ORIGINAL ARTICLE

Achieving assessor accuracy on the International Standards for Neurological Classification of Spinal Cord Injury

AJ Armstrong^{1,2,7}, JM Clark^{3,7}, DT Ho¹, CJ Payne¹, S Nolan^{2,4}, LM Goodes², LA Harvey⁵, R Marshall³, MP Galea⁶, SA Dunlop² on behalf of the Clinical Trial Collaborators⁸

Study design: A retrospective audit of assessor accuracy using the International Standards for Neurological Classification of Spinal Cord Injury (ISNCSCI) in three multicentre randomised controlled trials (SCIPA: Spinal Cord Injury and Physical Activity) spanning 2010–2014 with standards revised in 2011.

Objectives: To investigate assessor accuracy of neurological classification after spinal cord injury.

Setting: Australia and New Zealand.

Methods: ISNCSCI examinations were undertaken by trained clinicians prior to randomisation. Data were recorded manually and ISNCSCI worksheets circulated to panels, consensus reached and worksheets corrected. An audit team used a 2014 computerised ISNCSCI algorithm to check manual worksheets. A second audit team assessed whether the 2014 computerised algorithm accurately reflected pre- and post-2011 ISNCSCI standards.

Results: Of the 208 ISNCSCI worksheets, 24 were excluded. Of the remaining 184 worksheets, 47 (25.5%) were consistent with the 2014 computerised algorithm and 137 (74.5%) contained one or more errors. Errors were in motor (30.1%) or sensory (12.4%) levels, zone of partial preservation (24.0%), motor/sensory scoring (21.5%), ASIA Impairment Scale (AIS, 8.3%) and complete/incomplete classification (0.8%). Other difficulties included classification when anal contraction/sensation was omitted, incorrect neurological levels and violation of the ‘motor follows sensory rule in non-testable myotomes’ (7.4%). Panel errors comprised corrections that were incorrect or missed or incorrect changes to correct worksheets.

Conclusion: Given inaccuracies in the manual ISNCSCI worksheets in this long-term clinical trial setting, continued training and a computerised algorithm are essential to ensure accurate scoring, scaling and classification of the ISNCSCI and confidence in clinical trials.

Spinal Cord (2017) **55**, 994–1001; doi:10.1038/sc.2017.67; published online 20 June 2017

INTRODUCTION

Three multicentre randomised controlled trials, collectively termed Spinal Cord Injury and Physical Activity (SCIPA), investigated the potential for various modes of intensive physical therapies to improve recovery after spinal cord injury (SCI).^{1,2} As for many SCI trials,^{3,4} SCIPA is using the American Spinal Injury Association (ASIA) International Standards for the Neurological Classification of Spinal Cord Injury (ISNCSCI)⁵ for stratification of participants.^{1,2,6} In addition, as for other SCI trials, elements of the ISNCSCI were used as outcome measures.^{1,2,6}

The ISNCSCI was first developed in 1982 to ensure precision in classification of SCI, as well as accurate communication between physicians and researchers, and has since undergone regular revision and improvement^{7–10}—in 2003 and 2011 with the most recent version being published in 2015, after the completion of this study.⁵ The ISNCSCI involves two broad clinical skills, the first being the psychomotor skill of physical examination of a patient involving

scoring of sensory and motor functions coupled with S4/5 dermatome and anorectal testing. A number of studies have examined both validity and reliability of neurological examinations and have highlighted the importance of examiner training.^{11–17} The overall conclusion is that the ISNCSCI is an appropriate instrument for use by multiple examiners to discriminate, as well as evaluate, patients with SCI and to do so longitudinally.^{13,14,16} Nevertheless, continuing investigation of the ISNCSCI has been recommended to examine, for example, how elements are related to minimal clinically important differences and functional measures such as activities of daily living as well as to standardise the correct classification of challenging cases.^{18–20} In addition, the International Standards committee of the ASIA has agreed to review, and, if necessary, revise the Standards every 3 years.⁷

The second skill, which is the focus of this study, is cognitive and involves classifying the injury using the examination data. The classification procedure involves establishing motor and sensory levels

¹Faculty of Medicine, Dentistry and Health Science, The University of Western Australia, Crawley, Western Australia, Australia; ²School of Biological Sciences, The University of Western Australia, Crawley, Western Australia, Australia; ³South Australian Spinal Cord Injury Research Centre, Centre for Orthopedics and Trauma Research, Faculty of Health Sciences, The University of Adelaide, Adelaide, South Australia, Australia; ⁴Department of Physiotherapy, Royal Perth Hospital, Perth, Western Australia, Australia; ⁵John Walsh Centre for Rehabilitation Research, Sydney School of Medicine, University of Sydney & Kolling Institute, Royal North Shore Hospital, St Leonards, New South Wales, Australia and ⁶Department of Medicine (Royal Melbourne Hospital), The University of Melbourne, Parkville, Victoria, Australia

⁷Share equal first authorship.

