Spine

SPINE Volume 40, Number 24, pp E1299–E1304
© 2015 Wolters Kluwer Health, Inc. All rights reserved

Novel Method Using Baseline Normalization and Area Under the Curve to Evaluate Differences in Outcome Between Treatment Groups and Application to Patients With Cervical Spondylotic Myelopathy Undergoing Anterior Versus Posterior Surgery

Shian Liu, BS,[*] Lindsay Tetreault, BSc,[†] Michael G. Fehlings, MD, PhD,[†] Vincent Challier, MD,[*] Justin S. Smith, MD, PhD,[‡] Christopher I. Shaffrey, MD,[‡] Paul M. Arnold, MD,[§] Justin K. Scheer, BS,[¶] Jens R. Chapman, MD,[||] Branko Kopjar, MD, PhD,[#] Themistocles S. Protopsaltis, MD,[*] Virginie Lafage, PhD,[*] Frank Schwab, MD,[*] Eric M. Massicotte, MD, MSc,[**] Sangwook T. Yoon, MD, PhD,[††] and Christopher P. Ames, MD[‡‡]

Study Design. Retrospective review of a prospective database.

Objective. To describe a novel method that uses baseline normalization and area under the curve (AUC) to compare surgical outcomes between patients surgically treated anteriorly versus posteriorly for cervical spondylotic myelopathy (CSM).

Summary and Background Data. It is important to control for baseline characteristics, especially disease severity, when evaluating differences in outcomes between 2 treatment groups. However, current methods of reporting outcomes are limited perhaps diminish the health impact of the entire postoperative recovery experience.

Methods. In the prospective, multicenter AO Spine North America CSM database, 147 patients had complete modified Japanese Orthopaedic Association (mJOA) data at baseline and at 6-, 12-, and 24-months postoperatively and were either treated anteriorly (n = 94) or posteriorly (n = 53). Each patient's follow-up mJOA scores were normalized by dividing them by the patient's baseline value. A graph was then plotted with the time point on the *x*-axis and the normalized score or ''recovery index'' on the *y*-axis. The AUC was calculated and then compared between the anterior and posterior surgical approach groups.

Results. The non-normalized recovery profile of the anterior group was better than that of the posterior group, as the patients treated anteriorly had less functional impairment at baseline. After normalization, patients in the anterior and posterior group had similar recovery indices and AUCs at 6-months following surgery. At 24-months, patients treated posteriorly had a significantly higher recovery index (1.32) and a larger AUC (16.3) than those treated anteriorly (1.11, 14.5, $P = 0.004$ and $P = 0.006$, respectively).

Conclusion. This is the first study to apply AUC analysis to patients with CSM. In surgical patients with CSM, those treated anteriorly achieved a higher mJOA score at all time points than those treated posteriorly. The recovery indices, however, were not significantly different between approach groups at 6 months.

Key words: area under curve, cervical, cervical spondylotic myelopathy, HRQOL, recovery.

Level of Evidence: 3

Spine 2015;40:E1299–E1304

From the [*]Department of Orthopaedic Surgery, NYU Hospital for Joint Diseases, New York, NY; [†]Toronto Western Hospital, University Health Network, University of Toronto, Toronto, Ontario, Canada ; [‡]Department of Neurosurgery, University of Virginia Health System, Charlottesville, VA; [§]Department of Neurosurgery, University of Kansas, Kansas City, KS; [¶]Department of Neurological Surgery, Northwestern University Feinberg School of Medicine, Chicago, IL; [||]Department of Orthopaedic Surgery, University of Washington, Seattle, WA; [#]Department of Health Services, University of Washington, Seattle, WA; [**]Division of Neurosurgery, Toronto Western Hospital, Toronto, Canada; [††]Department of Orthopaedics, Emory University, Atlanta, GA; and [‡‡]Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA

Acknowledgment date: April 22, 2015. Acceptance date: June 29, 2015.

The manuscript submitted does not contain information about medical device(s)/drug(s).

AOSpine North America, Inc., a nonprofit 501(c)(3) corporation, the International Spine Study Group Foundation, and Depuy Synthes funds were received in support of this work.

