ORIGINAL ARTICLE

Reliability of the radiographic variables in the International Spinal Cord Injury Spinal Column Injury Basic Data Set compared with the AO classification

C Lucantoni[1], RG Krishnan[1], M Gehrchen[1], DW Hallager[1], F Biering-Sørensen[2] and B Dahl[1]

Study design: Intra- and interrater reliability study for radiological variables of the International Spinal Cord Injury (SCI) Spinal Column Injury Basic Data Set.

Objectives: To test reliability of the radiological variables in the International SCI Spinal Column Injury Basic Data Set and compare it with the Arbeitsgemeinschaft für Osteosynthesefragen (AO) classification.

Setting: The database of Eastern Denmark Regional SCI Referral Center, Copenhagen, Denmark.

Methods: Ratings of the International SCI Spinal Column Injury Basic Data Set radiological variables and AO classification were obtained by two international observers for all the surgically treated spine trauma patients between 1st October 2010 and 31st December 2012 at the Spine Unit, Rigshospitalet, Denmark. Statistical analyses for intra- and interrater crude agreement and Cohen's unweighted kappa (κ) coefficients were performed.

Results: For 283 spine injuries, the intra- and interrater reliability for the individual radiological variables of the International SCI Spinal Column Injury Basic Data Set was at least substantial ($\kappa = 0.67$–0.97 for interrater, $\kappa = 0.79$–0.89 for the intrarater agreement). For the AO classification, intrarater reliability was moderate-to-substantial ($\kappa = 0.57$–0.75), whereas interrater reliability was substantial ($\kappa = 0.67$–0.69). The crude intra- and interrater agreement for a combined radiographic SCI Spinal Column Injury Basic Data Set variable showed no significant difference compared with the AO classification ($P = 0.067$–0.895).

Conclusions: The reliability of International SCI Spinal Column Injury Basic Data Set radiological variables is comparable to the AO classification system. We encourage its use for spinal column injury description, thus facilitating data collection and comparison between centres and countries.

INTRODUCTION

The ideal classification of spinal column injuries serves as a guidance for treatment, as well as a method of communication between clinical researchers. Also, the system should be easily implemented into clinical practice and show reliability among different users. The two most commonly used classification systems are the Arbeitsgemeinschaft für Osteosynthesefragen (AO) classification,[1] introduced in 1994 for thoracolumbar trauma and extended to cervical spine,[2] and the thoracolumbar injury classification system (TLICS).[3,4] One of the major differences between the two systems is the inclusion of the patient's neurological status in the TLICS, as well as consideration of integrity of the posterior ligamentous complex (PCI). Still, studies on observer variation with both scoring systems have indicated that the optimal system has not yet been established.[5,6]

Recently, an international committee of experts suggested the International Spinal Cord Injury (SCI) Spinal Column Injury Basic Data Set with variables reflecting trauma mechanism and describing structural injuries affecting the integrity of the spinal column.[7] The aim of this data set is to provide a uniform method for spinal column injury description, thus facilitating a common language among different SCI centres worldwide.

For all international data sets and classifications, reliability is of paramount importance to allow comparison of data from different facilities or countries. As the radiological variables in the International SCI Spinal Column Injury Basic Data Set have not yet been tested for reliability, the aim of the present study was to evaluate the intra- as well as the interrater reproducibility of these variables and have them compared with the reliability of the AO classification system.

METHODS

Patients and ratings

Radiological electronic database of a tertiary referral unit for spine trauma, Rigshospitalet, Copenhagen University, Denmark, was retrospectively reviewed. All patients undergoing surgical treatment for a traumatic spinal injury in the period 1st October 2010 through 31st December 2012 were included. The centre serves the Eastern Denmark, corresponding to a population of 2.6 million.

