Spine

SPINE Volume 36, Number 17, pp E1140–E1144

Reliability and Reproducibility of Subaxial Cervical Injury Description System

A Standardized Nomenclature Schema

Christopher M. Bono, MD,* Andrew Schoenfeld, MD,† Giri Gupta, MD,‡ James S. Harrop,§ Paul Anderson, MD,¶ Alpesh A. Patel, MD,‖ John Dimar, MD,** Bizhan Aarabi, MD,†† Andrew Dailey, MD,‡‡ Alexander R. Vaccaro, MD,§§ Ralf Gahr, MD,¶¶ Christopher Shaffrey, MD,*** David G. Anderson, MD,§§ and Raj Rampersaud, MD,†††

Study Design. Radiographic measurement study.

Objective. To develop a standardized cervical injury nomenclature system to facilitate description, communication, and classification among health care providers. The reliability and reproducibility of this system was then examined.

Summary of Background Data. Description of subaxial cervical injuries is critical for treatment decision making and comparing scientific reports of outcomes. Despite a number of available classification systems, surgeons, and researchers continue to use descriptive nomenclature, such as "burst" and "teardrop" fractures, to describe injuries. However, there is considerable inconsistency with use of such terms in the literature.

Methods. Eleven distinct injury types and associated definitions were established for the subaxial cervical spine and subsequently refined by members of the Spine Trauma Study Group. A series of 18 cases of patients with a broad spectrum of subaxial cervical spine injuries was prepared and distributed to surgeon raters. Each rater was provided with the full nomenclature document and asked to select primary and secondary injury types for each case. After receipt of the raters' first round of classifications, the cases were resorted and returned to the raters for a second round of review. Interrater and intrarater reliabilities were calculated as percent agreement and Cohen kappa (κ) values. Intrarater reliability was assessed by comparing a given rater's diagnosis from the first and second rounds.

Results. Nineteen surgeons completed the first and second rounds of the study. Overall, the system demonstrated 56.4% interrater agreement and 72.8% intrarater agreement. Overall, interrater κ values demonstrated moderate agreement while intrarater κ values showed substantial agreement. Analyzed by injury types, only four (burst fractures, lateral mass fractures, flexion teardrop fractures, and anterior distraction injuries) demonstrated greater than 50% interrater agreement.

Conclusion. This study demonstrated that, even in ideal circumstances, there is only moderate agreement among raters regarding cervical injury nomenclature. It is hoped that more familiarity with the proposed system will increase reproducibility in the future. Additional research is required to establish the clinical utility of this novel nomenclature schema.

Key words: classification, computed tomography, dislocation, fracture, injury types, nomenclature, radiographic, subaxial cervical spine, trauma. **Spine 2011;36:E1140–E1144**

Treatment decision making for subaxial cervical spine injuries can be challenging. Among other factors, such as the presence of a neurological deficit or high-energy polytrauma, challenges have arisen from lack of a consistent, reproducible classification system that is capable of guiding management and predicting outcomes.[1,2]

Intrinsic in most classification systems[1–9], and arguably more commonly utilized, are nomenclature schemata by which injuries are verbally described. For example, in the Allen and Ferguson system[5], five stages of cervical flexion-compression injuries are described. While most spine practitioners may not be intimately aware of the distinguishing features of each stage, they are sufficiently familiar with the general term "flexion-compression," as it has become part of the colloquial vocabulary of spinal trauma. In another example,

From the **Orthopaedic Spine Service, Harvard Medical School, Brigham and Women's Hospital, Boston, MA; †Department of Orthopaedic Surgery, William Beaumont Army Hospital, El Paso, TX; ‡Department of Orthopaedic Surgery, Booneville Medical Clinic, Jackson, KY; §Department of Neurosurgery, Thomas Jefferson University, Philadelphia, PA; ¶Department of Orthopaedic Surgery, University of Wisconsin, Madison, WI; ||Department of Orthopaedic Surgery, University of Utah, Salt Lake City, UT; **Department of Orthopaedic Surgery, University of Kentucky, Louisville, KT; ††Department of Neurosurgery, University of Maryland, Baltimore, MD; ‡‡Department of Neurosurgery, University of Utah, Salt Lake City, UT; §§Department of Orthopaedic Surgery, Thomas Jefferson University, Philadelphia, PA; ¶¶Department of Orthopaedic Surgery, Trauma Centre, Leipzig, Germany; ***Department of Neurosurgery, University of Virginia, Charlottesville, VA; and †††Department of Orthopaedic Surgery, University of Toronto, Ontario, Canada.

