



External Multicenter Study of Reliability and Reproducibility for Lower Cervical Spine Injuries Classification Systems—Part 2: An Analysis of the Subaxial Cervical Spine Injury Classification and Cervical Spine Injury Severity Score Scale

Global Spine Journal
2021, Vol. 11(1) 99-107
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2192568219896546
journals.sagepub.com/home/gsj



Andrey Grin, Professor, MD, PhD^{1,2}, Vladimir Krylov, Professor, MD, PhD^{1,2}, Ivan Lvov, MD, PhD¹ , Aleksandr Talypov, MD, PhD¹, Dmitriy Dzukaev, MD³, Anton Kordonskiy, MD, PhD¹, Vladimir Smirnov, MD¹, Vasily Karanadze, MD¹, Boburmirzo Abdukhalikov, MD¹, Ulugbek Khushnazarov, MD¹, Artem Airapetyan, MD, PhD¹, Aleksandr Dmitriev, MD, PhD¹, Aleksandr Kaykov, MD, PhD¹, Alexander Peyker, MD³, Vitaliy Semchenko, MD³, Andrey Aksenov, MD³, Anton Borzenkov, MD³, Vladimir Gulyy, MD³, Soslan Torchinov, MD³, Sergey Bagaev, MD³, Anton Toporskiy, MD³, Alik Kalandari, MD, PhD², Denis Kasatkin, MD, PhD², Aleksey Sytnik, MD⁴, Valeriy Lebedev, MD⁵, Dmitriy Epifanov, MD⁵, Dmitriy Hovrin, MD, PhD⁶, Victor Feniksov, MD, PhD⁷, and Daniyar Choriev, MD⁷

Abstract

Study Design: A multicenter observational survey.

Objective: To quantify and compare inter- and intraobserver reliability of the subaxial cervical spine injury classification (SLIC) and the cervical spine injury severity score (CSISS) in a multicentric survey of neurosurgeons with different experience levels.

Methods: Data concerning 64 consecutive patients who had undergone cervical spine surgery between 2013 and 2017 was evaluated, and we surveyed 37 neurosurgeons from 7 different clinics. All raters were divided into 3 groups depending on their level of experience. Two assessment procedures were performed.

Results: For the SLIC, we observed excellent agreement regarding management among experienced surgeons, whereas agreement among less experienced neurosurgeons was moderate and almost twice as unlikely. The sensitivity of SLIC relating to treatment tactics reached as high as 92.2%. For the CSISS, agreement regarding management ranged from medium to substantial,

¹ Sklifosovsky Research Institute of Emergency Care, Moscow, Russia

² Evdokimov Moscow State University of Medicine and Dentistry, Moscow, Russia

³ Moscow Spine Center of City Hospital No. 67, Moscow, Russia

⁴ Moscow City Hospital No. 13, Moscow, Russia

⁵ Pirogov National Medical and Surgical Center, Moscow, Russia

⁶ Moscow City Hospital No. 7, Moscow, Russia

⁷ Moscow City Hospital No. 1, Moscow, Russia

Corresponding Author:

Ivan Lvov, Sklifosovsky Research Institute of Emergency Care, B. Suharevskaya Pl. 3, Moscow 107945, Russia.

Email: dr.speleolog@gmail.com



Creative Commons Non Commercial No Derivs CC BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-Non Commercial-NoDerivs 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

depending on a neurosurgeon's experience. For less experienced neurosurgeons, the level of agreement concerning surgical management was the same as for the SLIC in not exceeding a moderate level. However, this scale had insufficient sensitivity (slightly exceeding 50%). The reproducibility of both scales was excellent among all raters regardless of their experience level.

Conclusions: Our study demonstrated better management reliability, sensitivity, and reproducibility for the SLIC, which provided moderate interrater agreement with moderate to excellent intraclass correlation coefficient indicators for all raters. The CSISS demonstrated high reproducibility; however, large variability in answers prevented raters from reaching a moderate level of agreement. Magnetic resonance imaging integration may increase sensitivity of CSISS in relation to fracture management.

Keywords

SLIC, CSISS, cervical spine trauma, interobserver agreement, intraobserver agreement, intraclass correlation

Introduction

Point scales for measuring cervical stability after injury were first developed by White and Panjabi in 1990.¹ Based on the x-ray data and clinical assessment, the degree of damage to the anterior and posterior elements, level of dislocation and rotation, neurological deficit, disc injury, and results of stretch test were measured in points. This scale was not widely used because of the broad application of computed tomography (CT) and magnetic resonance imaging (MRI) in routine practice. A retrospective study was also inconvenient because of the occasional use of the stretch test in patients with mild or controversial spine injuries. Moreover, this procedure is incompatible with current standards of medical care in patients with acute cervical injuries.² After 2006, the Subaxial Cervical Spine Injury Classification (SLIC) and Cervical Spine Injury Severity Score (CSISS) scales were developed for point measuring of cervical stability after injury.^{3,4} Although these scales were developed more than 10 years ago, they have never been used widely despite the fact that the effectiveness of the SLIC has been proven in clinical work.^{5,6} According to Chhabra et al,⁷ only 35% of experts use the SLIC scale, and 5% use the CSISS scale daily.

