The International Standards for Neurological Classification of Spinal Cord Injury: Intra-Rater Agreement of Total Motor and Sensory Scores in the Pediatric Population

Ross S. Chafetz, PT, DPT, MPH¹; John P. Gaughan, PhD²; Lawrence C. Vogel, MD³; Randal Betz, MD¹; M.J. Mulcahey, PhD, OTR/L¹

¹Shriners Hospitals for Children, Philadelphia, Pennsylvania; ²Temple University Medical School, Philadelphia, Pennsylvania; ³Shriners Hospitals for Children, Chicago, Illinois

Received March 31, 2008; accepted November 29, 2008

Abstract

Background: The International Standards for Neurological Classification of Spinal Cord Injury (ISCSCI) is the gold standard for evaluating and classifying the neurological consequence of spinal cord injury (SCI).

Objective: To determine the within-rater agreement for total scores of light touch (LT), pin prick (PP), and total motor (TM) in children and youth.

Design: Part of a larger cross-sectional study to determine the intra-rater reliability of the standards when applied to children and youth.

Participants/Methods: A total of 187 subjects participated in 2 repeated examinations performed by the same rater. A total of 7 raters participated in this study. Intraclass correlations coefficients (ICCs), with 95% CI were calculated to determine agreement between the 2 examinations for LT, PP, and TM.

Results: With the exception of subjects younger than 6 years, agreement on repeated total PP, LT, and TM scores were good to excellent, as shown by ICC values of 0.92 or higher. Although agreement was high for the youngest age group for LT (ICC = 0.920), PP (ICC = 0.957), and TM (ICC = 0.971), all of the lower 95% CI values fell well below 0.66, indicating poor precision. All subgroups had good to high agreement for total PP, LT, and TM scores, as indicated by ICC values of 0.87 and higher. There were lower 95% CI (LCI) values for the 6- to 11-year-old group with incomplete paraplegia due to the low number of subjects in that subgroup (N = 4). The LCI values were poor for PP for the subgroups with 6- to 11-year-olds with incomplete tetraplegia (LCI = 0.675) and the 12- to 15-year-old group with incomplete paraplegia (LCI = 0.707) and for TM for 16- to 21-year-old group with complete paraplegia (LCI = 0.706).

Conclusions: In children as young as 6 years, within-rater agreement on LT, PP, and TM exceeded recommended values for clinical measures. With the exception of 6- to 11-year-olds with incomplete injuries, type of injury and severity of injury were not factors in agreement. Although more work is needed to define the lower age limit in which the ISCSCI have utility, these data represent growing evidence supporting the use of the ISCSCI when evaluating the neurological consequence of SCI in children.

J Spinal Cord Med. Apr 2009;32(2):157-161

Key Words: Spinal cord injuries, complete, incomplete; Child; Adolescence; Classification; Tetraplegia; Paraplegia; Physical examination, neurological; Measurement; International Standards for Neurological Classification of Spinal Cord Injury

Please address correspondence to M. J. Mulcahey, PhD, OTR/L, Rehabilitation Services/Research, Shriners Hospitals for Children, 3551 North Broad Street, Philadelphia, PA 19140; p: 215 430 4081; f: 215 430 4141 (e-mail: mmulcahey@shrinersnet.org).

This study was funded by the Shriners Hospitals for Children Research Advisory Grant no. 8956 (M. J. M., primary investigator).

© 2009 by the American Paraplegia Society

INTRODUCTION

The usefulness of measurement in clinical assessment, clinical research, and decision making depends upon the extent to which clinicians can rely on data as accurate and meaningful indicators of a behavior or an attribute (1). Intra-rater reliability refers to the stability of the data recorded by one individual across 2 or more trials. In contrast, inter-rater reliability concerns variation between 2 or more raters. Both intra-rater and inter-rater reliability are important, the former providing an indication of the

stability of the data and the latter providing an indication of the generalizability of data.