Acknowledgement date: February 11, 2009. Revision date: December 13, 2009. Acceptance date: January 15, 2010.

The manuscript submitted does not contain information about medical device(s)/drug(s). Corporate/Industry funds were received in support of this work.

One or more of the author(s) has/have received or will receive benefits for personal or professional use from a commercial party related directly or indirectly to the subject of this manuscript: e.g., honoraria, gifts, consultancies, royalties, stocks, stock options, decision making position.

Address correspondence and reprint requests to Christopher M. Bono, MD, Harvard Medical School, Chief, Orthopaedic Spine Service, Brigham and Women's Hospital, 75 Francis St, Boston, MA 02115; E-mail: bonocm@prodigy.net

DOI: 10.1097/BRS.0b013e318221a56d

TABLE 1. Interpretation of κ Statistic

Kappa Value	Agreement
>0.80	Near perfect
0.61–0.80	Substantial
0.41–0.60	Moderate
0.21–0.40	Fair
0.00–0.20	Slight

the system proposed by Daffner *et al*[10] denotes five variants of axial compression injuries, one being a burst fracture. Even newly proposed classification systems, such as the Subaxial Injury Classification system of the Spine Trauma Study Group (STSG),[1,11] rely on conventional names, such as "burst fractures," to categorize injuries and assign scores.

What is extrinsic to these classification systems is an exact and precise definition or description of the meaning of each nomenclature term. Despite the widespread and almost universal use of the terms burst fracture, teardrop fracture, and facet jump, the authors could find no original or primary source definition for any of these terms. This deficiency in the literature can, and likely already has, led to confusion in interpreting published outcomes of subaxial cervical injury treatment. In following, there is no certainty that any two papers reporting treatment outcomes for an injury type, (*e.g.*, cervical teardrop fracture) are referring to the same injury. Notwithstanding, published studies, textbooks, and treatment guidelines continue to use a system of injury nomenclature that has no single or identifiable origin.

Prompted by multiple cases of disagreement during review of prospectively collected clinical data from a multicenter database, the members of the STSG developed by consensus a uniform nomenclatural list of common injury types called SCIDS (Subaxial Cervical Injury Description System). The purpose of this study was to evaluate the reliability and reproducibility of the respective definitions in the system. In doing so, it was the authors' goal to improve the accuracy and precision by which subaxial cervical injuries are categorized.

MATERIALS AND METHODS

Phase I: Development of the List of Cervical Injury Types and Their Definitions

The development of SCIDS was effected by a modified Delphi technique. The initial phase of this study was to develop a list of the injury names (or types) and their corresponding written definitions. To serve as a template, two of the authors (C.M.B. and J.S.H.) formulated a preliminary document with a list of definitions. This was subsequently sent to each STSG member *via* e-mail. Members then changed the document using the track changes option in Microsoft Word (Microsoft Corporation, Redmond, WA). These suggested changes were reviewed in detail. The document language was then amended to accommodate all changes, as best as possible, without being contradictory or duplicative. This included addition of injury types as well alterations to the definition language.

The revised document was once again sent to all of the members for reevaluation. This process was repeated until each member expressed a uniform consensus and agreement with the established definition. In the end, 11 distinct injury types and associated definitions were established in the nomenclature schema. Of note, the group did not attempt to define stability or instability.

Phase II: Assessing the Reliability and Reproducibility of the Nomenclature Schema

After institutional review board approval was obtained, blinded, de-identified radiographic images, including plain radiographs and computerized tomography, of 18 patients with subaxial cervical spine injuries from multiple institutions were distributed *via* mailed compact discs to 21 STSG members. The selected cases represented a broad variety of injury types. Source computed tomographic images were available in reformatted two-dimensional axial, sagittal, and coronal views in their original DICOM (NEMA, Rosslyn, VA) format. The viewing software was eFilm Lite (eFilm Medical Inc., Toronto, ON), which allowed the surgeons to scroll through image slices and adjust contrast as they would in usual practice.

Each surgeon rater was provided with the full nomenclature document and precise instructions on how to complete the review. Surgeons were asked to select the primary and secondary injury type for each case, with the primary injury defined as that, which would mostly direct the treatment plan. No information regarding clinical findings, additional injuries, or subsequent treatment was provided. The order of review was randomized among observers. After receipt of the raters' selections, the compact discs were mailed for a second round of review.