Only 5 published articles have estimated the reliability of the SLIC and CSISS scales,^{3,4,8-10} and only 3 of those were external studies.⁸⁻¹⁰ The CSISS scale has been examined by surgeons with different levels of experience; however, all studies had a single-centre design (Table 1). The SLIC scale has been examined by only experienced surgeons. The reliability and reproducibility of this scale have never been estimated for residents and young surgeons.

Therefore, our study aimed to measure and compare the inter- and intraobserver reliabilities of the SLIC and CSISS scales among neurosurgeons who have different levels of experience and work in different clinics.

Materials and Methods

Patient Cohort

We used data of 64 consecutive patients who underwent surgery between January 2013 and December 2017. All patients underwent surgery in the study initiator's institute. This study was approved by an institutional review board.

We used anonymized CT (all cases) and MRI (58%) data in our study. In 22 patients, operated on in 2013 and 2014, cervical spine CT was performed with a slice thickness of 2 mm. In the remaining 42 cases, CT was performed with a slice thickness of 0.5 mm. Personalized information was removed from the Digital Imaging and Communications in Medicine (DICOM) files, and each case was assigned a unique number. Each case also had a description of the patient's neurological status. Each rater's folder included all 64 cases in a random order—this ruled out the risk of raters copying each other's answers.

Raters

Thirty-seven surgeons from 7 different clinics were included in the study. Five clinics were level 1 trauma centers, and 2 were university clinics. All raters were divided into 3 groups depending on their level of experience. The group of beginners (20 surgeons) included residents, nonspinal neurosurgeons, and junior spinal surgeons with less than 5 years of experience. This division was due to the peculiarities of the residency program in our country, which only lasts 2 years. In relation to Europe and the United States, our beginners group correspond to final-year residents or surgeons with 1 to 2 years of experience, that is, surgeons who perform some operations under the periodic supervision of a more experienced surgeon. The intermediate group comprised 10 neurosurgeons with between 5 and 10 years of experience in spinal surgery and who were able to independently perform all types of anterior surgery of the cervical spine, and who had participated in multiple surgeries. The experienced group comprised 7 surgeons with >10 years of experience in spinal surgery, and who were able to perform all types of anterior and posterior surgery for cervical spine trauma.

Each rater received a unique package that included the following: (1) a USB flash drive with the blinded DICOM archive and information concerning patients' neurological status, (2) a reference material booklet containing a detailed description of each studied classification, and (3) an application form to fill in the answers.

Assessment Process

Two assessment procedures were performed. The initial assessment included 36 raters from all clinics. Then the order and

Table 1. Current Studies of the SLIC and CSISS Scales.

Study	Scale	Number of Cases	Patient Cohort	Number of Raters	Types of Raters	Number of Participating Centers
Anderson et al, ³ 2007	CSISS	34	Selected from database	15	Residents Fellows Attending surgeons Radiologist	Single center
Zehnder et al, ¹⁰ 2009	CSISS	50	Selected from database	15	6 junior residents 6 senior residents 3 spine attendings	Single center
Middendorp et al, ⁸ 2013	SLIC	51	Selected from database	12	5 spine surgeons 7 orthopedic/trauma surgeons	Multiple centers
Stone et al, ⁹ 2010	SLIC CSISS	50	Consecutive patients	5	Unknown	Single center
Vaccaro et al, ⁴ 2007	SLIC	11	Selected from database	20	5 neurosurgeons 15 orthopedic surgeons	Multiple centers
Present study	SLIC CSISS	64	Consecutive patients	37	20 beginners and nonspinal neurosurgeons 10 intermediate spine surgeons 7 experienced spine surgeons	Multiple centers

Abbreviations: SLIC, Subaxial Cervical Spine Injury Classification; CSISS, Cervical Spine Injury Severity Score.

serial numbers of cases were randomly changed. The second assessment was conducted in each clinic separately within 1.5 months after the first assessment. Twenty-four raters were included in the second assessment.

Description of the Scales

The SLIC scale⁴ is a simultaneous assessment of the morphology of damage, integrity of the discoligamentous complex, and degree of damage to neural structures. If the total score is 5 or more, then surgical treatment is indicated. With a score of 3 or less, conservative therapy is indicated. With a score of 4, the choice of treatment is based on the surgeon's experience and consideration of the patient's condition and individual characteristics.

The main purpose of the CSISS is to determine the surgical tactic at the stage of CT. The authors identified 4 anatomical zones.³ Bone damage was evaluated according to the amount of displacement of the fragments. Ligament damage was based on the degree of divergence of the corresponding bone landmarks: 0 points, no damage; 1 point, damage without displacement; and 5 points, the maximum possible damage at this level. Intermediate points of damage to the ligamentous apparatus were set in accordance with the surgeon's opinion.

Statistical Analysis

We used the Fleiss kappa and intraclass correlation coefficient (ICC) indicators to estimate the interrater reliability. For the kappa statistic (K) calculation, we used Microsoft Excel 2011 for Mac (Microsoft Corp) with the VBA program AgreeStat 2015.6 (Advanced Analytics, LLC).