Relevant financial activities outside the submitted work: board membership, consultancy, employment, grants, patents, royalties, expert testimony, payment for lectures, stocks, travel/accommodations/meeting expenses.

Address correspondence and reprint requests to Christopher P. Ames, MD, Department of Neurosurgery, University of California, San Francisco, Medical Center, 400 Parnassus Avenue, A850, San Francisco, CA 94143; E-mail: Christopher.Ames@ucsf.edu

DOI: 10.1097/BRS.0000000000001152

Copyright © 2015 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Clinicians and researchers are often interested in comparing outcomes between treatment groups. Some of these comparisons include early versus delayed treatment, drug versus placebo and conservative versus operative intervention. Even in randomized controlled trials, differences between groups may exist with respect to baseline characteristics that can bias statistical tests. It is therefore essential to ensure that baseline parameters, especially disease severity, are similar or adjusted for between groups before evaluating any differences in outcomes. This is especially important if such baseline characteristics may be significant predictors of outcome.

In the present study, we propose a novel method to adjust for baseline characteristics by first normalizing each patient's follow-up scores to his/her baseline score, plotting these points on a graph and then computing the area under the curve (AUC). In general, a receiver operating curve (ROC) gauges the accuracy of a predictive model or diagnostic test against a gold standard.[1,2] This concept originated from the field of "signal detection therapy" during World War II to assess how well radar receiver operators distinguished between an enemy target, a friendly ship and "noise." Since the 1970s, signal detection therapy has been applied to interpret medical results and is used throughout the fields of cardiology, surgery, and critical care.[3] The area under the ROC curve measures the accuracy of a diagnostic or prognostic test and its ability to discriminate between patients with or without a disease or a certain outcome.[1–3] Both ROC and AUC analyses have proven to be invaluable in diagnostic testing and in predictive modeling. Currently, there is ongoing development of AUC analysis and interest in expanding these concepts to compare outcomes.

This study aims to describe a novel method in patients with cervical spondylotic myelopathy (CSM) undergoing surgical intervention that uses baseline normalization and AUC to evaluate differences in outcome between treatment groups. In a study by Fehlings *et al.*,[4] the main objective was to determine whether the anterior or posterior surgical approach was more effective at improving clinical and neurological outcomes in patients with CSM. Upon analysis, key differences in demographics were identified between patients treated anteriorly and those treated posteriorly. The most significant of these was baseline severity score: based on the mJOA score, patients who underwent an anterior surgery had less severe myelopathy ($mJOA = 13.47 \pm 2.46$) than those approached posteriorly ($mJOA = 11.84 \pm 2.86$).[4] In the present study, we will use a novel method to first adjust for this difference in baseline severity score and then to evaluate differences in outcomes across treatment groups.

METHODS

Study Design and Inclusion Criteria

Patients with clinically and radiographically confirmed CSM were enrolled in this multicenter, prospective cohort study at 12 surgical spine centers across North America.[5,6] Each participating center was a member of the AOSpine North America clinical research network (SpineNet), a nonprofit consortium dedicated to research advancement in spinal pathologies. All centers obtained internal review board approval and had the resources required to conduct prospective clinical research.

Patients enrolled in this study met the following inclusion criteria: (1) presenting with symptomatic CSM with at least 1 clinical sign of myelopathy; (2) image-evidence of cervical cord compression on magnetic resonance imaging or computed tomography myelography; (3) 18 yr or older in age; (4) no prior surgical treatment for CSM; 5) no concomitant symptomatic lumbar stenosis. All patients underwent surgical intervention at their respective sites. The approach (posterior, anterior or circumferential), surgical technique and number of levels to decompress was at the discretion of the attending surgeon. This cohort was divided into 2 groups: patients who underwent only an anterior surgery and those who were treated only posteriorly. Patients treated with combined anterior and posterior approaches were not included in this study. All patients included in the current analysis had complete severity scores at baseline and at 6-, 12- and 24-months following surgery.