Over the last 15 years, there has been much attention on the evaluation and development of the psychometric parameters of the motor and sensory examinations of International Standards for Neurological Classification of Spinal Cord Injury (ISCSCI) (2–6). This attention, which is largely based on the need for standardized and sensitive measurement tools for clinical trials, is evidenced by notable publications (7–12) and by a commitment from ASIA to make ISCSCI training modules available.

Our research program is equally interested in establishing the ISCSCI motor and sensory examinations as reliable methods to evaluate the neurological consequence of spinal cord injury (SCI) in children. We have demonstrated that training improves pediatric clinicians' skills needed for both the ISCSCI examination (13) and classification (14). Our early reliability work suggested that 4 years may be the youngest age for which the neurological examinations have clinical utility; that many children as old as 8 years had difficulty with the pin prick test; and that the anorectal examination was difficult, particularly for young children and those injured prior to being toilet trained (15). We also showed that although intra-rater agreement of summed pin prick (PP), light touch (LT), and total motor (TM) scores of children with SCI was good to high, they had poor precision based on wide CI and lower CI (LCI) values that reflected poor agreement (15). The latter study was a catalyst to a multicenter effort designed to carry out another reliability study in a larger sample of children to evaluate agreement of the ISCSCI as a factor of age, type of injury (paraplegia/tetraplegia), and severity of injury (complete/incomplete). This paper reports the results for intra-rater reliability for total sensory and motor scores.

METHODS

This was a cross-sectional repeated measures multicenter study. The study protocol, consent, assent, and Health Insurance Portability and Accountability Act forms were all reviewed and approved by a research ethics/institutional review boards at each participating center. Written informed consent was obtained from all parents or legal guardians of all subjects younger than 18 years. In accordance with institutional review board guidelines, subjects 7 to 18 years old also provided written informed assent; subjects older than 18 years provided their own written consent. For subjects who spoke Spanish as their primary language, Spanish consent forms were used and an interpreter assisted in the informed consent process and testing procedures. The institutional review board–approved Health Insurance Privacy and Portability Act HIPPA forms were also reviewed with subjects for their consent. We certify that all applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed throughout the duration of this study.

Sample

A convenience sample of patients 3 months to 21 years of age was invited to participate in this study. These participants had chronic SCI (defined as ≥ 3 months after injury with no neurological changes) and no evidence of concomitant brain injury that would interfere with cognitive ability to participate in the examination.

Data Collection

There were 7 raters trained in the examination (13) and classification (14) techniques of the ISCSCI (16) who collectively were responsible for conducting the neurological examinations. Subjects participated in up to 4 repeated neurological examinations, with 2 raters performing 2 examinations each on 4 separate days. Average time between examinations was 2.4 days (± 2.3). Of the 7 raters, 4 had more than 3 years of experience conducting the ISCSCI, and the remaining 3 had less than 1 year of experience (13).

The ISCSCI motor and sensory examinations were performed using the standardized techniques described in the ISCSCI training package (16). To ensure standardization among the raters, a certain protocol was followed. For PP, subjects were scored “2” (normal) if they could consistently distinguish 8 of 10 trials of sharp/dull, “1” (impaired) if sensation was impaired and the subject could discriminate between sharp and dull, and “0” (absent) if there was no appreciation of pinprick testing. For LT, subjects were given a score of “2” if 4 of 5 trials tested as feeling the same as on the check; “1” if there was appreciation of LT, but it was different from when tested on the check; and “0” if there was no appreciation of LT. Muscle strength was evaluated manually according to the guidelines described in the testing manual (16).

Scores were documented immediately throughout the examination by either the examiner or a scribe using an ISCSCI form modified with permission specifically for this investigation. Scoring and classification were performed by the trained raters and reviewed by the senior investigator (M.J.M.) based on the guidelines set forth in the ASIA manual (16). Data were entered into a secure database by 2 independent assistants blinded to the study. Final mathematical calculations for TM and sensory scores were conducted using computer logic specifically designed for this study.