⁸A full List of SCIPA (Spinal Cord Injury and Physical Activity) Trial Collaborators are available in the supplementary information.

Correspondence: Professor SA Dunlop, School of Biological Sciences, The University of Western Australia, (M092) 35 Stirling Highway, Crawley, Western Australia 6009, Australia. E-mail: sarah.dunlop@uwa.edu.au

Received 19 April 2017; accepted 2 May 2017; published online 20 June 2017

as well as a single neurological level of injury, total motor and sensory scores, an ASIA Impairment Scale (AIS) and severity of injury (complete/incomplete). In addition, for complete injuries, ISNCSCI involves establishing zones of partial preservation (ZPP).^{7,9} Earlier studies have revealed challenges in achieving correct classification,¹² whereas others have shown that training results in significant improvements in all components of the classification procedure.^{11,21–23} Notably, a recent European Multicentre Study on Human Spinal Cord Injury (EMSCI) showed that, regardless of trainee experience in SCI medicine, 2-day formal training sessions significantly improved classification skills.²³

The SCIPA trials variously have involved all 8 spinal units in Australia and New Zealand with patient recruitment spanning 2010–2014. Blinded assessors employed at participating sites were qualified physiotherapists and occupational therapists. All had extensive experience in working with SCI patients and experience in SCI rehabilitation. Undergraduate and postgraduate training on the ISNCSCI physical examination and assessment was variable. Some Australian blinded assessors attended undergraduate lectures and practical classes on the ISNCSCI, and then, as postgraduates, participated in specialist practicums under the direction of senior staff during 6-month rotations on spinal wards. The majority of blinded assessors at New Zealand sites had limited under- or postgraduate training, as medical practitioners typically conduct the ISNCSCI. Prior to the start of each trial, blinded assessors from both Australia and New Zealand received training on the ISNCSCI (see below). As part of quality control, we formed panels for each trial comprising experienced SCI clinicians (rehabilitation specialists and physiotherapists) as well as clinician and scientific researchers who were chief investigators and who were charged with checking blinded assessor manual ISNCSCI charts.

Given the challenges in ISNCSCI classification^{11,12,21,22} and the fact that, in other studies, errors occurred even after formal training,²³ we undertook a pilot study to assess the accuracy of a subset of ISNCSCI worksheets in one of the SCIPA trials.²⁴ Of the 20 worksheets examined, we identified inaccuracies in AIS classification, motor and sensory levels, and motor and sensory ZPP.²⁴ We therefore expanded the study to check the accuracy of the entire sample of 208 ISNCSCI worksheets generated manually in the three SCIPA trials. Previous studies have developed computerised algorithms^{25–27} that have been validated^{25,27} particularly for use in difficult cases (e.g. instances where several dermatomes/myotomes are classed as non-testable).²⁷ We took advantage of the freely available online computerised Rick Hansen Institute algorithm (v1.0.3, 2014) (<http://www.isncscialgorithm.com>)²⁸ that was only available at the conclusion of the three SCIPA trials and recently validated.²⁹

METHODS

SCIPA trials

Protocols have been published for each trial.^{1,2,6}

Use of the ISNCSCI

Prior to the start of each trial, a formal ISNCSCI training workshop was held at the Royal Talbot Rehabilitation Centre, Melbourne for staff from participating sites across Australia and New Zealand. Training involved a two-hour interactive seminar from a practising clinician routinely involved in ISNCSCI examination, classification and training followed by two hours of presenting example cases, testing and facilitated group discussions. For participants already experienced in the ISNCSCI assessment, the workshop provided an opportunity to refresh and update skills.

For each trial, full ISNCSCI assessments were performed at baseline prior to randomisation of participants to treatment group. A total of 208 ISNCSCI worksheets were completed manually, scanned, sent as soft copies to the SCIPA Programme Coordinator and then emailed to the respective panels involved in each trial. They were required to check the AIS grade because this was used to stratify participants in the SCIPA Full-On¹ and SCIPA Switch-On² trials and to describe the sample; for the Hands-On trial, stratification was by baseline ARAT score.⁶ Consensus amongst panels was reached via email and communicated to the Programme Coordinator. Where necessary, ISNCSCI worksheets were corrected by pen and classifications entered into the SCIPA database.

Because the trials spanned 2010–2014 and a revision of the ISNCSCI was published partway through the trials, that is, in 2011,⁹ we used the 2003 standards³⁰ for the two trials that had started before 2011 and the 2011 standards^{2,6} for the third trial that started after 2011.¹

To check the accuracy of the manually completed ISNCSCI worksheets, an audit team was formed (AJA, DTH and CJP, 5th-year medical students supervised by SAD). Their role was to enter the data from each manually completed ISNCSCI worksheet into the Rick Hansen Institute algorithm (v1.0.3). The audit team then compared the manual and computerised classifications and identified errors (see below). To ensure thorough understanding of the ISNCSCI, the audit team received training on ISNCSCI examination and classification at Royal Perth Hospital (RPH), Shenton Park Spinal Unit. RPH physiotherapists and senior spinal nursing staff who routinely use the ISNCSCI demonstrated several examinations, manual data recording on ISNCSCI worksheets, as well as clarifying any areas of difficulty in classification.