The interrater ICC was calculated using SPSS Statistics 23.0 for Mac (IBM Corp) and 2 formulas.^{11,12} For major

comparisons, we used a 2-way random model for ICC (ICC 2.1) assessment since the raters were assumed to be representative of other raters and that those other raters would produce a similar kind of consistency regardless of the participants. We used absolute agreement, as this was the only way to demonstrate the identity of cases to each other. The ICC form was a single measure, that is, the single reading of 1 rater was compared to a single reading of another rater. In order to compare our data with those in the literature, we analyzed the ICC using the 2-way random model, average measures, and absolute agreement (ICC 2, κ). Since each rater performed only 1 assessment in each stage of assessments, we averaged all the results in the first and second assessments.

Intrarater reliability was estimated for each rater using Cohen's kappa and ICC (3, 1; 2-way mixed model, single measures, and absolute agreement). The kappa statistic was interpreted using Landis and Koch's system¹³; if the kappa was less than 0.2, the degree of agreement was estimated as slight; if between 0.2 and 0.4, the degree of agreement was fair; if between 0.4 and 0.6, the degree of agreement was moderate; if between 0.6 and 0.8, the degree of agreement was substantial; if more than 0.8, the degree of agreement was excellent. During assessment of the ICC, the correlation was poor if the ICC was less than 0.50, moderate if between 0.50 and 0.75, good if between 0.75 and 0.90, and excellent if more than 0.90.

Results

The overall number of processed application forms was 60 (36 forms during the first assessment, and 24 forms during the second assessment). The included raters performed 3840 evaluations, and we received 7616 answers. In 21 cases, raters were unable to diagnose the injury, and in 11 cases, raters were unable to visualize the files because of flash drive failure.

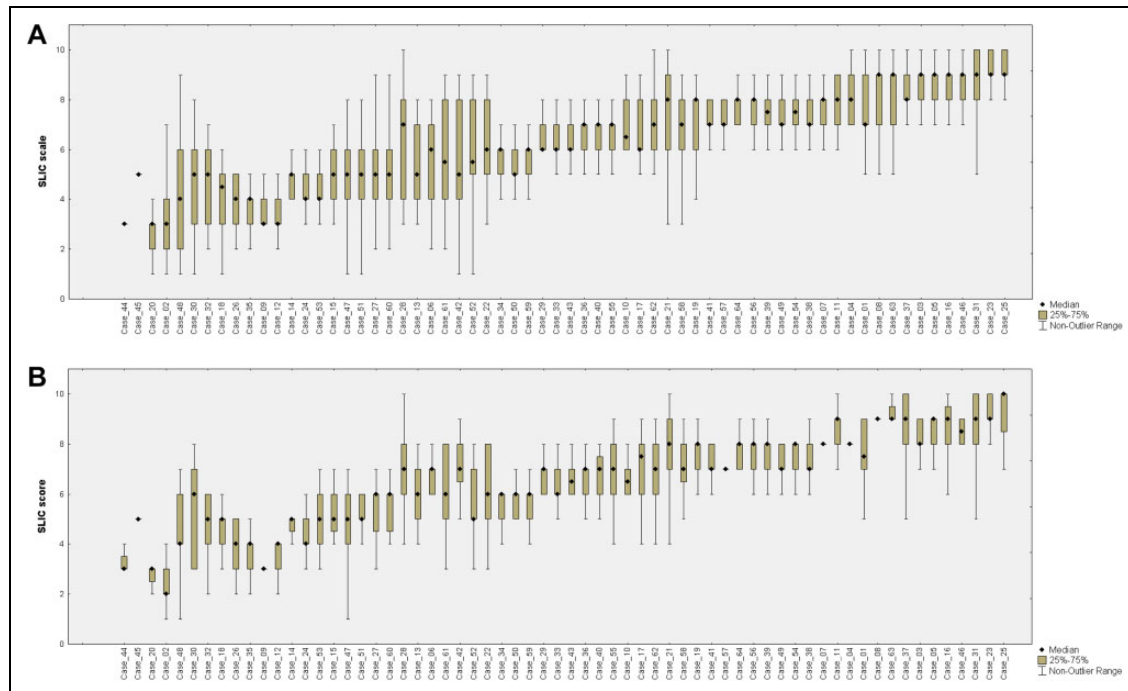


Figure 1. Variability of Subaxial Cervical Spine Injury Classification values during the first (a) and second (b) stages of the study.

Table 2. Interrater and Intrarater Agreement in the Beginners and Experienced Surgeon's Groups for SLIC Scale.

	First Assessment		Second Assessment	
	Beginners Group	Experienced Group	Beginners Group	Experienced Group
Interrater agreement				
ICC 2.1	0.66	0.75	0.64	0.85
ICC 2.k	n/a	n/a	0.95	0.96
Overall Fleiss kappa	0.14	0.38	0.19	0.46
Fleiss kappa, CT only	0.14	0.36	0.21	0.43
Fleiss kappa, CT + MRI	0.13	0.36	0.11	0.47
Management, kappa	0.43	0.66	0.46	0.85
Intrarater agreement				
Cohen's kappa		n/a	0.77	0.68
ICC 3.1		n/a	0.84	0.90

Abbreviations: SLIC, Subaxial Cervical Spine Injury Classification; CT, computed tomography; ICC, intraclass correlation; ICC 2.1, intraclass correlation, 2-way random, single measures, absolute agreement; ICC 2.k, intraclass correlation, 2-way random, average measures, absolute agreement; ICC 3.1, intraclass correlation, 2-way mixed model, single measures, absolute agreement; MRI, magnetic resonance imaging; n/a, not available.

