to discuss matters that were unclear. The raters were specifically advised to consider the latest version of the ISCSCI-92.[20] In order to improve the degree of

agreement, a standardising assessment was performed by all the raters and the first author together. It took place after the raters had classified 12 patients. It included a bedside assessment of one patient with the opportunity to discuss disparities between raters. This patient was not included in the study. Thereafter a written consensus on how to determine sensory and motor levels was distributed to the raters. In accordance with the 1996 revision a score of 5 was used as normal motor function.[20]

Statistics

Kappa was calculated to determine the degree of agreement between raters. Kappa is the ratio of the proportion of times that the raters agree (corrected for chance agreement) and the maximum proportion of times that the raters could agree (corrected for chance agreement). A value of 1.00 indicates 'a perfect agreement', 0 indicates 'no better agreement than chance', and a negative value indicates 'worse than chance' agreement.[21]

According to the guidelines presented by Altmann,[22] the strength of agreement is interpreted as: 0.81 – 1.00 very good, 0.61 – 0.80 good, 0.41 – 0.60 moderate, 0.21 – 0.40 fair, and <0.20 poor. In general, an acceptable degree of reliability should reach a degree of agreement that is good or very good.

Results

The results are shown in Table 1, presenting the Kappa values for each of the study parameters. By definition Kappa coefficients can not be calculated when the ratings in all patients are the same for all raters, ie when there were similar values, eg, for pin prick in one dermatome, for all individuals over all raters. In Table 1 this is marked by an asterix(*) and described as no variation in the table legend.

Pin-prick scores

The agreement for pin prick scores was very good to good in 4/46 dermatomes before the standardising assessment and in 22/50 dermatomes after, moderate in 25/46 before and in 19/50 after, fair to poor in 17/46 before and in 9/50 after the standardising assessment. There was no variation in 10 dermatomes before and in

Table 1 Kappa value for each rated segment and dermatome before and after a standardising assessment rated by four raters in 12 patients before and 11 patients after the assessment

	Pin prick				*Light touch*				*Motor*			
	Before		*After*		*Before*		*After*		*Before*		*After*	
Level	*Right*	*Left*	*Right*	*Left*	*Right*	*Left*	*Right*	*Left*	*Right*	*Left*	*Right*	*Left*
C2	*	*	*	*	*	*	*	*				
C3	*	*	*	*	*	*	*	*				
C4	*	*	*	*	*	*	**0.65**	*				
C5	*	*	**0.64**	0.14	*	*	**1.**	*	0.3	*	0.57	**0.75**
C6	*	0.09	0.57	0.06	*	0.27	0.44	0.23	0.33	0.46	**0.67**	**0.87**
C7	0.47	0.60	0.52	0.48	0.77	0.54	0.48	0.59	0.87	0.46	0.47	0.46
C8	0.33	0.38	0.59	**0.76**	**0.91**	0.39	0.57	**0.72**	**0.82**	**0.63**	0.57	**0.65**
T1	0.14	0.02	0.31	0.36	*	0.45	**0.61**	0.59	**0.76**	**0.86**	**0.62**	0.47
T2	0.08	0.28	0.33	0.49	*	*	0.37	0.58				
T3	0.33	*	0.31	**0.62**	*	*	**0.79**	**0.72**				
T4	0.12	0.24	**0.69**	**0.76**	**0.64**	0.17	**0.77**	**0.77**				
T5	0.13	0.32	0.59	**0.74**	0.4	0.44	**0.74**	**0.81**				
T6	0.32	0.42	**0.69**	**0.81**	**0.62**	0.59	**0.67**	**0.81**				
T7	0.41	0.52	**0.69**	**0.74**	**0.74**	**0.81**	**0.73**	**0.93**				
T8	0.44	0.59	**0.70**	**0.79**	**0.83**	**0.73**	**0.67**	**0.86**				
T9	0.47	0.48	0.60	**0.83**	**0.88**	**0.79**	**0.76**	**0.78**				
T10	0.47	**0.61**	**0.61**	**0.76**	**0.91**	**0.79**	**0.67**	**0.68**				
T11	0.47	0.48	**0.61**	**0.72**	**0.91**	**0.69**	0.55	0.46				
T12	**0.67**	0.41	0.57	0.51	**0.73**	0.60	0.40	0.44				
L1	0.44	0.53	**0.71**	0.56	**0.77**	0.60	0.60	0.55				
L2	0.55	0.52	**0.61**	**0.66**	**0.81**	**0.74**	**0.66**	**0.61**	0.54	0.48	0.56	**0.63**
L3	0.39	0.58	**0.65**	0.52	**0.79**	**0.74**	**0.68**	0.57	**0.7**	0.55	**0.7**	**0.72**
L4	0.34	0.34	0.52	0.58	**0.67**	0.53	0.43	0.50	0.53	0.49	**0.89**	**0.69**
L5	0.45	0.46	**0.62**	0.56	0.36	0.55	0.49	0.46	0.51	0.53	**0.66**	0.52
S1	0.49	0.52	0.55	0.38	0.46	0.51	0.49	0.30	**0.61**	0.55	**0.64**	**0.69**
S2	0.40	0.50	0.43	0.38	**0.76**	**0.66**	0.56	0.47				
S3	0.60	**0.69**	0.31	0.44	**0.61**	**0.71**	**0.74**	0.54				
S4 – 5	0.54	**0.63**	0.44	0.39	0.54	**0.61**	0.59	0.51				