Analysis

At the conclusion of the trials, hard copies of the ISNCSCI worksheets, whether corrected or not, were posted by the Programme Coordinator to the RPH site. The three members of the audit team used the Rick Hansen Institute ISNCSCI algorithm (v1.0.3)²⁸ to check the accuracy of ISNCSCI manual worksheets. Thus, each member of the audit team reviewed each of the 184 manual worksheets using the algorithm.

To take account of changes in the standards between 2003 and 2011, a separate audit team of physiotherapists trained in the ISNCSCI assessment (JMC, SN and MPG) checked the manual ISNCSCI worksheets again without the use of any computerised algorithm. The second audit team was directed to utilise the 2003 standards³⁰ for the 2 trials that started before 2011^{2,6} and the 2011 standards¹ for the trial that started after 2011. The data were then cross-checked against those derived from the 2014 computerised algorithm (Figure 1).

Variables were coded into 5 broad categories as follows: motor and sensory levels, AIS, complete/incomplete status, ZPP (sensory & motor), and summed sensory and motor scores (Figure 2); these follow categories and their abbreviations as described previously.^{7,9} For all variables, a yes/no classification was assigned based on accuracy of the original examiner judgement. Where errors were made in motor and sensory levels and summed motor/sensory scores, the degree of inaccuracy was documented as, for example, left motor level two levels too high, right motor score five points too low and so on. In instances where the same error was made for both the left and right sides (that is, motor and sensory levels and ZPP), these were counted as one, and not two, errors. A total of 16 different subclasses of errors were identified and evaluated (Figure 2, Table 1).

Data were tabulated using Microsoft Excel with correctness marked with yes/no (y/n). Simple, descriptive statistics (counts, percentages) were applied to nonparametric data. Copies of all computerised ISNCSCI worksheets were submitted to the SCIPA Programme Coordinator and any errors corrected on the SCIPA database.

Statement of ethics

We certify that all applicable institutional and hospital regulations concerning the ethical use of human volunteers were followed during the SCIPA trials.^{1,2,6} For this audit, patient and examiner consent was obtained for training, but ethical approval was not required because training was for educational purposes and not data collection.

Figure 1 Flow diagram depicting the auditing process showing 100% consistency between second ‘manual’ audit team utilising the pre-2011 and post-2011 standards as appropriate, and the Rick Hansen Institute 2014 algorithm results generated by the first audit team.

RESULTS

Identification of errors

Of the 208 ISNCSCI manual worksheets from the three SCIPA trials, 12 were excluded because of illegible handwriting and 12 were excluded because there were dermatomes/myotomes that were ‘non-testable’, and thus motor and sensory levels as well as motor and sensory scores were classified as ‘unable to determine’ (Figure 1). The remaining 184 manual charts were reviewed by the audit team (Figure 1). Forty-seven manual worksheets had no identifiable errors as revealed by the ISNCSCI algorithm (25.5%) and the remaining 137 (74.5%) worksheets had one or more, giving a total of 242 errors.

Three sets of errors were identified, namely examiner errors that were correctly changed by the panels (118 errors; Table 1, column A), examiner errors that persisted after panel review because they were either incorrectly changed, missed or not checked (124 errors; Table 1, column B) and errors where the original examiner ISNCSCI worksheets were correct, but were incorrectly changed by the panels (30 errors; Table 1, column C).

As we are primarily interested in the overall error rate of examiner classification, we included the first 2 sets of errors (242 cases) but excluded incorrect changes to accurate worksheets made by the panels. For example, when considering classification of motor level across the

three trials, a total of 73 errors were identified (Table 1, column D, $55+18=73$). Of these, 42 were corrected following panel review (Table 1, column A, $40+2=42$), 31 remained incorrect either because errors were incorrectly changed, not checked or missed (Table 1, column B, $15+16=31$) and 15 that were initially accurate but were incorrectly changed (Table 1, column C, $11+4=15$). The motor level subclass error therefore constituted 30.1% (Table 1, column E, $22.7\%+7.4\%=30.1\%$; Figure 3) of the total number of errors in the three trials ($(55+18)/242$). Because some ISNCSCI worksheets had more than one error (88/184; 47.8% data not shown), we also calculated the proportion of ISNCSCI worksheets with each subclass of error. For example, when considering classification of motor levels across the three trials, 39.7% of ISNCSCI worksheets had errors for this subclass (Table 1, column F, $29.9\%+9.8\%$; Figure 4).