SLIC Scale

Interobserver Agreement. The raters' primary assessments were highly variable (Figure 1a), and the estimated values ranged between 3 and 8 points. However, we observed a moderate correlation among all raters (ICC 2.1 = 0.57; 95% confidence interval [CI], 0.49-0.67) and good correlation among experienced surgeons (ICC 2.1 = 0.75; 95% CI, 0.67-0.82; Table 2). The variability of answers was much lower during the second assessment (Figure 1b) and this affected the ICC indicator. However, ICC 2.1 was almost the same in the groups of beginners and intermediates (ICC 2.1 = 0.64; 95% CI, 0.54-0.73 and ICC 2.1 = 0.62; 95% CI, 0.51-0.72, respectively). At the same

time, ICC 2.1 was almost excellent in the group of experienced surgeons (ICC = 0.85; 95% CI, 0.79-0.90).

During the first assessment, the interobserver kappa of all raters reached 0.15 (95% CI, 0.13-0.17) with the best values in the group of experienced surgeons (K = 0.38; 95% CI, 0.34-0.42). The second assessment demonstrated the increase of the overall K indicator up to 0.22 (95% CI, 0.19-0.25) and up to moderate in the group of experienced surgeons (K = 0.46; 95% CI, 0.38-0.55). Agreement for the beginners' group was slight and did not exceed 0.19 for both assessments (Table 2). In the group of intermediate surgeons, agreement was fair for the first and second assessments (K = 0.22; 95% CI, 0.18-0.26 and K = 0.25; 95% CI, 0.21-0.29, respectively).

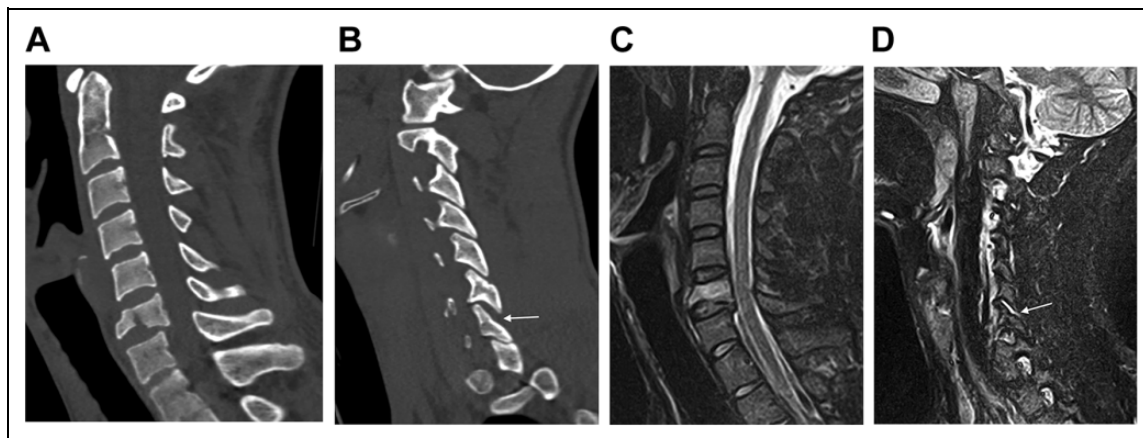


Figure 2. Case 20. Computed tomography scans (a, b) reveal a fracture of the C6 vertebral body in the frontal plane with expansion of the C6-C7 facet joint on the right side (indicated by an arrow). Magnetic resonance imaging in the short T1 inversion recovery sequence reveals damage to the upper wall of the fibrous ring of the C6-C7 disc in the fracture area (c) and fluid accumulation in the cavity of the right C6-C7 joint (d), which is indicated by an arrow. Median values for the Subaxial Cervical Spine Injury Classification are 3 points during both assessments.

The use of MRI did not significantly affect the interpretation of the injury; however, in the group of patients where both CT and MRI were used, there was a clear tendency for the agreement coefficient to decrease, with a difference in values from 0.03 to 0.05 depending on the raters' level of experience (Table 2).

Reproducibility. The intraobserver agreement and correlation were excellent for the SLIC scale among all raters ($K = 0.80$ [range 0.49-0.98] and $ICC = 0.90$ [range 0.65-0.98]).

Average Values. We averaged the values for each studied patient during the first and second assessments. $ICC\ 2, \kappa$ was excellent for all surgeons (0.98; 95% CI, 0.96-0.99).

Management Based on the SLIC. The level of agreement among all raters in management based on the SLIC during both assessments was moderate ($K = 0.42$; 95% CI, 0.30-0.54 and $K = 0.49$; 95% CI, 0.34-0.64, respectively). However, it was substantial and excellent in the group of experienced surgeons ($K = 0.66$; 95% CI, 0.52-0.80 and $K = 0.85$; 95% CI, 0.72-0.98, respectively; Table 2). Agreement for the beginners and intermediate groups was moderate and did not reach $K = 0.55$ during both assessments.

SLIC Sensitivity in Relation to Surgical Tactics. Of all the studied cases, only 5 cases (02, 09, 12, 20, and 44) had a median value of 3 points or less during one of the assessments (Figure 1). This observation was associated with compression fractures of the vertebral bodies of type A2 or A4 according to the AOSpine classification,¹⁴ without neurological deficit. Thus, the sensitivity of the SLIC scale was 92.2%. Of the features, it should be noted that MRI was performed in 3 cases (02, 09, and 20), and CT imaging was of poor quality in 2 cases (09 and 12). Examples of such cases are presented in Figure 2.