[a]Dermatomes and segments for which Kappa not could be calculated due to high agreement. Values in bold typeface shows good to very good agreement

six after the assessment. Fifteen of these non-calculated dermatomes were in the C2 to C6 region (Table 1).

Light-touch scores
The agreement for light touch scores was very good to good in 25/42 dermatomes before the standardising assessment and in 24/50 after, moderate in 12/42 before and in 22/50 after, fair to poor in 5/42 before and in 4/50 after the standardising assessment. There was no variation in 14 dermatomes before and in six after the standardising assessment. Fifteen of the non-calculated dermatomes were in the C2 to C6 region (Table 1).

Motor scores
The agreement for motor scores was very good to good in 7/19 segments before the standardising assessment and in 13/20 after, moderate in 10/19 before and in 7/20 after, fair to poor in 2/19 before and in 0/20 after the standardising assessment. There was no variation in one segment before the standardising assessment (Table 1).

Neurological levels
The results showed that the agreement between raters classifying sensory and motor levels was poor in all four levels before and fair in three levels and poor in one level after the standardising assessment. The observed agreement was seen in 2/48 levels before and 6/44 levels after the standardising assessment (Table 2).

Discussion

This study was performed in order to evaluate the inter-rater reliability of the ISCSCI-92 in scoring sensory and motor function and in defining sensory and motor levels in patients with incomplete SCI. The study design differs from previous studies using no golden standard but comparing different raters degree of agreement and including mainly incomplete SCI patients. The study design means that both examination and classification skills might influence the results.

The agreement, according to Altmann,[22] for pin prick scoring was acceptable in only 26/96 calculated dermatomes and for light touch scoring in 49/92 calculated dermatomes (Table 1). Thus, determination of sensory functions showed higher agreement for light touch than for pin prick scoring before the standardising assessment and an improvement in pin prick scoring after. This is in contrast to a previous study presented in the Reference manual where pin prick agreement was reported to be stronger than for light touch.[12] The raters in the present study noted that it was more difficult to determine absence or impairment in pin prick sensation than in light touch sensation. This problem was specifically addressed during the standardising assessment, which might explain the observed improvement for pin-prick scoring. Probably the discrepancies observed, reflect both raters skills and the problems inherent in grading sensation in the clinical setting. It is not clear how much further training and standardised testing conditions might reduce the discrepancies in this area.

The agreement in determination of motor scores was better than for sensory scores and improved after the standardising assessment. In contrast to the reliability study presented in the Reference manual, showing excellent agreement for motor scores we included scores of 4 and 5, while these were excluded in the reliability analysis in that study.[12] Even if some of the discrepancies observed in our study probably reflect a need for more training, it must also be pointed out that the instructions used to determine the scores of 4 and 5 might probably need some further refinement. Grade 4 is described in the Reference manual, to our understanding, as an active movement against some resistance through a full range of movement (ROM) (or maximum available ROM). According to the figure texts illustrating testing positions concerning the scores 4 and 5, the instructions do not clarify whether the raters resistance should be based on concentric, isometric or eccentric muscle strength and the testing position for grade 4 and 5 are illustrated with figures in only part of ROM. Further the ASIA *ad hoc* Committee recommends that a so called 'break test' is not to be used to differentiate a score of 5 from a score of 4, a recommendation which interferes with some of the figure texts in the reference manual.[12]

The degrees of agreement for sensory and motor levels in the present study were lower than those previously presented in the Reference manual[12] and by Donovan upon the 1992 version of the ASIA Standards.[9] Thus, according to the Reference manual it was good for one level, moderate for two and fair for one and according to Donovan good for one level and fair for three levels, while in the present study it was fair to poor. Kappa was used to calculate the agreement for sensory and motor levels in all these studies. Another study made by Cohen *et al* used percentage of correct classification which for determi-

Table 2 Kappa value for neurological levels before and after a standardising assessment rated by four raters in 12 patients before and 11 patients after the assessment

Sensory level right		*Sensory level left*		*Motor level right*		*Motor level left*	
Before	*After*	*Before*	*After*	*Before*	*After*	*Before*	*After*
0.12	0.23	0.07	0.24	0.14	0.14	0.11	0.25

nation of sensory and motor levels ranged from 21% to 97% after a training session, with no improvement after the training session.[16] The main reason for the observed lack of agreement in the definition of motor and sensory levels in our study was the disparity in the sensory scoring, which also has an impact on the motor level classification in the thoracic levels.