Distribution of errors for the five categories of variable

Errors for the 5 categories of variable in all 3 trials are summarised (Table 1, column E; Figure 3). For neurological levels, we identified a total of 103/242 (42.5%) errors with 73/242 (30.1%) being for motor levels and 30/242 (12.4%) for sensory levels. Of the 73 ($55+18=73$) errors in motor level classification, 18/73 (24.7%) were attributable to errors pertaining to assessment of clinically non-testable myotomes

Figure 2 Tree diagram depicting five key classes of variables (grey boxes, left-hand column) namely (I) neurological level (that is, motor level and sensory level); (II) AIS; (III) zone of partial preservation; (IV) scoring and (V) individually identified errors. Errors in 8 major subclasses of variables (coloured boxes, centre column+‘other’ in left-hand column) are shown in Figures 3 and 4. Errors for all subclasses of variable (numbers 1–16 to bottom right of coloured boxes) are shown in Table 1.

(violations of the ‘motor follows sensory’ rule). We also noted that 49/103 (47.6%) of the errors in motor or sensory levels were $\geq$ two levels (data not shown). There were 20/242 (0%+5.4%+2.9% = 8.3%) errors for the AIS grade, 2/242 (0.8%) instances of incorrect injury severity (complete/incomplete) and 58/242 (6.6%+12.4%+5.0% = 24%) errors for the ZPP.

Errors in motor and sensory scores totalled 52/242 (2.5%+7.4%+2.1%+9.5% = 21.5%) and arose either because they were incorrectly summed, not done or recorded in the wrong box. Three of these errors were identified by the panels (Table 1, column A). For the 41 cases where motor and sensory scores were summed incorrectly, 24/41 (58.5%) were erroneous by $\geq$ five points (data not shown). There were 4 instances where the single neurological level of injury

was erroneously recorded in the AIS classification box and 3 instances where deep anal pressure or voluntary anal contraction was not recorded (Table 1, columns A and B).

Revisions of the standards between 2003³⁰ and 2011⁹ are summarised in Table 2. In brief, there are 5 revisions of which only 2 are pertinent to the current study. These were C2/C1 classification, for which we found only 2 worksheets that were inaccurate, as well as when omitting deep anal pressure assessment is permissible, for which we found only 3 errors. Thus, the changes between the 2003 and 2011 standards had virtually no impact on the accuracy of the 2014 algorithm, which was supported by the findings of the second audit team. The remaining revisions were not pertinent to the current study and pertain to changing terminology from ‘deep anal sensation’ to ‘deep anal pressure’, the definition of ZPP and changes to the worksheet format. In addition, there were 5 other topics for which clarification was provided in the revision as follows: (1) the examination procedure itself; (2) clarification of the motor follows sensory rule, which did not change between the 2 standards but nevertheless remains a contentious issue; (3) what to document when a complete injury lacked a ZPP; (4) clarification regarding AIS classification; and (5) utilising non-key muscles in defining the ZPP in otherwise not-testable muscles.

When comparing the second audit to that generated using the 2014 algorithm, there was an overall consistency of 192/208 (91.3%) across the datasets (Figure 1). For the 16 manual charts for which discrepancies were found, 5 were due to errors made by the first audit team when tabulating and coding the data generated by the algorithm. The remaining 11 were made by second audit team and were due to missing errors that were identified by the algorithm.

DISCUSSION

The ISNCSCI instrument is widely used in clinical trials^{3,4} and accurate interpretation of neurological assessment following SCI is crucial in quantifying possible therapeutic responses.³¹ Here, we identify significant error rates in manual classification procedures by examiners (153) that were detected by panels. Strikingly, despite training provided prior to each trial and after panel review, 124 errors persisted and 30 new errors were introduced. Some, but not all, errors were detected by the panels, but a *post hoc* discussion revealed that not all panels were checking all aspects of the worksheets. The findings provide compelling evidence about the inherent challenges in ISNCSCI assessment and classification and for the fundamental importance of thorough and continued training and the use of the most current, freely available online computerised algorithms to ensure classification accuracy while conducting clinical trials.

Previous studies investigating the accuracy of classification procedures for the ISNCSCI have examined the effect of training in specific training settings rather than while undertaking large multicentre clinical trials. Three studies used relatively small numbers of trainees and SCI cases. In one study, two written cases from the 1994 reference manual were used and 15 house officers were tested two months apart.²² The remaining two studies also predated the development of the computerised ISNCSCI algorithm, and we assume that they used written worksheets; one study involved two SCI cases and four experts¹² and the other two cases and 106 professionals.¹¹ All three studies showed significant improvements after training. A more recent study from the EMSCI reported the benefits of training 106 participants who classified five challenging SCI cases manually in a total of 10 formal training workshops.²³

In contrast to the above studies which when compared before and after training, another study took a different approach to determine

Table 1 Counts (*n*) and percentages (%) of errors for the different subclasses of variables (numbers 1–16, see Figure 2) for the three SCIPA trials