CSISS Scale

Interobserver Agreement. Both assessments demonstrated a comparatively high level of variability in the CSISS estimation, and in some cases, the difference between assessments reached 10 points (Figure 3). During the first assessment, we observed moderate correlation between the answers of all raters ($ICC\ 2.1 = 0.56$; 95% CI, 0.48-0.66) and a slight agreement for each separate assessment ($K = 0.10$; 95% CI, 0.08-0.12). During the second assessment, we observed no significant differences in the results: slight agreement ($K = 0.11$; 95% CI, 0.09-0.13) and moderate correlation ($ICC\ 2.1 = 0.50$; 95% CI, 0.41-0.60).

In the group of experienced surgeons (Table 3), moderate correlation ($ICC\ 2.1 = 0.65$; 95% CI, 0.54-0.76) was accompanied by fair agreement ($K = 0.34$; 95% CI, 0.30-0.38). For the beginner surgeons group, agreement was slight with moderate correlation during the first ($K = 0.14$; 95% CI, 0.11-0.18 and $ICC\ 2.1 = 0.62$; 95% CI, 0.57-0.67) and second ($K = 0.12$; 95% CI, 0.09-0.15 and $ICC\ 2.1 = 0.53$; 95% CI, 0.49-0.56) assessments. Agreement for the intermediate group was also slight during the first and second assessments ($K = 0.13$; 95% CI, 0.10-0.16 and $K = 0.12$; 95% CI, 0.09-0.15, respectively) with the lowest values of intraclass correlation ($ICC\ 2.1 = 0.50$; 95% CI, 0.45-0.55 and $ICC\ 2.1 = 0.41$; 95% CI, 0.37-0.45, respectively).

The use of MRI did not significantly affect the interpretation of the injury (Table 3). In the group of patients where both CT and MRI were used, there was a clear tendency for the agreement coefficient to decrease for beginners and intermediates. For experienced surgeons, the use of MRI increased the interrater kappa to 0.1 and 0.03 during the first and second assessments, respectively.

Reproducibility. The overall intraobserver agreement level was substantial ($K = 0.76$; range 0.39-0.98) accompanied by excellent correlation ($ICC = 0.97$; range 0.92-0.99). The analysis of

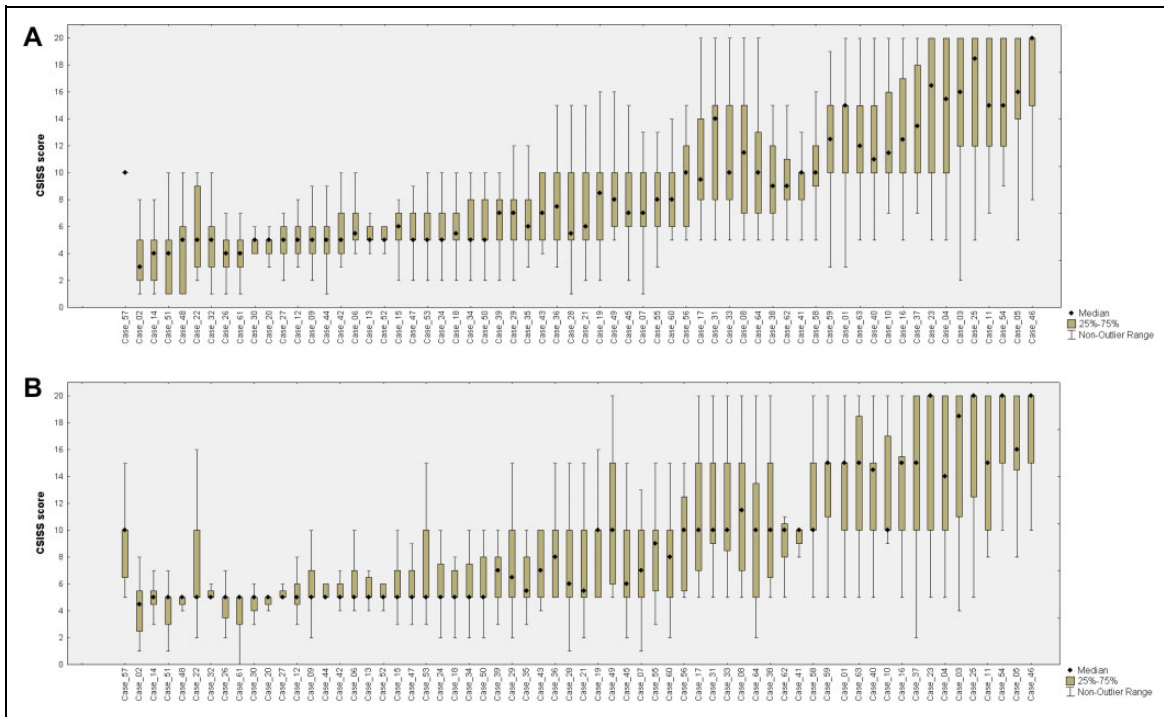


Figure 3. Variability of Cervical Spine Injury Severity Score values during the first (a) and second (b) stages of the study.

Table 3. Interrater and Intrarater Agreement in the Beginners and Experienced Surgeon’s Groups for CSISS Scale.