Statistical Analysis

Inter- and intrarater reliabilities were calculated as both (1) percent agreement and (2) Cohen kappa (κ) statistic. Intrarater reliability was assessed by comparing each rater's selections from the first and second rounds. κ statistic values were graded according to the recommendations of Landis and Koch[12] (Table 1). All statistical calculations were performed using SPSS software (SPSS Inc., Chicago, IL).

RESULTS

Phase I

Table 2 lists the injury types and respective definitions upon which STSG members agreed to include in SCIDS.

Phase II

Twenty-one surgeons agreed to participate in the reliability evaluation process, with nineteen completing the first and second rounds of the study protocol. Participants included fourteen orthopedic and five neurological surgeons. Two of the participants were orthopedic spine surgical fellows. All participants were affiliated with a teaching institution.

Overall, the nomenclature schema demonstrated 56.4% interrater agreement while the intrarater agreement was 72.8%. The interrater κ statistic was 0.438 (95% CI, 0.418–0.459), indicating moderate agreement. The intrarater κ statistic was 0.648 (95% CI = 0.584–0.713) suggesting substantial agreement.

As there were no "correct" selections for injury types, it was not feasible to calculate agreement for an individual case. However, agreement for a different injury name was performed. Lateral mass fractures and burst fractures demonstrated the greatest interrater agreement (70.8% and 68.1%, respectively), while facet subluxations, spinous process fractures, and transverse process fractures had the lowest (Table 3). In total, only four injury types (lateral mass fractures, burst fractures, flexion teardrop fractures, and anterior distraction injuries) were found to have more than 50% interrater agreement.

DISCUSSION

Terms, such as burst fracture, teardrop fracture, and lateral mass fracture, are pervasive in the published spine literature and integral in guiding patient care.[1–11,13] Repetitive historical reference to these terms has led to entrenchment of nomenclature without standardized definitions of commonly encountered subaxial cervical injury types.[1,2,8] Therefore, it has been assumed that what one surgeon describes, for example, as a burst fracture would be similarly named by another surgeon. However, the current data indicate that this assumption may be incorrect.

This study attempted to maximize the reproducibility and reliability of a subaxial cervical spine nomenclature schema. All of the participating surgeons were directly involved with the formulation of the list of injuries and definitions. In addition, the surgeons were provided a written copy of the nomenclature schema to which they could refer. Despite these conditions, percent agreement was, at best, 56.4% between raters and 72.8% between the first and second rounds (intrarater).

These values are similar to those found for a multitude of other classification systems in the spine and nonspine literature.[14–17] In assessing the thoracolumbar injury severity score and thoracolumbar injury classification and severity score systems, Whang *et al*[14] recorded interrater κ statistics between 0.45 and 0.74, respectively. Wood *et al*[15] compared the Arbeitsgemeinschaft für Osteosynthesefragen and Denis systems for thoracolumbar fracture classification, finding κ

TABLE 2. The 11 Distinct Injury Types and Associated Definitions of SCIDS

Injury Type (Name)	Definitions (Description)
Spinous process fracture	Fracture that detaches a portion of the spinous process, or the entire spinous process, from the cervical lamina.
Isolated lamina fracture	Fracture that extends through the lamina medial to the facet joint/lateral mass and lateral to the base of the spinous process. No subluxation or kyphosis should be present in this injury pattern.
Unilateral facet dislocation	Disruption of a single facet joint in which the inferior articular process of the cranial vertebra has translated anterosuperiorly over the superior articular process of the caudal vertebra.
Bilateral facet dislocation	Disruption of both facet joints in which the inferior articular processes of the cranial vertebra have translated anterosuperiorly over the superior articular processes of the caudal vertebra. This pattern of injury may be associated with comminution, or fracture, of the facet joint complex. Perched facets, in which the tip of the inferior articular process abuts the superior articular process, qualify as dislocations as long as there is no articular surface apposition.
Facet subluxation	Misalignment of two adjacent vertebrae resulting in less than full apposition of facet articular surfaces (unilateral or bilateral).
Flexion teardrop fracture	Vertebral body fracture characterized by a triangular, or quadrangular, bone fragment derived from the anteroinferior vertebral body. Coexistent anterior cranial-caudal vertebral body height loss must also be present.
Lateral mass fracture	Fracture of any portion of the lateral mass complex, including the articular processes and the pedicle. This categorization includes the so-called "floating lateral mass," where ipsilateral fractures of the lamina and pedicle result in superior and inferior articular processes that are in discontinuity with the native vertebrae.
Compression fracture	Vertebral body fracture with loss of craniocaudal height. No involvement of the posterior cortical margin and no translational or rotational deformity are allowed.
Burst fracture	Vertebral body fracture with loss of craniocaudal height and involvement of the posterior cortical margin. This fracture pattern is often associated with retropulsion of bone fragments into the spinal canal.
Anterior distraction injury	Bone, ligament, or disc injury that results in craniocaudal distraction of the anterior disc space to a greater extent than the posterior disc space. By definition, this injury pattern represents disruption of the anterior tension band.
Transverse process fracture	Fracture along any portion of the transverse process, including fractures that extend into the foramen transversarium.