Data Collection and Outcome Assessment

Extensive data were collected at baseline, including age, gender, medical co-morbidities, symptomatology, and causative pathology. Functional status and quality of life were evaluated preoperatively and at 6-, 12-, and 24-months following surgery using the Nurick grade, mJOA, 30-m walking test, neck disability index (NDI), and Short-Form-36 (SF-36). This study will examine differences in functional outcomes between anterior and posterior surgical cohorts using the mJOA. The mJOA is a clinician-administered CSM-specific index that evaluates a patient's functional status by separately assessing upper and lower motor extremity function, sensation, and sphincter disturbance. It ranges from 0 (worst) to 18 (best) and demonstrates construct validity and responsiveness to change.[7]

Data were uploaded into an electronic data capture system and processed through the AOSpine North America Clinical Research Network data management center. Professional clinical research monitors performed periodic site visits to help ensure the accuracy and completeness of the collected data.

Statistical Analysis

The data were analyzed using SPSS version 17.0 software (Chicago, IL). Descriptive statistics were used to summarize baseline demographic information, surgical summary and surgical outcomes at 6-, 12-, and 24- months following surgery. MJOA scores were compared between approach groups at each time point using independent sample *t*-tests. Statistical significance was set at a *P*-value of <0.05.

A similar analysis was conducted after controlling for baseline severity score using the following methods:

Copyright © 2015 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

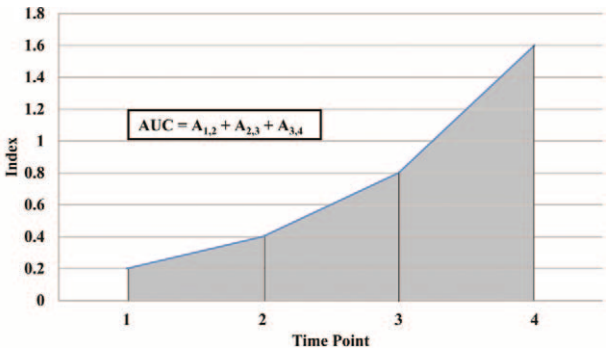

Figure 1. AUC calculation is the summation of the areas, calculated as trapezoids, between each time point. AUC indicates area under the curve.

1. Each mJOA score (baseline, 6-, 12-, and 24-months) was normalized by dividing it by the respective patient's baseline value. Therefore, at baseline, each patient started at 1. Any changes in functional status observed at follow-up visits were either >1, equal to 1, or <1 depending on whether the patient improved, stayed the same or deteriorated from baseline. This normalized value is referred to as the recovery index.
2. These scores were plotted on a line graph, where the *x*-axis was the time point and the *y*-axis was the recovery index.
3. The area under the line graph was divided into trapezoids at each time period (*e.g.* Baseline to 6 months; 6–12 months; 12–24 months) (Figure 1).
4. The area of each trapezoid was computed using the following equation:

$$A = \frac{\Delta x}{2}(y_{Ta} + y_{Tb})$$

where A is the area, Δx is the time between visit A and visit B, y_{Ta} is the normalized score for visit A and y_{Tb} is the normalized score for visit B.

The total AUC was calculated by summating the areas from each time period.

Steps 1 to 5 were done for each patient.

The means (±standard deviations) for the AUCs were computed for each approach group and statistically compared using independent *t*-tests. For the mJOA, the AUC represents postoperative recovery and a greater area reflects superior improvement.

RESULTS

Subjects

Two-hundred and seventy-eight patients were enrolled in the AOSpine North America CSM study at 12 surgical spine centers across North America. One hundred and thirty-one patients were excluded from our analysis because they either underwent a 2-stage anteroposterior surgery (n = 19) or did not attend their follow-up appointment at 6, 12, and/or 24 months (n = 123). Patients with a perfect preoperative score of 18 on the mJOA were also excluded as they did not have capacity for improvement (n = 7).

The study cohort consisted of 84 men and 63 women, with ages ranging from 29 to 86 years (mean age: 56.5 yr). Ninety-four were treated anteriorly whereas 53 underwent a posterior operation. In both the anterior and posterior groups, patients exhibited preoperative functional impairment as measured by the mJOA score (mean: 12.88, range: 4–17). However, on average, patients in the posterior group (mean: 11.8 ± 3.1) had a significantly lower baseline mJOA score than those in the anterior group (mean 13.5 ± 2.1) ($P < 0.001$). Both treatment groups demonstrated substantial improvements in functional status at 6-months following surgery and sustained these gains to the final follow-up visit at 24 months.

