Reliability analysis of the epidural spinal cord compression scale

Clinical article

MARK H. BILSKY, M.D.,1,2 ILYA LAUFER, M.D.,2 DARYL R. FOURNEY, M.D., F.R.C.S.C.,3 MICHAEL GROFF, M.D.,4 MEIC H. SCHMIDT, M.D.,5 PETER PAUL VARGA, M.D.,6 FRANK D. VRIonis, M.D., M.P.H., Ph.D.,7 YOSHIYA YAMADA, M.D.,8 PETER C. GERSZTEN, M.D.,9 AND TIMOTHY R. KUKLO, M.D., J.D.10

Departments of 1Neurosurgery and 4Radiation Oncology, Memorial Sloan–Kettering Cancer Center; 2Department of Neurosurgical Surgery, Weill Cornell Medical College, New York-Presbyterian Hospital, New York, New York; 4Department of Neurosurgery, Beth Israel Deaconess Medical Center, Boston, Massachusetts; 5Department of Neurosurgery, University of Utah, Salt Lake City, Utah; 6Department of Neurosurgery, H. Lee Moffitt Cancer Center, Tampa, Florida; 7Department of Neurosurgery, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania; 8Department of Orthopaedics, Washington University, St. Louis, Missouri; 9Department of Neurosurgery, University of Saskatchewan, Saskatoon, Saskatchewan, Canada; and 10Department of Orthopaedics, National Center of Spinal Disorders, Budapest, Hungary

Objective. The evolution of imaging techniques, along with highly effective radiation options has changed the way metastatic epidural tumors are treated. While high-grade epidural spinal cord compression (ESCC) frequently serves as an indication for surgical decompression, no consensus exists in the literature about the precise definition of this term. The advancement of the treatment paradigms in patients with metastatic tumors for the spine requires a clear grading scheme of ESCC. The degree of ESCC often serves as a major determinant in the decision to operate or irradiate. The purpose of this study was to determine the reliability and validity of a 6-point, MR imaging–based grading system for ESCC.

Methods. To determine the reliability of the grading scale, a survey was distributed to 7 spine surgeons who participate in the Spine Oncology Study Group. The MR images of 25 cervical or thoracic spinal tumors were distributed consisting of 1 sagittal image and 3 axial images at the identical level including T1-weighted, T2-weighted, and Gd-enhanced T1-weighted images. The survey was administered 3 times at 2-week intervals. The inter- and intrarater reliability was assessed.

Results. The inter- and intrarater reliability ranged from good to excellent when surgeons were asked to rate the degree of spinal cord compression using T2-weighted axial images. The T2-weighted images were superior indicators of ESCC compared with T1-weighted images with and without Gd.

Conclusions. The ESCC scale provides a valid and reliable instrument that may be used to describe the degree of ESCC based on T2-weighted MR images. This scale accounts for recent advances in the treatment of spinal metastases and may be used to provide an ESCC classification scheme for multicenter clinical trial and outcome studies. (DOI: 10.3171/2010.3.SPINE09459)

Key Words • epidural spinal cord compression • spinal metastases • thoracic spine • stereotactic radiosurgery • lumbar spine • metastatic spine tumor

This article contains some figures that are displayed in color online but in black and white in the print edition.
Reliability analysis of the ESCC scale

Historically, myelography was used to define ESCC based on the degree of myelogram dye block as none, partial or incomplete, or complete.\(^8,10,18\) The transition to MR imaging as the diagnostic modality of choice for spine tumors has left a void in defining the degree of ESCC, thus causing difficulty in interpreting the more recently published literature.\(^3\)

The importance of a clearly defined grading system for ESCC has become even more important with the development of SRS of the spine.\(^5,6\) Despite a steep dose gradient, tumors that compress or even abut the spinal cord may be excluded from consideration of SRS because of the constraints of spinal cord tolerance. A validated ESCC grading system will provide consistent reporting in the radiation, surgery, and radiology literature.

This study presents the validation of a 6-point grading system to standardize reporting regarding ESCC. A 4-grade, MR imaging–based grading system was previously reported based on the degree of impingement of the CSF space and spinal cord compression.\(^4\) This system was sufficient to report surgical decision making, but the introduction of SRS necessitated expansion to a 6-point grading system to account for gradations of thecal sac impingement. This grading system can be used in single and multiinstitutional trials and may ultimately be used to help guide decision making.