Data Analysis

Data were deidentified and exported to a spreadsheet for analysis by a biostatistician blinded to the study. Data were normalized using a ranked transformation to accommodate for subject/rater nesting and analyzed using SAS software (SAS Institute Inc, Cary, NC). Intra-rater agreement was evaluated by using a nested ICC model and 95% CI for total PP, LT, and TM scores based on age, type of injury (paraplegia/tetraplegia), and severity of injury (complete/incomplete). ICC values greater than 0.90 reflect high agreement, values from

Table 1. Characteristics of Sample by Age, Severity of Injury, and Type of Injury

	Years	
Average time since injury	4.9 ± 4.4	
Average age at time of exam	14.6 ± 4.2	
	Number	Percent
Gender		
Male	110	58.8
Female	77	41.2
Severity		
Complete	92	49.2
Incomplete	95	50.8
Type of injury		
Tetra	99	52.9
Para	88	47.1
ASIA impairment scale (AIS)		
AIS A	92	49.2
AIS B	49	26.2
AIS C	26	13.9
AIS D	19	10.2
Scoring not possible	1	0.5
Age grouping (y)		
0–5	3	2.1
6–11	43	23
12–15	49	26.2
16–21	92	48.7

0.75 to 0.90 reflect moderate agreement, and ICC values less than 0.75 are reflective of poor agreement. The 95% CI provides an indication of precision of the coefficients, whereas a narrow CI reflects good precision and a wide CI reflects weak precision.

RESULTS

A total of 211 subjects consented to participate. A total of 24 subjects from 1 to 8 years of age were excluded from the analysis due to their inability to participate in the examination, leaving 187 subjects with data for the analysis. As summarized in Table 1, the sample included 187 subjects from 4 to 21 years old. There were slightly more subjects with tetraplegia ($N = 99$) than paraplegia ($N = 88$) and relatively equal numbers of subjects with complete ($N = 92$) and incomplete ($N = 95$) injuries. Among participants with incomplete injuries, there was equal distribution between subjects with sensory incomplete injuries (AIS A impairment scale [AIS] B; $N = 49$) and subjects with motor incomplete injuries (AIS C, D; $N = 45$). AIS designation for one subject was not determined due to missing data. Average age of subjects was 14.6 ± 4.2 years with an average of 4.9 ± 4.4 years since injury (Table 1).

Table 2 provides ICC values and 95% CI grouped by age, type of injury (tetraplegia/paraplegia), and severity of injury (complete/incomplete). With the exception of subjects younger than 6 years, agreement on repeated

Table 2. ICC for TM Scores Type of Injury, Age, and Severity of Injury

		N	ICC	95% Confidence Interval
Tetraplegia	Total LT	95	0.965	0.948–0.9770
	Total PP	95	0.947	0.921–0.964
	TM score	95	0.990	0.985–0.993
Paraplegia	Total LT	89	0.980	0.970–0.9870
	Total PP	89	0.975	0.963–0.9841
	TM score	90	0.959	0.939–0.9731
0–5 y	Total LT	NA	NA	NA
	Total PP	3	0.957	0.079–0.999
	TM score	3	0.971	0.272–0.999
6–11 y	Total LT	42	0.979	0.961–0.989
	Total PP	42	0.960	0.928–0.979
	TM score	43	0.984	0.970–0.991
12–15 y	Total LT	49	0.984	0.973–0.9915
	Total PP	49	0.984	0.972–0.9912
	TM score	48	0.995	0.992–0.9976
16–21 y	Total LT	90	0.983	0.974–0.989
	Total PP	90	0.976	0.9640–0.984
	TM score	91	0.993	0.990–0.996
Complete	Total LT	95	0.999	0.999–0.999
	Total PP	95	0.987	0.980–0.991
	TM score	96	0.983	0.976–0.989
Incomplete	Total LT	89	0.968	0.952–0.979
	Total PP	89	0.958	0.937–0.973
	TM score	89	0.986	0.980–0.991

ICC, interclass correlations coefficients; TM, total motor; LT, light touch; PP, pin prick; and NA, not calculated due to insufficient sample.

total PP, LT, and TM scores was good to excellent, as evidenced by ICC values of 0.920 or higher. Although agreement was high for the youngest age group for LT (ICC = 0.920), PP (ICC = 0.957), and TM (ICC = 0.971), all of the lower 95% CI values fell well below 0.660, indicating poor precision.