TABLE 3. Percent Agreement by Injury Type

Injury Type	% Agreement
Lateral mass fractures	70.8
Burst fracture	68.1
Flexion teardrop fractures	57.3
Anterior distraction injury	55.7
Isolated lamina fracture	45.3
Facet dislocation (Bilateral)	33.6
Facet dislocation (Unilateral)	28.6
Compression fractures	28.3
Facet subluxation	20.0
Spinous process fracture	14.3
Transverse process fracture	4.0

statistics of 0.475 and 0.598, respectively. Similar values have been found for nonspine classification systems. In a study of the Sanders classification for calcaneus fractures, Humphrey *et al*[16] reported an average κ value of 0.41 with values ranging from 0.07 to 0.64. Finally, Malek *et al*[17] reported a κ statistic of 0.61 for the commonly used Weber classification of ankle fractures. Though not ideal, the current values of 0.438 and 0.648 for inter- and intrarater κ statistics are comparable to these previously reported assessments of fracture classification systems.

There are a number of findings in this study that deserve further explanation. A number of injuries, specifically facet subluxations, spinous process fractures, and transverse process fractures, had fairly low agreement. This might not be unexpected for facet subluxation, as some surgeons, despite the clear definitions provided, might have considered these injuries to be dislocations or assigned them a different injury type altogether. What was surprising was the low agreement for spinous process (14.3%) and transverse process (4.0%) fractures. It seems unlikely that, given a single image of a vertebra with a fractured spinous process or transverse process, participants would disagree with the presence, absence, or the type of injury. However, surgeons were asked to review many images of oftentimes complex injuries. In this situation, many of the participants may have overlooked a relatively minor secondary fracture, as they may not consider them to be important in the overall management of the injury. It is also possible that fractures at the base of the spinous processes were "incorrectly" categorized as lamina fractures. This distinction probably has little clinical significance, however.

To better understand these inconsistencies, an informal analysis of a random sampling of seven of the respondents' individual answers. This revealed a number of interesting findings. Ignoring the delineation of primary or secondary diagnosis, seven of seven surgeons agreed that a burst fracture was present in case 12 and a lateral mass fracture was present in case 18, indicating perfect agreement. In case 2, 4, 6, 9, 10, and 11, six of seven surgeons agreed upon at least one injury being present. There were particularly troubling cases as well. In case 7, for instance, four of seven agreed that an anterior distraction injury was present, while three of seven believed that only a lamina fracture was present. Considering the differences between these two injury types, it is difficult to resolve this disparity. Other disparities are more easily resolved, however. In case 5, five of seven noted a flexion-teardrop fracture while two of seven noted a burst fracture. With an unclear distinction between teardrop and burst fractures in previously published literature, this inconsistency is not unexpected, despite the clear distinction in the SCIDS definitions between these two injuries.

There are several limitations to this study. First, because of the complexity of the statistics, a comparison of primary and secondary nomenclature selections was not performed. It is possible, therefore, that the calculated reliabilities were underestimated. In subsequent studies, participants could be asked to list all of the injury types present, obviating observers' weighing one injury's importance over another. For example, if all of the participants listed "spinous process fracture" as one of the injury types for a particular case, this could be interpreted as 100% agreement.

This study might have overestimated the reliability of the system. The raters were involved in the development of the SCIDS nomenclature system, a process that required extensive time, effort, and debate. The reproducibility of the nomenclature schema among the spine community at-large, neurosurgeons, orthopedic traumatologists, emergency department providers, or general surgical traumatologists, and any other group of physicians involved in care of spine trauma is unknown. It may be expected that the κ statistic and percent agreements would be different in groups unfamiliar with this system or its development.