	First Assessment		Second Assessment	
	Beginners Group	Experienced Group	Beginners Group	Experienced Group
Interrater agreement				
ICC 2.1	0.62	0.65	0.53	0.72
ICC 2.k	n/a	n/a	0.93	0.91
Overall Fleiss kappa	0.14	0.34	0.12	0.32
Fleiss kappa, CT only	0.13	0.27	0.11	0.28
Fleiss kappa, CT + MRI	0.15	0.37	0.10	0.31
Management, kappa	0.50	0.62	0.41	0.57
Intrarater agreement				
Cohen’s kappa		n/a	0.70	0.72
ICC 3.1		n/a	0.97	0.97

Abbreviations: CSISS, Cervical Spine Injury Severity Score; CT, computed tomography; ICC, intraclass correlation; ICC 2.1, intraclass correlation, 2-way random, single measures, absolute agreement; ICC 2.k, intraclass correlation, 2-way random, average measures, absolute agreement; ICC 3.1, intraclass correlation, 2-way mixed model, single measures, absolute agreement; MRI, magnetic resonance imaging; n/a, not available.

surgeon groups depending on the level of experience also demonstrated excellent reproducibility (Table 3).

Average Values. Values averaged (ICC 2, κ) for the first and second assessments demonstrated excellent correlation both in the overall and group analyses. ICC 2, κ reached 0.96 (95% CI, 0.94-0.97) among all raters.

Management Based on the CSISS. The agreement in management based on the scale was moderate among all raters ($K = 0.41$; 95% CI, 0.33-0.48) and substantial among experienced surgeons ($K = 0.62$; 95% CI, 0.52-0.73) during the first

assessment. This indicator was lower almost in all surgeon groups during the second assessment (Table 3).

CSISS Sensitivity in Relation to Surgical Tactics. Of 64 cases, according to the CSISS scale, 29 cases had a median value of 6 points or less during both assessments. Most often, a value of less than 7 points was found in cases of compression fractures with the compression-extension mechanism of trauma, distraction injuries of the ligamentous complex without displacement (type B3 according to the AOSpine classification, Figure 4), traumatic disc herniation, and isolated unilateral unstable facet fracture (types F2 or F4 according to the AOSpine

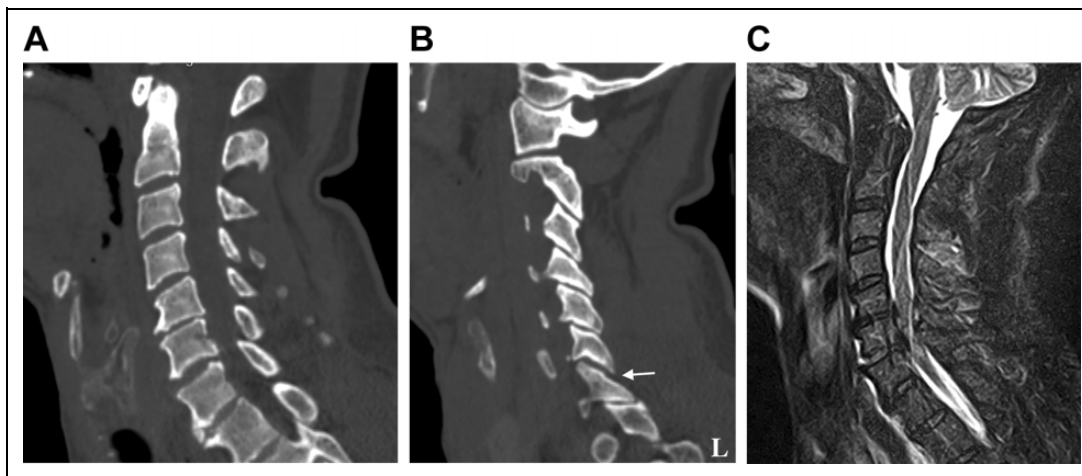


Figure 4. Case 32. Computed tomography scans demonstrate minimal retrolisthesis of C6 (a), which could be degenerative in nature, minimal diastasis for 1 to 1.5 mm of the left C6-C7 facet joint (b). Median values for the Cervical Spine Injury Severity Score are 5 points during both assessments. Magnetic resonance imaging in the short TI inversion recovery sequence (c) reveals intervertebral disc disruption with spinal cord compression.

classification). Thus, the sensitivity of the CSISS scale was 54.7%. Of 29 cases with a value of 6 or less, 20 cases (69%) had MRI of the cervical spine. This cohort also included 11 patients with CT data with a slice thickness of 2 mm.

Discussion

The major disadvantage of most morphological scales is the absence of a clear protocol for patients' management including the necessity for surgery. These scales clearly demonstrate the type of injury and dislocation in the segment but do not suggest any further management of the injury. Further tactics depend on only the knowledge of modern standards and surgeons' experience. In cases of insufficient experience in spinal surgery or in complex cases, point scales may simplify patient management determination. Therefore, in our study, we analyzed the agreement evaluation among less experienced neurosurgeons who are usually the first to see patients on admission.

The SLIC and CSISS scales were developed almost simultaneously.^{3,4} The main difference between them is that the SLIC scale estimates the injury structure in light of morphology and clinical manifestation, whereas the CSISS scale considers the CT data of the injury. The SLIC value might reach 10 points, but the CSISS value can reach 20 points. This allows for a large spread of answers under certain conditions.