Table 3 provides ICC and 95% CI values for the sample stratified by type of injury, severity, and age. Due to an insufficient number of subjects, it was not possible to calculate ICCs for the youngest age group (<6 years). All others subgroups had good to high agreement for total PP, LT, and TM scores, as indicated by ICC values of 0.87 and higher. There were low LCI values for the 6- to 11-year-old group with incomplete paraplegia due to the low number of subjects in that subgroup ($N = 4$). The lower CI values were poor for PP for the subgroup of 6- to 11-year-olds with incomplete tetraplegia (LCI = 0.675), for the 12- to 15-year-old group with incomplete paraplegia (LCI = 0.707), and for TM for the 16- to 21-year-olds with complete paraplegia (LCI = 0.706).

DISCUSSION

This study evaluated the intra-rater reliability of the total sensory and motor scores from ISCSCI examinations

Table 3. ICC for Summed PP, LT, and TM Scores Type of Injury, Age, and Severity of Injury

												95% Confidence		
Age (y)					N	ICC	Interval				N	ICC	95% Confidence	
												Coefficient	Interval	
Tetra	0–5	Comp	LT	NA	NA	NA	Para	0–5	Comp	LT	NA	NA	NA	NA
			PP	NA	NA	NA				PP	NA	NA	NA	
			TM	NA	NA	NA				TM	NA	NA	NA	
		Incomp	LT	NA	NA	NA		Incomp	LT	NA	NA	NA		
			PP	NA	NA	NA			PP	NA	NA	NA		
			TM	NA	NA	NA			TM	NA	NA	NA		
	6–11	Comp	LT	12	0.942	0.811–0.982	6–11	Comp	LT	12	0.973	0.911–0.992		
			PP	12	0.957	0.858–0.987			PP	12	0.944	0.820–0.983		
			TM	12	0.999	0.995–0.999			TM	13	0.948	0.841–0.984		
		Incomp	LT	14	0.946	0.839–0.982		Incomp	LT	4	0.882	0.020–0.992		
			PP	14	0.883	0.675–0.960			PP	4	0.959	0.514–0.997		
			TM	14	0.994	0.980–0.997			TM	4	0.919	0.213–0.994		
Tetra	12–15	Comp	LT	8	0.948	0.763–0.989	Para	12–15	Comp	LT	17	0.969	0.919–0.989	
			PP	8	0.956	0.797–0.991				PP	17	0.987	0.965–0.995	
			TM	8	0.998	0.991–0.999				TM	17	NA	NA	
		Incomp	LT	9	0.964	0.851–0.991		Incomp	LT	8	0.963	0.895–0.987		
			PP	9	0.985	0.933–0.996			PP	15	0.891	0.707–0.962		
			TM		0.988	0.941–0.997			TM	15	0.974	0.926–0.991		
	16–21	Comp	LT	17	0.977	0.936–0.991	16–21	Comp	LT	26	0.983	0.964–0.992		
			PP	17	0.980	0.945–0.992			PP	26	0.973	0.940–0.987		
			TM	17	0.995	0.984–0.998			TM	26	0.856	0.706–0.933		
		Incomp	LT	33	0.959	0.919–0.979		Incomp	LT	14	0.965	0.896–0.988		
			PP	33	0.942	0.886–0.970			PP	14	0.971	0.913–0.990		
			TM	34	0.987	0.974–0.993			TM	14	0.984	0.953–0.995		

ICC, interclass correlations coefficients; PP, pin prick; LT, light touch; TM, total motor; tetra, tetraplegia; comp, complete; NA, not calculated due to insufficient sample; para, paraplegia, and incomp, incomplete.

administered by 7 raters to 187 children from 4 to 21 years old with chronic SCI. Overall, the results of this study demonstrated a high agreement (ICCs ≥ 0.094) in total PP, LT, and TM scores. These results support previous work in adults with SCI that suggested good to excellent total scores for LT, PP, and muscle strength.