The present data presents the initial assessment of a newly developed schema. With continued use over a longer period of time, the reliability may improve. Previously, investigations of spine injury classification system have demonstrated that as physician familiarity increases so does the percent agreement and κstatistic.[1,18] Although raters involved were intimately involved with the development of the nomenclature schema, the current data represent the first two trials of its usage. It can be postulated that additional use of the nomenclature would produce higher inter- and intrarater reliability values. Further investigation is needed to assess the schema's performance among physicians other than those involved in its development.

To plan effective treatment, a common nomenclature system, such as SCIDS, is important. However, it cannot stand alone. The stability of an injury and the presence of a neurological deficit are critical factors. It is hoped that recently proposed systems that address these issues, such as the Cervical Spine Injury Severity Score and Subaxial Injury Classification, can be used in concert with SCIDS to comprehensively characterize injuries to allow the best possible treatment decisions to be made.

CONCLUSION

A common language for describing subaxial cervical spine fractures and dislocations is an integral part of effective treatment decision making, determining prognosis, and clinical research. Despite its widespread use and assumption of uniformity, the current data indicate that, even in ideal circumstances, there is less than ideal agreement regarding injury nomenclature. It is hoped that more routine use and familiarity with the proposed system will increase reliability in the future. However, further research is needed to more broadly demonstrate its generalizability and clinical utility.

➢ Key Points

- ❑ There was moderate overall agreement among raters regarding cervical injury nomenclature.
- ❑ Four of eleven injury types were found to have greater than 50% interrater agreement.
- ❑ Lateral mass fractures and burst fractures were the most reliably named, while facet subluxations, spinous process fractures, and transverse process fractures were the least.
- ❑ Further study is needed to fully demonstrate the generalizability and clinical utility of the proposed nomenclature schema.

References

1. Vaccaro AR, Hulbert RJ, Patel AA, et al. 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* 2007;32:2365–74.
2. Anderson PA, Moore TA, Davis KW, et al. Cervical spine injury severity score. Assessment of reliability. *J Bone Joint Surg Am* 2007;89:1057–65.
3. Holdsworth F. Fractures, dislocations, and fracture-dislocations of the spine. *J Bone Joint Surg Am* 1970;52:1534–51.
4. Bohlman HH. Acute fractures and dislocations of the cervical spine. An analysis of three hundred hospitalized patients and a review of the literature. *J Bone Joint Surg Am* 1979;61:1119–42.
5. Allen BL Jr, Ferguson RL, Lehmann TR, et al. A mechanistic classification of closed, indirect fractures and dislocations of the lower cervical spine. *Spine* 1982;7:1–27.
6. Louis R. Spinal stability as defined by the three-column spine concept. *Anat Clin* 1985;7:33–42.
7. Harris JH, Edeiken-Monroe B, Kopansiky DR. A practical classification of acute cervical spine injuries. *Orthop Clin North Am* 1986;17:15–30.
8. Moore TA, Vaccaro AR, Anderson PA. Classification of lower cervical spine injuries. *Spine* 2006;31:S37–43.
9. White AA III, Panjabi M. Kinematics of the spine. In:White AA III, Panjabi MM, eds. *Clinical biomechanics of the spine*. 2nd ed. Philadelphia, PA: Lippincott, Williams & Wilkins; 1990:85–278.
10. Daffner RH, Brown RR, Goldberg AL. A new classification for cervical vertebral injuries: influence of CT. *Skeletal Radiol* 2000;29:125–32.
11. 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* 2007;32:2620–9.
12. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977;33:363–74.
13. Guidelines for management of acute cervical injuries. Introduction. *Neurosurgery* 2002;50(suppl 3):S1.
14. Whang PG, Vaccaro AR, Poelstra KA, et al. The influence of fracture mechanism and morphology on the reliability and validity of two novel thoracolumbar injury classification systems. *Spine*. 2007;32(7):791–5.
15. Wood KB, Khanna G, Vaccaro AR, et al. Assessment of two thoracolumbar fracture classification systems as used by multiple surgeons. *J Bone Joint Surg Am* 2005;87(7):1423–9.
16. Humphrey CA, Dirschl DR, Ellis TJ. Interobserver reliability of a CT-based fracture classification system. *J Orthop Trauma* 2005;19:616–22.
17. Malek IA, Machani B, Mevcha AM, et al. Inter-observer reliability and intra-observer reproducibility of the Weber classification of ankle fractures. *J Bone Joint Surg Br* 2006;88:1204–6.
18. Patel AA, Vaccaro AR, Albert TJ, et al. The adoption of a new classification system: time-dependent variation in interobserver reliability of the thoracolumbar injury severity score classification system. *Spine* 2007;32:E105–10.