			A # Examiner errors CORRECTLY changed by panel r/v	B # Examiner errors STILL INCORRECT after panel r/v	C # Accurate classifications INCORRECTLY changed by panel r/v	D Total # of exam- iner errors (that is, excluding column C)	E % each subclass constitutes total error number of 242	F % ISNCSCI charts with errors in each subclass
<i>ISNCSCI variables for all trials (see also Figure 1)</i>								
<i>Neurological levels (that is, motor and sensory levels)</i>						Total = 103	Total = 42.5	
1	1a	Motor level incorrect	40	15	11	55	22.7	29.9
2	1ai	Violation of 'motor follows sensory' rule	2	16	4	18	7.4	9.8
3	1b	Sensory level incorrect	17	13	8	30	12.4	16.3
<i>AIS classification and severity (complete/incomplete)</i>						Total = 22	Total 2a = 8.3 Total 2b = 0.8	
4	2a	2ai AIS A/B error	0	0	0	0	0	0
5		2aii AIS B/C error	8	5	0	13	5.4	7.1
6		2aiii AIS C/D error	3	4	0	7	2.9	3.8
7	2b	Complete/incomplete error	2	0	0	2	0.8	1.1
<i>Zone of partial preservation</i>						Total = 58	Total = 24	
8	3a	ZPP incorrectly included/omitted	9	7	1	16	6.6	8.7
9	3bi	ZPP incorrectly calculated (motor)	22	8	0	30	12.4	16.3
10	3bii	ZPP incorrectly calculated (sensory)	7	5	6	12	5.0	6.5
<i>Motor (a) and sensory (b) scores</i>						Total = 52	Total = 21.5	
11	4a	4ai Score not calculated	0	6	0	6	2.5	3.3
12		4aii Summed incorrectly	2	16	0	18	7.4	9.8
13	4b	4bi Score not calculated	0	5	0	5	2.1	2.7
14		4bii Summed incorrectly	1	22	0	23	9.5	12.5
<i>Other</i>						Total = 7	Total = 2.9	
15	5	DAP/VAC not recorded	1	2	0	3	1.2	1.6
16		SNLI recorded in AIS box	4	0	0	4	1.7	2.2
Totals			118	124	30	242	100.0	NA ^a

Abbreviations: AIS, American Spinal Injury Association Impairment Scale; DAP, deep anal pressure; ISNCSCI, International Standards for the Neurological Classification of Spinal Cord Injury; NT, not tested; SCIPA, Spinal Cord Injury and Physical Activity; SNLI, single neurological level of injury; VAC, voluntary anal contraction; ZPP, zone of partial preservation.

^aThe % values in column F do not add up to 100% because some ISNCSCI worksheets had more than one error (see also Figure 4).

reliability and repeatability of the ISNCSCI by training eight physicians and eight physical therapists who were then required to evaluate three patients for motor and sensory scores on the same day as the training day and one patient the following day.¹⁶ The results showed overall high inter-rater reliability and better repeatability for complete compared to incomplete patients in this setting.

However, the above studies examined accuracy in specific training settings over short periods. Our study differs because, while we included training for all examiners in the trial protocols,^{1,2,6} we did not include either pre- or post-training testing of the examiners, nor of the panels and, for practical and financial reasons, neither did we repeat the training during the course of the trials. Rather, once we detected errors in our pilot study,²⁴ we decided to use a computerised ISNCSCI algorithm to examine accuracy of manual classification procedures undertaken in a real-life situation of multicentre randomised controlled trials across 8 spinal units in 2 countries over a

period of 5 years. To our knowledge, this is the first study of this nature, and our findings highlight the importance of continued training combined with using now freely available computerised ISNCSCI algorithms.

Our findings are supported by a 5-year EMSCI study.²³ Direct comparison is difficult because the EMSCI study examined the effect of training on participants' ability to correctly classify five cases during a 2-day workshop while we examined 184 cases in a real-time randomised control trial scenario over a period of 5 years. The EMSCI study also has two sets of data—i.e. pre- and post-training. Because our examiners received training, we compared our data to post-training data from the EMSCI study.²³ Comparison is also difficult because types of error have been attributed somewhat differently in the two studies. Nevertheless, in both the EMSCI study and ours (Figure 4—for which we converted 'correct answers' into 'error'), the greatest number of errors was found in classifying motor

levels (EMSCI: 18.1% of worksheets; this study: 39.7%). The EMSCI study also concluded that clinically non-assessable myotomes were the most prominent source of difficulty in the ISNCSCI classification procedure.²³ Violation of the 'motor follows sensory' rule in the levels without testable motor function (C1-4, T2-L1) accounted for 69.7% of the EMSCI motor level errors.²³ In our study, we found that violation of this rule accounted for 25% of the motor level errors, emphasising the need to practice this rule in ISNCSCI training programmes.