Indeed, we performed the agreement assessment using not only the ICC indicator but also the kappa statistic. We believe that kappa can demonstrate the chance-corrected agreement of the specific values between all provided answers.¹⁵ Since every classification must be minimally controversial for each injury subtype characteristic, an ideal point scale suggests that all answers must be significantly similar to each other.¹⁶

In the assessment of interrater agreement in the SLIC scale, we observed a rather high level of variability in the answers during both assessments of the study (Figure 2). This resulted in low rater agreement among all specific values. We observed

a moderate level of interrater agreement only in the group of experts. The same agreement level was demonstrated in an internal study by Vaccaro et al.⁴ During assessment of the SLIC scale, most raters described the assessment of intervertebral disc injury without MRI data as well as differential diagnostics of some distraction and translational injuries as complicating. The same problems were observed by the authors of 2 aforementioned studies.^{4,8} Analysis of ICC demonstrated a moderate agreement level among all raters and good/excellent level among experts. The same results were reported in all existing articles dedicated to assessment of the SLIC scale^{4,8,9} that included experienced surgeons.

Finally, we observed moderate to excellent agreement in respect of the defined management among experienced surgeons. At the same time, agreement among less experienced neurosurgeons was found to be almost 2 times less than that of experienced surgeons, which could complicate the decision-making regarding the correct surgical strategy to adopt. Many difficulties in determining the appropriate surgical approach occurred for patients with type A2 and A4 fractures according to the AOSpine classification. In the absence of a neurological deficit in some surgically treated cases, most surgeons gave a rating of 3 points, which indicates conservative treatment based on the SLIC scale. Despite this, the sensitivity of SLIC in relation to the treatment tactic was high and reached 92.2%. In conjunction with high sensitivity, regular use of this scale in routine practice may increase the level of surgeons' agreement for a reliable and appropriate surgical strategy.

We also compared agreement coefficients in patient groups with and without MRI data. We found no significant difference, which indicated a possibility of determining which treatment to apply on the basis of CT data only. Moreover, the quality of CT images also had no effect on determining the need for further treatment, which indicated the versatility of the scale and its possible use in almost any hospital, regardless of imaging equipment.

During 2 assessments, variability of answers in analysis of the CSISS scale was also high (Figure 3a and b). Interrater agreement between specific values was slight even among experienced surgeons and never exceeded fair. Additionally, we observed a rather low ICC near the poor correlation level among all raters. Agreement of the raters regarding management ranged from medium to substantial, depending on the surgeon's experience. For less experienced neurosurgeons, the level of agreement concerning surgical management was the same as for the SLIC scale and did not exceed a moderate level. However, this scale had insufficient sensitivity that only slightly exceeded 50%. In 29 operated patients, most surgeons gave a rating of 6 points or less, which suggested conservative treatment. Most raters in our study suggested that this happens because of imprecise interpretation of ligament injuries. In cases of compressive fractures or translational injuries, variability of answers was low. Additionally, variability was much higher if any evidence of ligament injury was revealed (reduction of the facet joint contact area, widening of the interspinous distance, and irregular widening of the interbody space). In most patients with a low CSISS score, MRI was performed so additional damage to the ligamentous complex and spinal cord was visualized, and this was the reason for the surgical intervention. This aspect indicated a possible need to expand the CSISS scale to include an MRI-based interpretation of damage to the discoligamentous complex. Furthermore, in 11 patients with an inappropriate interpretation of damage according to the CSISS scale, CT was performed with a slice thickness of 2 mm, which resulted in 2- and 3-dimensional low-quality reformations during the assessment. This fact indicates a high sensitivity of this scale to the quality of CT images and a high probability of erroneous interpretation of damage in cases with CT images of poor quality.

However, we found that the use of MRI did not significantly affect the agreement coefficients for the CSISS. In most groups, the agreement coefficient did not change or there was a slight downward trend. One exception concerned assessment of the CSISS by the experienced surgeons where, during both stages, kappa growth values were noted as 0.1 and 0.03, which also indicated a potential use of MRI along with the CSISS scale.

Despite a relatively low level of interobserver agreement, the reproducibility of both scales was excellent among all raters regardless of their experience.

Thus, on the basis of our study's results, we think that the SLIC scale is the optimal numeric scale. A high level of agreement observed among the experienced surgeons indicated that greater surgical experience was likely to result in an increase in the SLIC agreement level. Our result also corresponds with the data of the Chhabra et al⁷ study. Unfortunately, the application of any scale in clinical practice is very limited now in most clinics. The use of classifications greatly simplifies communication between specialists, and we believe that their use at the level of the emergency department is mandatory for all patients with spinal injury. The SLIC scheme is used much more often than the CSISS because of its simplicity and convenience.⁷

However, we presume that the CSISS has high clinical potential. We expect that MRI integration and unification of different assessments of bone and ligament injuries should increase the interrater kappa value in the CSISS scale and increase sensitivity for further surgical management. Perhaps assessment of the injury type using the CSISS scale performed by radiologists during primary visualization might reduce the time of decision making in debatable cases for surgeons.