Observers involved in the study were two fellows in spine surgery at Rigshospitalet, Copenhagen, RB (rater 1) and CL (rater 2), respectively, one orthopaedic surgeon and one neurosurgeon, with their education and clinical

[1]Spine Unit, Department of Orthopaedic Surgery, Rigshospitalet, Copenhagen, Denmark and [2]Clinic for Spinal Cord Injuries, Neuroscience Center, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark
Correspondence: Dr C Lucantoni, Spine Unit, Department of Orthopaedic Surgery, Rigshospitalet, 9 Blegdamsvej, Copenhagen 2100, Denmark.
E-mail: clucantoni@gmail.com

Table 1 Spinal Column Injury Data Set radiological variables

Variable	Values
Spinal column injury(ies)	Yes/No/Unknown
Single or multiple spinal column level injury(ies)	Single/Multiple/Unknown
Spinal column injury number	1/2/3/…n
Spinal column injury level	vC00–vC07: Cervical (C0–C7); vT01–vT12: Thoracic (T1–T12); vL01–vL05: Lumbar (L1–L5); vS01–vS05: Sacrum (S1–S5); vC99: Unknown Cervical (C0–C7); vT99: Unknown Thoracic (T1–T12); vL99: Unknown Lumbar (L1–L5); vS99: Unknown Sacral (S1–S5); vX99: Unknown level
Disc/posterior ligamentous complex (PLC) injury	Yes/No/Unknown
Traumatic translation	Yes/No/Unknown

experiences from different countries (India and Italy). Observers had not been involved in the treatment of the patients, and they were not aware of the patients' clinical information. Before the study, the observers were provided with the original description of the International SCI Spinal Column Injury Basic Data Set and the AO classification system.[1,2,7]

The International SCI Spinal Column Injury Basic Data Set includes seven elements identifying and describing the spinal column injury(ies).[7]

Six out of these seven variables, considered the 'radiological variables' of the data set, were used to classify all the identified column injuries between C0 (occiput) and sacrum, following the proposed scheme[7] (Table 1). The spinal column injury(ies) variable documents 'whether there is any disruption of the spinal column', including bone, ligaments, capsules, discs and other supportive soft tissue. The single or multiple spinal column injury variable documents whether the injury involves one or more levels separated by at least a motion segment (i.e. two adjacent vertebrae and interconnecting disc and ligaments) of intact spinal column. The disc/PLC injury variable documents whether there is a disc and/or a PLC injury for each level. Several criteria have been proposed to define this lesion: radiographic widening of the spinous processes, subluxation, displaced fracture or dislocation of the facets in the presence of an acute deformity. The traumatic translation variable documents whether there is or not any traumatic sagittal and/or coronal translation (at least 3.5 mm malalignment in sagittal and coronal plane for cervical spine and 2.5 mm for thoracolumbar spine were suggested as cutoff values).[7]

All injuries caudal to C2 were also classified according to the AO classification system including type and group.[1,2] Axial cervical injuries (C0, C1, C2) were excluded from the AO classification process, given the different classification schemes usually adopted for this anatomic region.

Ten sets of radiographs were used for pre-evaluation and were not included in the final study. The two observers independently rated all the available preoperative radiographs, computed tomography (CT)-scans and, if available, MRIs. The procedure was repeated after 3 weeks with case order scrambled between the two assessments. When a case had multiple spinal injuries, variables were classified for each injury individually. Observers were blinded against each other's ratings. The observers were two international spine surgery fellows from different countries but with similar experience regarding clinical assessment and treatment of patients with spinal injuries.

To compare agreements for the International SCI Spinal Column Injury Basic Data Set to the AO classification, four radiographic variables (Single or Multiple Level Spinal Column Injury, Spinal Column Injury Level, Disc and/or PCI, Traumatic Translation) for injuries caudal to C2 were aggregated into a composite variable named 'D(ata)S(et) Entire Classification'. Agreement for this latter variable was considered when there was agreement in all the four individual radiological variables.

Statistical analysis

Including all injuries, crude agreement (%) of inter- and intrarater agreement and Cohen's unweighted Kappa coefficients (κ) with 95% confidence intervals were calculated for the four radiographic variables of the International SCI Spinal Column Injury Basic Data Set separately, for the AO classification and the DS entire classification. To test for differences in proportions of cases of agreement versus non-agreement between DS entire classification and the AO classification, McNemar's test for marginal homogeneity was used. Rater bias,

defined as a systematic difference in the distribution of classifications between raters, was assessed using McNemar's or Bhapkar's tests for marginal homogeneity in contingency tables with two or more categories, respectively. During statistical calculations, patients with missing data for a rating of the variable in question were excluded from the respective analysis. According to Landis and Koch,[8] κ coefficients of agreement < 0.00 were considered poor, 0.00–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial and 0.81–1.00 almost perfect. The statistical tests were calculated using the software packages 'irr'[9] and 'gmodels'[10] for statistical language R version 3.1.0.[11] A significance level of $P < 0.05$ was chosen.