**Methods**

**Survey Administration**

A 6-point grading system was established to determine the degree of ESCC. The original 4-point system ranged from 0 to 3. In the original system, a grade of 0 indicated bone involvement only; 1, epidural impingement; 2, spinal cord compression but CSF visible; and 3, spinal cord compression but no CSF seen. Based on the need to further delineate epidural impingement for SRS purposes, the Grade 1 category was subdivided into Grades 1a–c where 1a indicated epidural impingement but no deformation of the thecal sac; 1b, deformation of the thecal sac, but without spinal cord abutment; and 1c, deformation of the thecal sac with spinal cord abutment, but without compression (Figs. 1 and 2).

A survey consisting of axial MR images was distributed by mail to 7 members of the Spine Oncology Study Group consisting of neurological and orthopedic surgeons who specialize in the treatment of spine tumors. A series of 25 metastatic tumors were compiled. Included for each case were 1 sagittal T1-weighted image and 3 axial images, including T1-weighted, T2-weighted, and T1-weighted Gd-enhanced images. An axial image from an adjacent normal spinal segment of the same signal intensity was included for comparison on each slide (Fig. 3). The surgeons were asked to grade the degree of epidural spine cord compression using the 6-point grading scheme. This survey was administered 3 times at 2-week intervals. Each time the order of the images was changed in a random fashion.

**Statistical Analysis**

Statistical analysis was conducted using SPSS 15.0 software (SPSS, Inc.). Interrater agreement and test-retest reliability were assessed using the ICC. The test-retest reliability assessed the degree of agreement between the first and second and the second and third administrations of the survey for each surgeon. The data were further stratified to determine if specific MR imaging sequences provided superior reliability in the measurement of ESCC. An ICC less than 0.39 was considered consistent with poor correlation, 0.4–0.59 with fair correlation, 0.6–0.74 with good correlation, and 0.75–1.0 with excellent correlation. All CIs were computed with 95% certainty.

**Results**

The overall interrater reliability of the ESCC grading scale ranged between fair and good (0.569–0.665) during the 3 iterations of the survey. While the ICC ranged between poor and good for the T1-weighted images with or without contrast administration (0.395–0.633), the ICC for T2-weighted images yielded superior results ranging from good to excellent (0.701–0.782). The intrarater reliability ranged between good and excellent (0.619–0.819), with T2-weighted images resulting in excellent correlation during both comparisons. Furthermore, the confidence intervals of the T1-weighted images without contrast and

---

**Fig. 1.** Schematic representation of the 6-point ESCC grading scale. A grade of 0 indicates bone-only disease; 1a, epidural impingement, without deformation of the thecal sac; 1b, deformation of the thecal sac, without spinal cord abutment; 1c, deformation of the thecal sac with spinal cord abutment, but without cord compression; 2, spinal cord compression, but with CSF visible around the cord; and 3, spinal cord compression, no CSF visible around the cord.
of the T2-weighted images did not overlap, indicating a statistically significant difference between the reliability of measurement provided by the 2 sequences.

Discussion

The need for a reliable, validated MR imaging-based ESCC grading system is important for consistency in reporting single and multiinstitutional outcomes of surgical and radiotherapeutic studies. Prior to MR imaging, decision making was based on myelography findings.8,10,18 Because of the invasive nature of myelography, these studies were often withheld until patients presented with advanced myelopathy. Magnetic resonance imaging provides a more tolerable diagnostic test than myelography, and, thus, patients often undergo imaging when there are symptoms of biological spine pain before the onset of neurological symptoms.1,7 More effective surgical and radiotherapeutic options offered prior to the development of myelopathy have helped improve outcome in patients with spine tumors.

Multiple surgical series have used high-grade spinal cord compression as indications for surgery, but most have not defined this term.9,13,14 Patchell et al.12 conducted a randomized, multiinstitutional study comparing the outcome of surgery and conventional external-beam radiotherapy with external-beam radiotherapy alone in patients with metastatic ESCC and symptomatic, solid tumors. In this study, ESCC was defined as displacement of the spinal cord. The surgical group experienced improved maintenance and recovery of ambulation, bladder function, and even survival.