AIS and complete/incomplete classifications were less prone to error and similar in both studies (EMSCI: AIS: 11.9%; this study: 10.9%; complete/incomplete: EMSCI: 3.8%; this study: 1.1%). However, errors in sensory levels were over fivefold greater in our study (16.3%) compared to those in the EMSCI study (3.2%).²³ Our identification of high error rates in ZPP (31.5%) suggests that this aspect also requires attention, along with the protocol regarding the implications of calculating scores when there are dermatomes/

myotomes 'not-testable', and how to manage cases where voluntary anal contraction and deep anal pressure sensation are non-examinable.

We also identified miscalculation and summation errors in motor (13.1%) and sensory (15.2%) scores (Figure 4). Some of these errors were associated with scores being summed incorrectly (9.8% motor and 12.5% sensory). This is important because other construct validity studies have identified that clinically significant improvement correlates with increases in motor scores of five points or more as well as changes in single neurological levels of two levels or more.²⁰ For some studies, errors of this degree in summed motor and sensory scores and single neurological levels could potentially affect results and therefore overall interpretation of study effects. However, use of blinded assessors prevents any impact on between-group comparisons because any error will not systematically bias the results. Of note, for the three SCIPA trials, only parts of the ISNCSCI were used as secondary outcome measures. These were change in upper-limb sensory scores for 'Hands-On',⁶ change in motor score for 'Full-On'¹ and changes in motor and sensory scores for 'Switch-On'.²

Overall, although EMSCI found slightly under one-half the percentage of errors in motor level (18.1%) compared to this study (39.7%), we found approximately two-thirds fewer (24.7%) violations of the motor follows sensory rule compared to EMSCI (69.7%). Similarly, although errors in sensory level were more than fivefold higher in this study (16.3%) compared to those in the EMSCI study (3.2%), errors in AIS and complete/incomplete classification were similarly low in both studies. The findings suggest that, although training in an EMSCI workshop setting reviewing five patients over two days results in fewer errors in, for example, motor and sensory levels compared to this study, errors still occur even over this short period. Errors in ISNCSCI classification therefore might not be unexpected over a considerably longer clinical trial period of 5 years. In addition, although continued training on ISNCSCI assessments has been recognised as important,³¹ this was not included in our design. Of concern also are the errors made by the panels. However, *post hoc* discussion revealed that not all panel members checked all aspects of the worksheet; therefore, their error rate may not be a true reflection of their skills. To our knowledge, this is the first study that

Figure 3 Percentage (%) distribution of errors in the audited ISNCSCI worksheets ($N=242$). Pie-chart shows each of the eight major subclasses of error ($N=184$; Figure 2). Data are for all three trials and refer to (Table 1, column E).

Figure 4 Percentage (%) distribution of audited ISNCSCI worksheets ($N=184$) with errors in each major subclass. Note: the total is >100% because a proportion (48%) of ISNCSCI worksheets had more than one error. Data are for all three trials and refer to (Table 1, column F).

Table 2 Adapted from Kirshblum *et al.*⁷ reference for the 2011 revision of the ISNCSCI

Topic	Pre-2011 ³⁰	Post-2011 ⁸	Number of worksheets affected
<i>Revisions relevant to the current study</i>			
S4/S5 sensory testing terminology	S4/S5 sensory terminology of 'deep anal sensation' utilised.	Revised to 'DAP' terminology.	Not relevant to ISNCSCI audit.
C2 and C1 sensory classification.	Nil direction pertaining to the classification of C2 and higher neurological impairment in the pre-2011 standards.	Advised that if sensation is abnormal at C2, the neurological level should be designated as 'C1'.	4 patients with impairment at C1 sensory dermatome: sensory NLI quoted as 'proximal to C2' ($n=2$). Correctly assessed as C1 ($n=2$).
When omitting DAP assessment is permissible.	Pre-2011, DAP constitutes a mandatory aspect of the ISNCSCI assessment.	Clarifying that in patients who have light touch or pin prick sensation at S4–S5, examination for DAP is not required.	Three patients were flagged as incorrect for omitting the DAP assessment: the assessment was reviewed, and deemed the same both pre- and post- 2011 ISNCSCI standards.
ZPP definition.	Amount of ZPP sensory/motor sparing defined by the number of motor/sensory level differences between the most caudal preserved motor/sensory function with reference to the SNLI.	Defining the ZPP with reference to the corresponding NLI (to be consistent with InSTeP training).	Not relevant.
Worksheet format changes.	Nil comment.	2011 changes adjusted the worksheet for ease of use.	Not relevant.