RESULTS

The two raters separately classified 271 surgically treated patients with 283 spinal column injuries according to the International SCI Spinal Column Injury Basic Data Set and 259 injuries caudal to C2 according to the AO classification.

Patient characteristics

The mean age of patients at injury was 51.8 years (s.d. = 19.5), with a median of 52 years (range 12–90). A total of 184 (67.9%) patients were male and 87 (32.1%) were female. A preoperative CT-scan was available on all patients and 141 (52%) out of 271 had an MRI of the spine.

About one-third of the 283 injuries (36%) were located in the cervical region, 91 (32%) in the thoracic region and 89 (31%) in the lumbar region, with total agreement by the two raters.

Across all four readings, an average of 37 (16%) of 271 patients were classified as having multiple level injuries, 47 (16%) of 283 injuries as having traumatic translation (missing data for 8 ratings) and 211 (74%) as having PLC (missing data for 7 ratings). All patients were classified as having a spinal column injury across all four readings; thus, this variable, although being one of the radiological variables of the SCI Data Set, was excluded by statistical analysis. According to the AO classification, the average distribution across all four readings of the 259 injuries showed the majority of cases classified as B1 and B2 (Figures 1 a–d).

Rater bias

After the first reading, the distribution of classifications differed significantly between raters. Proportion of cases classified as multiple vs single injuries was 10% and 14% for rater 1 and 2, respectively ($P = 0.025$), whereas proportion of cases classified as having PLC vs not or unknown was 71% and 78%, respectively ($P < 0.001$), and the distributions of the AO classification were significantly different ($P < 0.001$). After the second reading, the proportion of cases classified as having PLC vs not or unknown was 74% and 78% for rater 1 and 2, respectively ($P = 0.007$), and the distributions of the AO classification likewise still differed significantly ($P = 0.004$). For rater 1, the distribution of classifications differed significantly between the two
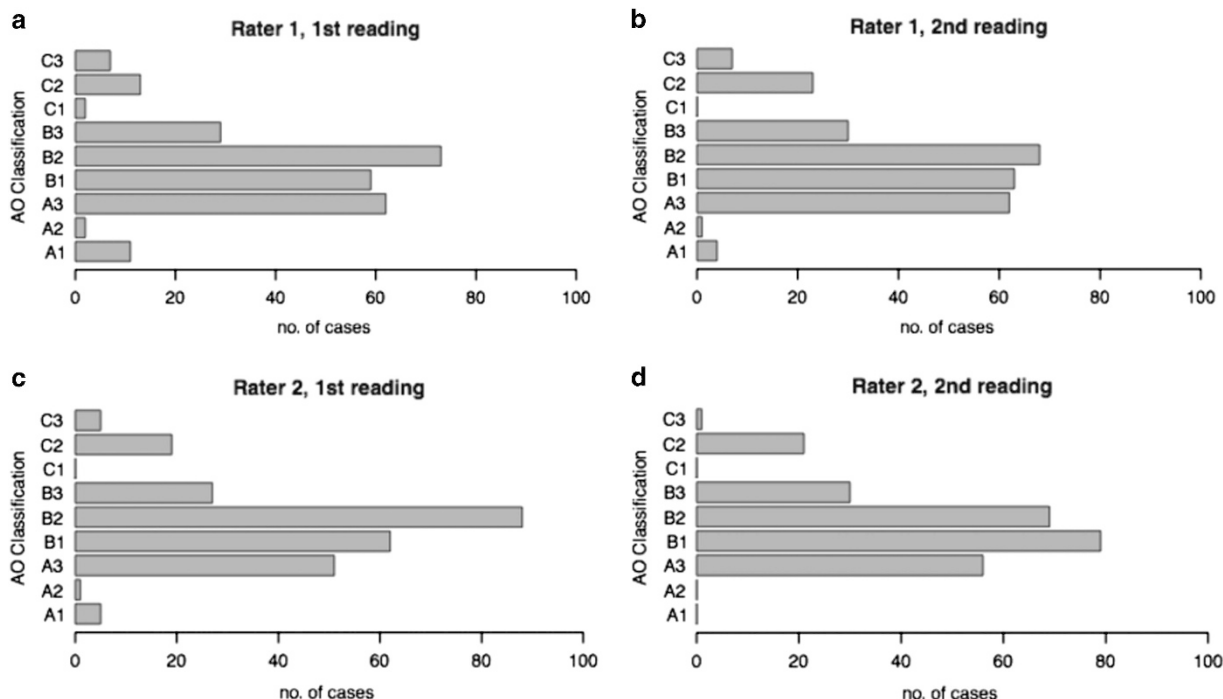
Figure 1 (a–d) AO classification distribution in the four readings.