In adults, Cohen and Bartko conducted a reliability study as part of the GM-1 clinical trials (6). Their study examined the intra-rater and inter-rater reliability for 32 patients who were examined by 29 raters from 9 institutions. All 32 patients participated in the inter-rater reliability, and 11 of those returned on a second separate day to participate in the intra-rater reliability study. Similar to our study, their intra-rater reliability for LT, PP, and TM was excellent, that is, all scored higher than 0.98. The authors did not report 95% CI, and their limited sample size makes it difficult to generalize findings to other populations.

In our previous study with 78 children, we demonstrated good to excellent reliability (ICC > 0.98 ; CI > 0.98) in sensory scores regardless of age, type of injury, or severity. There was poorer precision for motor scores (ICC 0.711; 95% CI 0.226–0.892), particularly in patients younger than 10 years. This finding differs from our current work in which TM scores had high reliability and precision in patients older than 6 years, regardless of

severity or type of injury. The reduced width of the CI and, hence, better precision in this study are most likely a product of the larger sample size; the study by Mulcahey et al had only 18 subjects in the 6- to 11-year-old group, in contrast to this study, which had 43 subjects in that age group (15).

Unlike all previous studies, due to our large sample size, we were able to group subjects, as well as stratify by severity, age, and type of injury. When further stratifying, our reliability and precision, for the majority of comparisons, remained good to excellent (>0.90). However, although the ICC values were high, there was a widening of the lower range of the 95% CI, particularly in the 6- to 11-year-old group with incomplete injuries. In patients aged 6 to 11 years with incomplete with tetraplegia or paraplegia, ICC values were higher than 0.900 but 95% CI for PP, LT, and TM. The low numbers ($n = 14$ and $n = 4$, respectively) in these age groups are likely the cause of the wide 95% CI. For the 0- to 5-year age group, calculations of ICC were not achieved due to an inadequate number of children. There are 2 plausible explanations for small numbers in these subgroups. First, as discussed by Mulcahey et al, it is possible that the distribution of pediatric injury may be such that there are fewer incomplete injuries in children younger than 10 years (15,17). It has been suggested that this age group

sustains lap belt injuries and SCI without radiographic abnormalities, both of which have been associated with a higher incidence of complete injuries (18,19). The second reason for the low number of subjects in the lower age groups is the inability of children to complete the ISCSCI examinations. In this study, 24 subjects younger than 10 years were not testable due to their age. In several cases, subjects were unable to understand the concept of “same or different during PP or LT testing to the face” or did not fully understand the difference between “normal” and “abnormal.” Without the cognitive ability to understand these concepts, we could not administer the examinations. Further analysis into subjects classified as “not testable” is needed.

A limitation of this study is that the testing was completed during the same week, typically over 4 days. When possible, testers alternated days between each examination to maximize time between examinations. For this study, there was an average of 2.4 days between a rater’s first and second examination. Having the examinations so close together may subject the results of this study to tester-recall bias.

CONCLUSIONS

These results support the use of the motor and sensory examinations in children 5 years of age and older who are able to follow the test instructions when PP and LT are applied to their cheek. High intra-rater reliability was maintained when patients were stratified by age, level of injury (tetraplegia/paraplegia), and severity of injury. Strong intra-rater reliability provides preliminary support for the use of the ISCSCI motor and sensory exams for clinical care in pediatrics and builds on the growing evidence for its use in future pediatric clinical trials.