Abbreviations: DAP, deep anal pressure; ISNCSCI, International Standards for the Neurological Classification of Spinal Cord Injury; SNLI, single neurological level of injury; ZPP, zone of partial preservation. Adapted from Kirshblum *et al.*⁷ depicting the modifications that were made following the 2011 revision of the ISNCSCI. Adapted from text contained within 'Reference for the 2011 revision of the ISNCSCI'.⁷

incorporated panel consensus in an attempt to ensure quality control. However, whereas panel consensus improved accuracy in 51.2% of cases, in the remainder errors were missed and sometimes introduced by erroneously changing initially correct charts, suggesting that while continued training should be promulgated, the ISNCSCI is inherently complex and challenging, even for senior practitioners with considerable experience working with this instrument.

Despite correction for updated consensus guidelines,^{7,8} our audit identified multiple formative and systematic assessor errors in ISNCSCI-written worksheets, and it did so under protocol-controlled conditions. The challenges are to introduce a standardised method to index-written worksheets to *in silico* classifications, as exemplified by the online Rick Hansen Institute ISNCSCI classification algorithm and also, to improve the sensitivity and specificity of the next generation of *in silico* systems to discriminate complex neurology. The introduction of a standardised method for the categorisation of 'complex' and 'non-complex' neurology within the ISNCSCI-written worksheet and its electronic derivatives, individually or interactively, may have an additional or complementary role. The ISNCSCI instrument is the best that we have, and any significant advance in its sensitivity or specificity to discriminate clinically significant improvements may pave the way for appropriate therapies.

Taken together, the findings point to the need for emphasis on continued blinded assessor training coupled with the development of computerised algorithms with capability for differential diagnosis to reduce assessor error and ensure accurate classification.

CONCLUSION

Given that 75% of ISNCSCI worksheets had one or more errors when completed manually while undertaking randomised controlled trials, both continued training and free online computational algorithms are key to improving accuracy and therefore confidence in trial outcomes. Although possibly impractical for real-time assessment, computational

algorithms should be used to check manually completed worksheets. Although computer algorithms cannot replace clinical judgement, for example in complex cases where a brachial plexus injury may complicate scoring of an otherwise normal myotome,¹⁹ they provide an effective means for both validating ISNCSCI examination classification in a consistent and standardised way thus ensuring confidence in outcome measures involving the ISNCSCI as well as educating future generations of healthcare professionals.

DATA ARCHIVING

There were no data to deposit.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The work was funded by the Transport Accident Commission (Victorian Neurotrauma Initiative), the Lifetime Care and Support Authority NSW, the Australian Government National Health and Medical Research Council (APP1002347 and APP1028300) and the Neurotrauma Research Program (Western Australia).

- 1 Galea MP, Dunlop SA, Marshall R, Clark J, Churilov L. Early exercise after spinal cord injury ('Switch-On'): study protocol for a randomised controlled trial. *Trials* 2015; **16**: 7.
- 2 Harvey LA, Dunlop SA, Churilov L, Hsueh Y-SA, Galea MP. Early intensive hand rehabilitation after spinal cord injury ('Hands On'): a protocol for a randomised controlled trial. *Trials* 2011; **12**: 14.
- 3 Fehlings MG, Theodore N, Harrop J, Maurais G, Kuntz C, Shaffrey CI *et al*. A phase I/IIa clinical trial of a recombinant Rho protein antagonist in acute spinal cord injury. *J Neurotrauma* 2011; **28**: 787–796.
- 4 Fehlings MG, Vaccaro A, Wilson JR, Singh A, W Cadotte D, Harrop JS *et al*. Early versus delayed decompression for traumatic cervical spinal cord injury: results of the Surgical Timing in Acute Spinal Cord Injury Study (STASCIS). *PLoS ONE* 2012; **7**: e32037.