Conclusion

This study indicates limited inter-rater reliability of the revised ISCSCI-92 for scoring sensory and motor function and in defining sensory and motor levels in incomplete SCI. We suggest that further efforts to develop training programs as well as the reference manual and computer based classification aids[16] are important in order to make the standards as valuable in assessing incomplete SCI as for complete SCI.

Acknowledgements

This study was supported by the Committee for the Health and Caring Sciences (NVV), the Swedish Association of registered Physiotherapists and the Karolinska Hospital. PT Mats Sternhag and Anna-Karin Åstrand at the Karolinska Hospital SCI are gratefully acknowledged for valuable assistance.

References

1 Maaron JC, Abla AA. Classification of acute spinal cord injury, neurological evaluation, and neurosurgical considerations. *Crit Care Clin* 1987; **3:** 655–677.
2 Heinemann W *et al*. Prediction of Rehabilitation outcomes with disability measures. *Arch Phys Med Rehabil* 1994; **75:** 133–143.
3 Bracken MB, Webb SB, Wagner FC. Classification of the severity of acute spinal cord injury: implications for management. *Paraplegia* 1977–78; **15:** 319–326.
4 Lucas JT, Ducker TB. Motor classification of spinal cord injuries with mobility, morbidity and recovery indices. *Am Surgeon* 1979; **45:** 151–158.
5 Roaf R. International classification of spinal injuries. *Paraplegia* 1972; **10:** 78–84.
6 Frankel HL *et al*. The value of postural reduction in the initial management of closed injuries of the spine with paraplegia or tetraplegia. *Paraplegia* 1969; **7:** 179–192.
7 Ota T *et al*. Functional assessment of patients with spinal cord injury: measured by the motor score and the Functional Independence Measure. *Spinal Cord* 1996; **34:** 531–535.
8 American Spinal Injury Association. Standards for neurological classifications of spinal injured patients. ASIA. Chicago 1982.
9 Donovan WH *et al*. A test of the ASIA guidelines for Classification of Spinal Cord Injuries. *J Neuro Rehab* 1990; **4:** 39–53.
10 Priebe MM, Waring WP. The interobserver reliability of the revised American Spinal Injury Association standards for neurological classification of spinal injury patients. *Am J Phys Med Rehab* 1991; **70:** 268–270.
11 Ditunno JF, Young W, Donovan WH, Creasy G. The International Standards Booklet for Neurological and Functional Classification of Spinal Cord Injury. *Paraplegia* 1994; **32:** 70–80.
12 Reference manual and videotapes for the international standards for neurological and functional classification of spinal cord injury. American Spinal Injury Association. Chicago, Illinois. 1994.
13 Cohen ME, Bartko JJ. Reliability of the ISCSCI-92. In: Ditunno JF, Donovan WH, Maynard FM (eds). *Reference manual for the international standards for neurological classification of spinal cord injury 1994*. ASIA, Chicago.
14 Cohen ME, Sheehan TP, Herbison GJ. Content validity and reliability of the international standards for neurological classification of spinal cord injury. *Topics Spinal Cord Injury Rehab* 1996; **4:** 15–31.
15 El Masry WS *et al*. Validation of the American Spinal Injury Association (ASIA) motor score and the National Acute Spinal Cord Injury Study (NASCIS) Motor Score. *Spine* 1996; **21:** 614–619.
16 Cohen ME, Ditunno Jr JF, Donovan WH, Maynard Jr FM. A test of the 1992 International Standards for Neurological and Functional Classification of Spinal Cord Injury. *Spinal Cord* 1998; **36:** 554–560.
17 Kvalitetsindikatorer i ryggmärgsskadevård. Socialstyrelsen, Stockholm 1995. (in Swedish).
18 Marino RJ, Ditunno JF, Donovan WH, Maynard F. Neurological recovery after traumatic spinal cord injury: Data from the model spinal cord injury systems. *Arch Phys Med Rehabil* 1999; **80:** 1391–1396.
19 Vårdprogram för Stockholmsregionen. Traumatiska hjärn-och ryggmärgsskador, prevention, akut behandling och rehabilitering. Institutionen för klinisk neurovetenskap, Karolinska Institutet, Stockholm 1996. (in Swedish).
20 Maynard Jr FM *et al*. International Standards for Neurological and Functional Classification of Spinal Cord Injury. *Spinal Cord* 1997; **35:** 266–274.
21 Siegel S. *Nonparametric statistics for the behavioural sciences*. 2nd edn. McGraw-Hill book company: Singapore, 1989, pp 284–291.
22 Altmann DG. *Practical statistics for medical research*. Chapmans Hall: London, 1991, pp 403–409.