Study Limitations

The limitations of this study need to be acknowledged. First, we described only the features of calculating the sensitivity of scales in relation to surgical tactics. All the patients in our study had been treated surgically. Our study would have benefited from consideration of a group of patients who have been conservatively treated for stable fractures, along with obtaining more precise responses from each rater without requiring median values. This strategy would make it possible to not only obtain a more accurate sensitivity value but also calculate the specificity of these scales in relation to further surgical tactics. Nevertheless, even in such indicative conditions, the advantage of the SLIC scale was clearly demonstrated.

Second, the study participants had to independently examine the CT and MRI images in the DICOM archives and build the reformations and 3-dimensional reconstruction. This may have affected the results of our study, especially the results obtained from the less experienced neurosurgeons because of considerably varying visualizations of the injury. Nevertheless, we consider that our results reflected challenges facing surgeons in real-world working situations.

The study participants did not separately mark scores for injury morphology, the disc-ligament complex, or the neurological status when using the SLIC scale. A more fine-tuned form of assessment could assist in evaluating the need for MRI in the SLIC scale in greater detail.

Another limitation is represented by the difficulties of comparing our data with ICC indicators in other studies. Among present studies that assessed the ICC,^{3,4,8-10} the model and/or form of the ICC are provided in only 2 articles. In one study, authors reported the application of a 2-way random model of the ICC⁹; however, no data concerning form or the agreement type were provided. In another study, the authors used averaging during the interobserver agreement assessment,³ but the model and agreement type were also not provided. Therefore, no adequate comparison can be performed between the mentioned studies as well as between these studies and our study.

Conclusions

Our study demonstrated better management reliability, sensitivity, and reproducibility of the SLIC scale. This scale provided moderate interrater agreement with moderate to excellent ICC. The CSISS scale demonstrated high reproducibility; however, the large variability in the responses using a visual assessment of ligament injuries prevented raters from reaching a

moderate level of agreement. Nevertheless, MRI integration may increase the sensitivity of the CSISS scale for fracture management. In future studies, a model of the applied ICC parameters should be provided to perform an adequate comparison. More studies concerning the clinical application of the SLIC and CSISS scales should be conducted to define the most convenient scale for use in clinical practice.

Acknowledgments

We would like to thank Editage (www.editage.com) for English language editing.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Ivan Lvov, MD, PhD  <https://orcid.org/0000-0003-1718-0792>

Ethical Approval

Ethical approval for conducting the study was provided by the appropriate institutional review board.

Informed Consent

As this was a retrospective study of anonymized patient records, the requirement of consent was waived by the appropriate ethics review board.

References

- White AA, Panjabi MM. Physical properties and functional biomechanics of the spine. In: White AA, Panjabi MM, eds. *Clinical Biomechanics of the Spine*. 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 1990.
- Walters BC, Hadley MN, Hurlbert RJ. Guidelines for the management of acute cervical spine and spinal cord injuries: 2013 update. *Neurosurgery*. 2013;60(suppl 1):82-91.
- Anderson PA, Moore TA, Davis KW, et al; Spinal Trauma Study Group. Cervical spine injury severity score. Assessment of reliability. *J Bone Joint Surg Am*. 2007;89:1057-1065.
- Vaccaro AR, Hurlbert RJ, Patel AA, et al; Spinal Trauma Study Group. The subaxial cervical spine injury classification system: a novel approach to recognize the importance of morphology, neurology, and integrity of the disco-ligamentous complex. *Spine (Phila Pa 1976)*. 2007;32:2365-2374.
- Dvorak MF, Fisher CG, Fehlings MG, et al. The surgical approach to subaxial cervical spine injuries: an evidence-based algorithm based on the SLIC classification system. *Spine (Phila Pa 1976)*. 2007;32:2620-2629.
- Patel AA, Hurlbert RJ, Bono CM, Bessey JT, Yang N, Vaccaro AR. Classification and surgical decision making in acute subaxial cervical spine trauma. *Spine (Phila Pa 1976)*. 2010;35(21 suppl):S228-S234.
- Chhabra HS, Kaul R, Kanagaraju V. Do we have an ideal classification system for thoracolumbar and subaxial cervical spine injuries: what is the expert's perspective? *Spinal Cord*. 2015;53:42-48.
- van Middendorp JJ, Audigé L, Bartels RH, et al. The subaxial cervical spine injury classification system: an external agreement validation study. *Spine J*. 2013;13:1055-1063.
- Stone AT, Bransford RJ, Lee MJ, et al. Reliability of classification systems for subaxial cervical injuries. *Evid Based Spine Care J*. 2010;1:19-26.
- Zehnder SW, Lenarz CJ, Place HM. Teachability and reliability of a new classification system for lower cervical spinal injuries. *Spine (Phila Pa 1976)*. 2009;34:2039-2043.
- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1:30-46.
- Trethethan R. Intraclass correlation coefficients: clearing the air, extending some cautions, and making some requests. *Health Serv Outcomes Res Methodol*. 2017;17:127-143.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
- Vaccaro AR, Koerner JD, Radcliff KE, et al. AOSpine subaxial cervical spine injury classification system. *Eur Spine J*. 2016;25:2173-2184.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155-163.
- van Middendorp JJ, Audigé L, Hanson B, Chapman JR, Hosman AJ. What should an ideal spinal injury classification system consist of? A methodological review and conceptual proposal for future classifications. *Eur Spine J*. 2010;19:1238-1249.