readings as proportion of cases classified as multiple vs single injuries was 11 and 13% for the first and second reading, respectively ($P=0.02$), and further the distributions of AO classifications differed ($P=0.001$). The distributions of classifications by rater 2 only differed significantly for the AO classification ($P=0.002$). Bhapkar's tests for differences in distribution for spinal column injury level and DS entire classification produced errors, but the distributions of the remaining variables did not differ significantly between raters or between first and second reading for each rater.

Interrater agreement

After the first reading, the two raters agreed on each of the classifications of the International SCI Spinal Column Injury Basic Data Set variables in 88–96% of cases ($\kappa=0.67$–0.88), they agreed on all four variables (DS entire classification) in 73% ($\kappa=0.71$) and the AO classifications were the same in 76% ($\kappa=0.69$). After the second reading, they agreed on each of the International SCI Spinal Column Injury Basic Data Set variables in 87–99% of cases ($\kappa=0.82$–0.97), the DS entire classification was the same in 79% ($\kappa=0.79$) and the AO classifications were consistent in 74% ($\kappa=0.67$) (Tables 2 and 3).

Intrarater agreement

Ratings for the four radiological International SCI Spinal Column Injury Basic Data Set variables were the same on both readings in 89–97% of the cases ($\kappa=0.82$–0.89) for rater 1 and in 85–95% ($\kappa=0.79$–0.84) for rater 2. The same rating for all four variables (DS entire classification) was observed in both readings in 78% ($\kappa=0.77$) for rater 1 and 73% ($\kappa=0.72$) for rater 2, and the AO classifications were the same in 80% ($\kappa=0.75$) for rater 1 and 66% ($\kappa=0.57$) for rater 2 (Tables 2 and 3).

Comparison of the DS entire classification and the AO classification

There were no significant differences in the proportions of cases with crude inter- or intrarater agreement between the two classification systems, with P-values ranging 0.067–0.895 (Table 2).

DISCUSSION

The present study shows that reliability of the radiographic variables in the International SCI Spinal Column Injury Basic Data Set is substantial to almost perfect and comparable to the rater agreement of the AO classification.

Strength of this study is that a considerable number of cases were included compared with previous interobserver studies. Previous studies have included a limited number of cases, often combined with a large number of observers.[5,6]

Furthermore, the patient population is representative and consecutive, as our centre is the only one serving the Eastern part of Denmark with surgical treatment of patients with spinal injuries. This could also represent a limitation, as all patients underwent surgical treatment, resulting in skewness towards worse classifications than a population including both surgically and non-surgically treated patients. The prevalence of disc injuries and posterior ligamentous injuries was 74%, supporting this assumption.

Interpreting the results from this study, certain limitations have to be considered. First, we had only two raters. The clinical experience of the raters, although coming from different countries and with different background, was almost the same (two spine fellows), and we did not compare their ratings with those of more or less experienced physicians in spine clinical practice. Our sample of patients was selected from a database of surgically treated fractures, with the implications that this could have in using classification systems created for surgically and non-surgically treated injuries. Radiological assessment of the fracture was obtained with different methods, because not in all cases spine MRI was available; this could have influenced the rating of some of the parameters included in our study. The AO classification has two separate systems for cervical and thoracolumbar spine fractures, whereas the International SCI Spinal Column Injury Basic Data Set includes all levels. We considered in our analysis the two AO systems as unique, although based on the same principles. A detailed analysis of the differences between the cervical

Table 2 Crude agreement

	Single vs multiple injuries (%)	Spinal column injury level (%)	Disc/PLC (%)	Traumatic translation (%)	DS entire classification (%)	AO classification (%)	P- value[a]
Interrater agreement							
Reading 1	93	88	89	96	73	76	0.233
Reading 2	99	87	93	97	79	74	0.067
Intrarater agreement							
Rater 1	97	89	93	96	78	80	0.895
Rater 2	95	85	93	95	73	66	0.104

[a]McNemar's test for differences in crude agreement rates between DS entire classification and AO classification.