Unadjusted Comparison between Treatment Groups

Figure 2 illustrates the recovery curves of patients in the anterior (red) and the posterior (blue) treatment groups and demonstrates substantial improvement from preoperative status regardless of approach. Patients treated anteriorly were less impaired at baseline and started at a higher mJOA

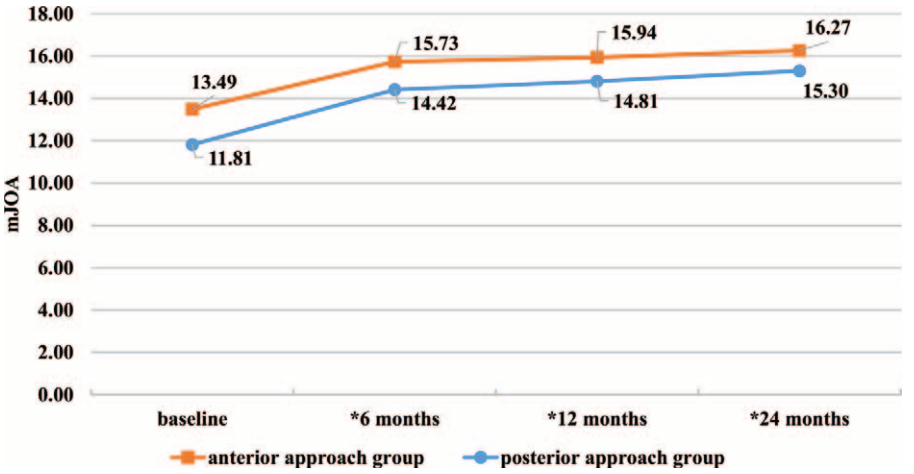

Figure 2. Baseline and postoperative course with mJOA by approach. *Simple stratification revealed the anterior approach group had significantly better mJOA scores at all time points ($P < 0.05$). mJOA indicates modified Japanese Orthopaedic Association.

Copyright © 2015 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

score than those treated posteriorly. At all time points, patients in the anterior group had a significantly better mJOA than those in the posterior group ($P < 0.05$). It is important, however, that each score is normalized to its baseline value before conclusions are made with respect to the relative efficacy of treatment approaches.

Patient X: An Example of Normalization

Patient X is a 63-year-old female who presented with severe myelopathy (mJOA = 10) secondary to spondylosis. This patient reported symptoms of numb and clumsy hands, impaired gait, bilateral arm paresthesia, and weakness. Her neurological signs included corticospinal distribution motor deficits, atrophy of intrinsic hand muscles, positive Hoffman sign, upgoing plantar responses, hyperreflexia, and a broad-based unstable gait. Following surgery, this patient improved to a mJOA of 12 at 6 months, 15 at 12 months, and 17 at 24 months.

These values are normalized by dividing each one by 10 as this was the patient's baseline score. The recovery index was computed as 1.20 at 6 months, 1.50 at 12 months and 1.70 at 24 months. The area under the curve was 6.60 between baseline and 6 months, 8.10 between 6 and 12 months, and 19.2 between 12 and 24 months. The following equation provides an example of how this area was computed between baseline and 6-months. $\frac{6\,months}{2}1.0 + 1.2 = 6.60$

The total area was calculated as 33.90 by summating the 3 individual areas.

Average of Normalized Scores: Posterior versus Anterior

The average of patients' normalized scores were computed and then graphed against time. Patients in the anterior group had a 1.06 recovery index at 6-months and a 1.11 recovery index at 24-months. These values were lower than those seen in the posterior group (1.15 at 6 months and 1.32 at 24 months). Figure 3 illustrates this difference as the recovery profile for the posterior group is above that of the anterior group. This difference, however, only reached statistical significance at the 24-month time point ($P = 0.004$).