REFERENCES

- Portney LG, Watkins MP. Statistical measures of reliability. In: *Foundations of Clinical Research: Applications to Practice*. Upper Saddle River, NJ: Prentice Hall; 2008:505–527.
- Mulcahey MJ, Gaughan J, Chafetz R, Betz R, Vogel LC. Testing a limited number of dermatomes (ASIAQUICK) as predictor of the 56 dermatome score [abstract #P25]. *J Spinal Cord Med*. 2007;30(2):408.
- Savic G, Bergstrom EM, Frankel HL, Jamous MA, Jones PW. Inter-rater reliability of motor and sensory examinations performed according to American Spinal Injury Association standards. *Spinal Cord*. 2007;45(6):444–451. epub 2007 Mar 27.
- Marino RJ, Jones LM, Kirshblum SC, Tal J. Reliability of the ASIA motor and sensory examination [abstract #P43]. *J Spinal Cord Med*. 2004;27(2):194.
- 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 Rehabil*. 1991;70(5):268–270.
- Cohen ME, Bartko J. Reliability of ISCSCI-92 for neurological classification of spinal cord injury. In: American Spinal Injury Association and International Medical Society of Paraplegia, eds. *Reference Manual for the Standards for Neurological and Functional Classification of Spinal Cord Injury*. Chicago, IL: American Spinal Injury Association; 1994:59–65.
- Bracken MB, Shepard MJ, Holford TR, et al. Administration of methylprednisolone for 24 or 48 hours or tirilazad mesylate for 48 hours in the treatment of acute spinal cord injury: results of the Third National Acute Spinal Cord Injury Randomized Controlled Trial. National Acute Spinal Cord Injury Study. *JAMA*. 1997;277(20):1597–1604.
- Bracken MB, Shepard MJ, Collins WF, et al. A randomized, controlled trial of methylprednisolone or naloxone in the treatment of acute spinal-cord injury: results of the Second National Acute Spinal Cord Injury Study. *N Engl J Med*. 1990;322(20):1405–1411.
- Baptiste DC, Fehlings MG. Pharmacological approaches to repair the injured spinal cord. *J Neurotrauma*. 2006; 23(3–4):318–334.
- Brown PJ, Marino RJ, Herbison GJ, Ditunno JF Jr. The 72-hour examination as a predictor of recovery in motor complete quadriplegia. *Arch Phys Med Rehabil*. 1991;72(8): 546–568.
- Blaustein DM, Zafonte R, Thomas D, Herbison GJ, Ditunno J. Predicting recovery of motor complete quadriplegic patients: 24 hour v 72 hour motor index scores. *Am J Phys Med Rehabil*. 1993;72(5):306–311.
- McDonald JW, Becker D, Sadowsky CL, Jane JA Sr, Conturo TE, Schultz LM. Late recovery following spinal cord injury: case report and review of the literature. *J Neurosurg*. 2002; 97(suppl 2):252–265.
- Mulcahey MJ, Gaughan J, Betz RR, Vogel LC. Rater agreement on the ISCSCI motor and sensory scores obtained before and after formal training in testing technique. *J Spinal Cord Med*. 2007;30(suppl 1):S146–S149.
- Chafetz RS, Vogel LC, Betz RR, Mulcahey MJ. International Standards for Neurological Classification of Spinal Cord Injury: training effect on accurate classification. *J Spinal Cord Med*. 2008;31(5):538–542.
- Mulcahey MJ, Gaughan J, Betz RR, Johansen KJ. The International Standards for Neurological Classification of Spinal Cord Injury: reliability of data when applied to children and youths. *Spinal Cord*. 2007;45(16):452–459. epub 2006 Oct 3.
- American Spinal Injury Association and International Medical Society of Paraplegia, eds. *Reference Manual of the International Standards for Neurological Classification of Spinal Cord Injury*. Chicago, IL: American Spinal Injury Association; 2003.
- Pang D. Spinal cord injury without radiographic abnormality (SCIWORA). In: Mulcahey MJ, Betz R, eds. *The Child With a Spinal Cord Injury*. Rosemont, IL: American Academy of Orthopaedic Surgeons; 1996;139–160.
- Launay F, Leet AI, Sponseller PD. Pediatric spinal cord injury without radiographic abnormality: a meta-analysis. *Clin Orthop Relat Res*. 2005 Apr;433:166–170.
- Mann DC, Dodds JA. Spinal injuries in 57 patients 17 years or younger. *Orthopedics*. 1993;16(2):159–164.