Table 3 Cohen's unweighted κ coefficients (95% confidence interval) for agreement

	Single vs multiple injuries	Spinal column injury level	Disc/PLC	Traumatic translation	DS entire classification[a]	AO classification
Interrater agreement						
Reading 1	0.67 (0.55–0.78)[b]	0.88 (0.85–0.90)	0.71 (0.61–0.82)[b]	0.84 (0.72–0.96)	0.71 (0.69–0.73)	0.69 (0.63–0.75)[b]
Reading 2	0.97 (0.85–1.00)	0.86 (0.83–0.89)	0.82 (0.72–0.93)[b]	0.90 (0.78–1.00)	0.79 (0.77–0.81)	0.67 (0.61–0.73)[b]
Intrarater agreement						
Rater 1	0.85 (0.74–0.97)[b]	0.89 (0.86–0.92)	0.82 (0.71–0.93)	0.87 (0.76–0.99)	0.77 (0.75–0.80)	0.75 (0.69–0.81)[b]
Rater 2	0.80 (0.69–0.92)	0.84 (0.81–0.87)	0.79 (0.69–0.90)	0.81 (0.70–0.93)	0.72 (0.70–0.74)	0.57 (0.50–0.63)[b]

[a]Test for rater bias not possible.
[b]Significant rater bias.

and the thoracolumbar AO classification systems is beyond the scope of the study; however, we did not notice a significant difference in reliability between these subgroups (data not shown). Moreover, from a statistical point of view, combining multiple categories in a single κ-value can have some bias, which we have to consider in interpreting the results.[12]

A number of different classification systems for spinal column injuries have been proposed as tools to guide diagnosis and treatment and facilitate data collection and communication among different centres and countries, but an optimal and an universally accepted classification has not yet been established. The complex anatomy of each vertebra and the variations between vertebrae in different parts of the spine itself make simple classifications of fractures very challenging.[7]

The previously proposed classification systems for traumatic spine injuries usually involved specific spine levels (cervical or thoracolumbar spine) and did not represent a complete overview of the extremely wide spectrum of injuries.

One of the most used classification system in the last two decades is the AO classification scheme proposed by Magerl et al.[1] in 1994 for thoracolumbar spine fractures, reviewing 1445 patients. The scheme is based on the morphologic appearance on radiographs, mechanism and severity of injury and proposes three types of injury, each one further divided into groups and subgroups. An extension of this scheme proposed by Blauth et al.[2] based on a scheme previously published by Aebi and Nazarian[13] has been adopted for subaxial cervical injuries. Although proposed later as a separate entity, the cervical system has the same basic principles of its counterpart in thoracolumbar spine, dividing compression, distraction and rotation injuries. The AO classification schemes have guided spine surgeons in decision making in different centres worldwide, with its wide range of spine injury types included. AO classification allows categorisation of injuries to the most relevant parts of the spine, providing categories for all types of injury patterns. This increases the complexity of the classification, making it sometimes difficult to use for communication among spine trauma physicians.

An international committee of experts in 2012 introduced the International SCI Spinal Column Injury Basic Data Set with the aim of facilitating 'comparisons of spinal column injuries among studies, centres and countries'.[7]

The role of an international classification/data set is to compress available information into categories without loss of significant content, and the cornerstone for this aim is reliability. This is of paramount importance to allow comparison of data from different clinicians, facilities or countries. The problem of reproducibility for spine fracture classifications has been widely reported in the literature.