- 5 Kirshblum SC, Waring Waring, Biering-Sorensen F, Burns SP, Johansen M *et al*. *International Standards for the Neurological Classification of Spinal Cord Injury, Revised 2015*, 8th edn, American Spinal Injury Association: Atlanta, GA, USA: 2015.
- 6 Galea MP, Dunlop SA, Davis GM, Nunn A, Geraghty T, Hsueh Y-SA *et al*. Intensive exercise program after spinal cord injury ('Full-On'): study protocol for a randomized controlled trial. *Trials* 2013; **14**: 291.
- 7 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**: 547–554.
- 8 Kirshblum S, Waring W 3rd. Updates for the International Standards for Neurological Classification of Spinal Cord Injury. *Phys Med Rehabil Clin N Am* 2014; **25**: 505–517 vii.
- 9 Kirshblum SC, Burns SP, Biering-Sorensen F, Donovan W, Graves DE, Jha A *et al*. International standards for neurological classification of spinal cord injury (revised 2011). *J Spinal Cord Med* 2011; **34**: 535–546.
- 10 Waring WP 3rd, Biering-Sorensen F, Burns S, Donovan W, Graves D, Jha A *et al*. 2009 review and revisions of the international standards for the neurological classification of spinal cord injury. *J Spinal Cord Med* 2010; **33**: 346–352.
- 11 Cohen ME, Ditunno JF Jr, Donovan WH, Maynard FM Jr. A test of the 1992 International Standards for Neurological and Functional Classification of Spinal Cord Injury. *Spinal Cord* 1998; **36**: 554–560.
- 12 Donovan WH, Brown DJ, Ditunno JF Jr, Dollfus P, Frankel HL. Neurological issues. *Spinal Cord*. 1997; **35**: 275–281.
- 13 Furlan JC, Fehlings MG, Tator CH, Davis AM. Motor and sensory assessment of patients in clinical trials for pharmacological therapy of acute spinal cord injury: psychometric properties of the ASIA Standards. *J Neurotrauma* 2008; **25**: 1273–1301.
- 14 Furlan JC, Noonan V, Singh A, Fehlings MG. Assessment of impairment in patients with acute traumatic spinal cord injury: a systematic review of the literature. *J Neurotrauma* 2011; **28**: 1445–1477.
- 15 Jonsson M, Tollback A, Gonzales H, Borg J. Inter-rater reliability of the 1992 international standards for neurological and functional classification of incomplete spinal cord injury. *Spinal Cord* 2000; **38**: 675–679.
- 16 Marino RJ, Jones L, Kirshblum S, Tal J, Dasgupta A. Reliability and repeatability of the motor and sensory examination of the international standards for neurological classification of spinal cord injury. *J Spinal Cord Med* 2008; **31**: 166–170.
- 17 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**: 444–451.
- 18 Curt A, Van Hedel HJ, Klaus D, Dietz VGroup E-SS. Recovery from a spinal cord injury: significance of compensation, neural plasticity, and repair. *J Neurotrauma* 2008; **25**: 677–685.
- 19 Kirshblum SC, Biering-Sorensen F, Betz R, Burns S, Donovan W, Graves DE *et al*. International Standards for Neurological Classification of Spinal Cord Injury: cases with classification challenges. *J Spinal Cord Med* 2014; **37**: 120–127.
- 20 Scivoletto G, Tamburella F, Laurenza L, Molinari M. Distribution-based estimates of clinically significant changes in the International Standards for Neurological Classification of Spinal Cord Injury motor and sensory scores. *Eur J Phys Rehabil Med* 2013; **49**: 373–384.
- 21 Chafetz RS, Vogel LC, Betz RR, Gaughan JP, Mulcahey MJ. International standards for neurological classification of spinal cord injury: training effect on accurate classification. *J Spinal Cord Med* 2008; **31**: 538–542.
- 22 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**: 268–270.
- 23 Schuld C, Wiese J, Franz S, Putz C, Stierle I, Smoor I *et al*. Effect of formal training in scaling, scoring and classification of the International Standards for Neurological Classification of Spinal Cord Injury. *Spinal Cord* 2013; **51**: 282–288.
- 24 Galea MP, Marshall R, Clark J, Dunlop SA, Hurley M, Beasy S. The international standards for neurological classification of spinal cord injury (ISNCSCI): consensus between expert examiners and clinicians (abstract). The 53rd ISCoS Annual Scientific Meeting; 2–4 September 2014; Maastricht, The Netherlands. International Spinal Cord Society: Aylesbury, Buckinghamshire, UK, 2014.
- 25 Chafetz RS, Prak S, Mulcahey MJ. Computerized classification of neurologic injury based on the international standards for classification of spinal cord injury. *J Spinal Cord Med* 2009; **32**: 532–537.
- 26 Linassi G, Li Pi, Shan R, Marino RJ. A web-based computer program to determine the ASIA impairment classification. *Spinal Cord* 2010; **48**: 100–104.
- 27 Schuld C, Wiese J, Hug A, Putz C, Hedel HJ, Spiess MR *et al*. Computer implementation of the international standards for neurological classification of spinal cord injury for consistent and efficient derivation of its subscores including handling of data from not testable segments. *J Neurotrauma* 2012; **29**: 453–461.
- 28 Rick Hansen Institute. ISNCSCI Algorithm <http://www.isncscialgorithm.com/>. International Spinal Cord Society 2014.
- 29 Walden K, Belanger LM, Biering-Sorensen F, Burns SP, Echeverria E, Kirshblum S *et al*. Development and validation of a computerized algorithm for International Standards for Neurological Classification of Spinal Cord Injury (ISNCSCI). *Spinal Cord* 2015; **54**: 197–203.
- 30 Marino R, Barros T, Biering-Sorensen F, Burns SP, Donovan WH, Graves D *et al*. International standards for neurological classification of spinal cord injury. *J Spinal Cord Med* 2003; **26** (Suppl 1): S50–S56.
- 31 Steeves JD, Lammertse D, Curt A, Fawcett JW, Tuszynski MH, Ditunno JF *et al*. Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP panel: clinical trial outcome measures. *Spinal Cord* 2007; **45**: 206–221.

Supplementary Information accompanies this paper on the Spinal Cord website (<http://www.nature.com/sc>)