In a number of studies, a 4-point ESCC grading system was used to further delineate degrees of spinal cord compression and justify the need for surgery, especially in patients who had symptomatic back pain or functional radiculopathy but in the absence of myelopathy.2,16 High-grade ESCC was defined as deformation of the spinal cord with partial (ESCC Grade 2) or complete (ESCC Grade 3) obliteration of the CSF space. Surgery was justified for neurological preservation in radiotherapy-resistant tumors with high-grade ESCC. Patients with ESCC Grade 0 (bone only) and Grade 1 (epidural impingement) were considered for radiation therapy even in the presence of radioresistant tumors.

The evolution of high-dose conformal photon therapy, such as SRS, necessitated a more sensitive ESCC grading system. Cytotoxic doses of radiation to the tumor are significantly greater than doses that result in toxicity to the spinal cord. Although spinal cord toxicity has not been precisely delineated for SRS, safe radiotherapy doses of 14 Gy to a single voxel on the spinal cord or 10 Gy to 10% of the spinal cord volume have been reported in the literature.17 In a recently reported dose escalation study, Yamada et al.17 showed that patients who received a 24-Gy single fraction achieved significantly better local tumor control than those who received less than 24 Gy. Additionally, those who received less than 15 Gy to the entire planning target volume had a significantly higher probability of recurrence.11 Both of these doses are above spinal cord tolerance.

In the original ESCC grading system, Grade 1 compression represented all degrees of spinal cord impingement, but the grade was not sensitive enough to differen-
Reliability analysis of the ESCC scale

Despite a steep dose gradient of approximately 10% per millimeter generated using current stereotactic radiosurgery techniques, the challenge remains whether one can deliver a cytotoxic dose to the tumor and remain within spinal cord constraints. On this basis, a tumor presenting as an ESCC grade of 1c with actual spinal cord abutment presents significantly greater challenges in radiation planning and delivery than an ESCC of Grade 1a or 1b that has a 1- or 2-mm distance from the spinal cord. In many cases, this small distance may change the safe therapeutic window. For instance, a renal cell carcinoma with an ESCC of Grade 1a or 1b may safely receive a minimum dose to the epidural margin of 15 Gy with a maximum dose of 14 Gy to a single voxel on the spinal cord. An ESCC grade of 1c with spinal cord abutment receiving a prescribed dose

**Fig. 3.** Examples of T1-weighted (A), T1-weighted Gd-enhanced (B), and T2-weighted (C) images used in the survey, with corresponding normal images presented for comparison.
of 15 Gy to the tumoral margin could overdose the spinal cord, possibly resulting in myelopathy. The subclassification of Grade 1 impingement improves the grading system as a tool to delineate potentially important differences in the ability to deliver effective SRS.

Conclusions

This paper represents a validation study of an ESCC grading system that takes into account the assessment of surgical and SRS candidates. Not unexpectedly, the T2-weighted images were the most reliable for both inter- and intrarater reliability. The T2-weighted images provide a myelogram-like effect in which the CSF spaces are readily identifiable. The T1-weighted images often show no differentiation between tumor and CSF, which are frequently both hypointense relative to normal vertebral body signal. Our study confirmed that T1-weighted contrast images were not as reliable as T2-weighted images. The proposed ESCC grading system, based on axial T2-weighted images, provides consistency in reporting and a basis for multisitutional trials.

A 6-point ESCC grading system represents a reliable and informative instrument that may guide treatment decisions. This scale takes into account recent advances in radiation oncology and the associated modifications in the surgical treatment of spinal tumors. Furthermore, it provides the basis for uniform ESCC reporting among various oncology centers and publications. The good inter- and intrarater reliability, especially when using T2-weighted images, validates this scale as a reliable grading device that may provide consistency when describing ESCC.

Disclosure

This study was supported by the Spine Oncology Study Group and funded by an educational/research grant from Medtronic Spinal and Biologics. Dr. Kuklo is a previous consultant for Medtronic Sofamor Danek.

Author contributions to the study and manuscript preparation include the following. Conception and design: Bilsky, Fourney, Groff, Schmidt, Varga, Vronis, Yamada, Gerszten, Kuklo. Acquisition of data: Bilsky, Fourney, Groff, Schmidt, Varga, Vronis, Yamada, Gerszten, Kuklo. Analysis and interpretation of data: all authors. Drafting the article: Bilsky, Laufer. Critically revising the article: Bilsky, Fourney, Groff, Schmidt, Varga, Vronis, Yamada, Gerszten, Kuklo. Reviewed final version of the manuscript and approved it for submission: all authors. Statistical analysis: Bilsky. Administrative/technical/material support: Bilsky. Study supervision: Bilsky.

References