In a study by Oner et al.[14] including 53 patients with thoracolumbar fractures, the reliability of the Magerl AO classification was tested by five raters. The interrater agreement using MRI was moderate for the type (A, B and C) classification of the AO scheme and substantial for the subclassification of the types. The type classification of AO system was only fairly reproducible using CT-scans, compared with moderate using MRI. The authors focus on the importance of MRI as a diagnostic tool to enhance the depiction of posterior ligament injuries, which has a paramount role in the classification process. The examination of MRI may have a role in detecting disc/PLC injuries. MRI can show subtle ligamentous lesion, given its higher sensitivity, specificity and accuracy in detecting these lesions compared with CT, but with low reliability.[4]

A study by Wood et al.,[5] classifying the AO type (A, B, C) of 31 thoracolumbar fractures among 19 spine surgeons, showed a κ-value of 0.48 for interobserver and 0.63 for intraobserver reliability.

Lenarz et al.[15] reported on the AO and thoracolumbar injury severity score[4] reliability, rating 97 thoracolumbar fractures (diagnosed with plain-film radiographs and CT-scan) by 21 raters with different degrees of clinical experience (junior residents, senior residents, non-spine attending, spine fellows, spine attending). For the AO classification (only types rating A, B, C), they found κ values for

interobserver reliability ranging from 0.52 to 0.77, corresponding to a moderate-to-substantial agreement according to Landis and Koch,[8] with higher values in groups routinely facing spine trauma.

Recently, a group of spine surgery experts proposed a new AO classification system for thoracolumbar spine injury.[4] This system, not yet available at the beginning of the present study, includes a morphologic classification of the fracture, a grading of the neurological status and a description of relevant modifiers. This new classification follows the principle of the original one, focusing on injury of the PLC as the key differentiating factor among severity groups. The reliability test for the identification of a morphologic injury type (A, B and C) showed a κ-value of 0.72 (substantial reliability). Compared with this comprehensive system, which aims to be a classification system to improve communication between clinicians and optimise treatment algorithms for clinical use,[4] the International SCI Spinal Column injury Basic Data Set is intended as a 'tool to facilitate standardisation of the data collection process',[7] including the minimal amount of useful basic descriptive information.

Recommendations for the reliability testing of international SCI Data Sets have previously been proposed.[16] Testing the classification scheme with spine fellows from different countries and different background clinical experience may represent an important step in developing international data sets.

In our population sample, the gender distribution was similar to other previously reported epidemiologic studies on spine fractures.[15,17]

In our study, the availability of MRI (52% of patients) could have influenced the detection of disc/PLC injury and the type rating of the AO scheme, but this was the same for both raters. Regarding disc/PLC injury classification, post-hoc analyses revealed a somewhat better average crude interrater agreement (96,vs 86%), but no difference in reliability when MRI was available (data not shown). Future studies, including a consecutive series of patients, will be necessary to better assess the role of MRI.

For the AO classification system, we found a substantial interrater agreement in both readings and a moderate-to-substantial intrarater agreement, respectively, for the two raters (Table 3), which is comparable to the previously reported agreements. However, we have to consider that in our study we classified both type and groups of fractures (for totally nine categories), differently from previously reported studies, in which only type was rated.

For the International SCI Spinal Column Injury Basic Data Set, we found a substantial to almost perfect inter- and intrarater agreement for the two raters in each of the four radiographic variables of the International SCI Spinal Column Injury Basic Data Set. Substantial inter- and intrarater agreement for the four radiological Spinal Column Injury Data Set variables combined (DS entire classification) was documented.

We found no statistical significant differences in proportions of cases with crude agreement between the DS entire classification and the AO classification (Table 2).

Although the crude interrater agreement of the AO classification was higher than the DS entire classification after the first reading, the interrater κ value was lower after both the first and the second reading. The same was seen for the intrarater agreement for rater 1. For rater 2, both crude intrarater agreement and κ values were higher in the DS entire classification than in the AO (Tables 2 and 3).

In the second reading, the interrater κ coefficients showed higher values for single/multiple, translation, disc/PLC Injury and DS entire classification than the first reading. This could indicate a learning process in variable rating (Table 3).

We have to note that κ statistics are highly influenced by trait prevalence, rater bias and the number of categories included.[18]

Moreover, we used only one κ value for multicategory classification systems (DS entire classification, AO classification and spinal column injury level), which could be interpreted as a weighted average of the individual category kappas, with weights determined by the corresponding prevalences.[12]

Thus, we think it could be inappropriate to infer that there is a true difference in reliability between the two classifications system. However, according to the commonly employed measurements of interobserver agreement,[8] the International SCI Spinal Column Injury Basic Data Set showed high intra- and interobserver agreement for all its variables and for a combination of these at least as high as the commonly employed AO classification system.

In conclusion, we were able to document a high intra- and interrater reliability for the international SCI spinal column injury basic data set and, in the need of worldwide communication about the severity of spine injuries and results of treatment strategies, we encourage its use for classification of spinal column injuries.

DATA ARCHIVING
There were no data to deposit.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

1 Magerl F, Aebi M, Gertzbein SD, Harms J, Nazarian S. A comprehensive classification of thoracic and lumbar injuries. Eur Spine J 1994; 3: 184–201.
2 Blauth M, Kathrein A, Mair G, Schmid R, Reinhold M, Rieger M. Classification of injuries of the subaxial cervical spine. In: Aebi M, Arlet V, Webb JK (eds). AO Spine Manual: Clinical Applications, vol. 2. Thieme: Stuttgart, 2007, pp 21–38.
3 Vaccaro AR, Lehman RA Jr, Hurlbert RJ, Anderson PA, Harris M, Hedlund R et al. A new classification of thoracolumbar injuries: the importance of injury morphology, the integrity of the posterior ligamentous complex, and neurologic status. Spine (Phila Pa 1976) 2005; 30: 2325–2333.
4 Vaccaro AR, Oner C, Kepler CK, Dvorak M, Schnake K, Bellabarba C et al. AO Spine Spinal Cord Injury & Trauma Knowledge Forum. AO spine thoracolumbar spine injury classification system: fracture description, neurological status, and key modifiers. Spine (Phila Pa 1976) 2013; 38: 2028–2037.
5 Wood KB, Khanna G, Vaccaro AR, Arnold PM, Harris MB, Mehbod AA. Assessment of two thoracolumbar fracture classification systems as used by multiple surgeons. J Bone Joint Surg Am 2005; 87: 1423–1429.
6 Whang PG, Vaccaro AR, Poelstra KA, Patel AA, Anderson DG, Albert TJ et al. The influence of fracture mechanism and morphology on the reliability and validity of two novel thoracolumbar injury classification systems. Spine (Phila Pa 1976) 2007; 32: 791–795.
7 Dvorak MF, Wing PC, Fehlings MG, Vaccaro AR, Itshayek E, Biering-Sorensen F et al. International spinal cord injury spinal column injury basic data set. Spinal Cord 2012; 50: 817–821.
8 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159–174.
9 Gamer M, Lemon J, Puspendra IF. irr: various coefficients of interrater reliability and agreement. Available at http://cran.r-project.org/package=irr. Accessed 20 November 2013.
10 Warnes GR, Bolker B, Lumley T, Johnson RC. gmodels: various R programming tools for model fitting. Available at http://cran.r-project.org/package=gmodels. Accessed 20 November 2013.
11 Core Team RR. A language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2014.
12 Kraemer HC, Periyakoil VS, Noda A. K statistics in medical research. Stat Med 2002; 21: 2109–2129.
13 Aebi M, Nazarian S. Classification of injuries of the cervical spine. Orthopaede 1987; 16: 27–36.
14 Oner FC, Ramos LM, Simmermacher RK, Kingma PT, Diekerhof CH, Dhert WJ et al. Classification of thoracic and lumbar spine fractures: problems of reproducibility. A study of 53 patients using CT and MRI. Eur Spine J 2002; 11: 235–245.
15 Lenarz CJ, Place HM, Lenke LG, Alander DH, Oliver D. Comparative reliability of 3 thoracolumbar fracture classification systems. J Spinal Disord Tech 2009; 22: 422–427.
16 Biering-Sørensen F, Alexander MS, Burns S, Charlifue S, DeVivo M, Dietz V et al. Recommendations for translation and reliability testing of international spinal cord injury data sets. Spinal Cord 2011; 49: 357–360.
17 Leucht P, Fischer K, Muhr G, Mueller EJ. Epidemiology of traumatic spine fractures. Injury 2009; 40: 166–172.
18 Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005; 85: 257–268.