Area Under the Curve

The total area under the recovery profile for the anterior group was 28.32 and was 31.00 for the posterior group (Table 1, Figure 3). The AUC between baseline to 6 months was not significantly different between the treatment groups (anterior: 6.52, posterior: 6.83, $P = 0.086$). The AUC between 6-months to 12-months (anterior: 7.17, posterior: 7.96, $P = 0.015$) and 12-months to 24-months (anterior: 14.53, posterior: 16.26, $P = 0.006$) was significantly greater for the posterior group than the anterior group. The AUC in total was significantly greater for the posterior group ($P = 0.013$).

DISCUSSION

In this study, we proposed a novel method to evaluate differences in outcome between treatment groups using baseline normalization and AUC. Using this method, we were able to confirm the results of the study by Fehlings *et al.*[4] and conclude that there is no significant difference in functional outcome at 6 months postoperatively in patients treated anteriorly and those treated posteriorly. At 12- and 24-months following surgery, however, patients treated posteriorly had a significantly greater recovery index and AUC than those treated anteriorly. This is likely because patients in the posterior surgery group were more severe preoperatively and had greater room for improvement. This is especially important to consider given that our outcome measure was the mJOA which may exhibit a ceiling effect at the upper end of the scale.

In our analysis, we were only able to adjust for baseline severity score. Although patients treated posteriorly are, on average, more severely impacted than those approached anteriorly, they also tend to be older, have multilevel degenerative disease and require decompressive surgery at a greater number of levels.[4] All of these factors may be

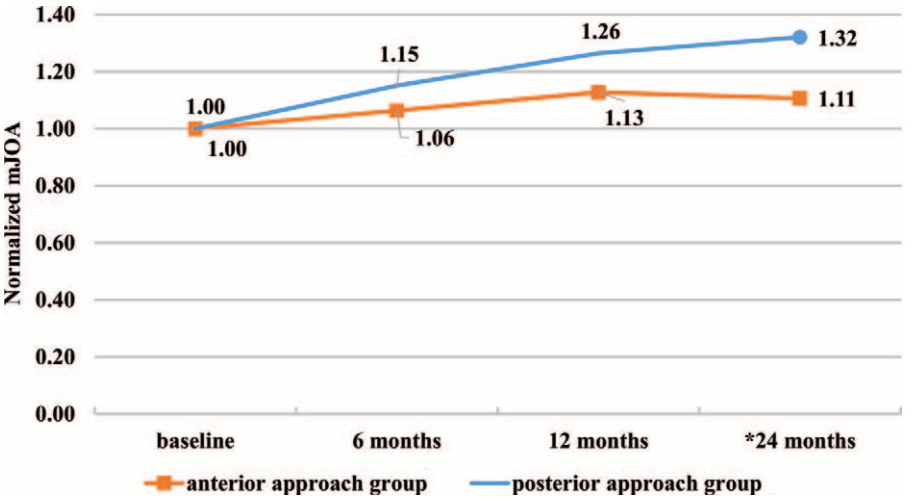

Figure 3. mJOA was normalized to each patient's baseline status. ∗Normalized mJOA was significantly different at 24 months only (P 0.004), with the posterior group demonstrating a significantly higher normalized mJOA. mJOA indicates modified Japanese Orthopaedic Association.

Copyright © 2015 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 1. AUC Analysis of mJOA by Surgical Approach

	BL to 6 month	6 month to 1 year	1 to 2 year	Total
Anterior AUC	6.52	*7.17	*14.53	*28.32 (total)
Posterior AUC	6.83	*7.96	*16.26	*31.00 (total)
P-value	0.086	0.015	0.006	0.013

*AUC analysis demonstrated that the posterior group had a higher AUC at all time intervals. This was statistically significant from the 6- to 12-month interval (P 0.015) and the 12- to 24-month interval (P 0.006). There was a significant difference in the overall recovery between the 2 cohorts (P 0.013).

AUC indicates area under the curve; BL, baseline; mJOA, mJOA, modified Japanese Orthopaedic Association.

considered confounders when evaluating differences in outcomes across the approach groups. However, since our method adjusted for baseline severity score, which is the most significant predictor of surgical outcome, our results likely provide a good estimate of differences in recovery rates between patients treated anteriorly versus posteriorly.

AUC was traditionally used in the context of ROC curves, which plot the true positive rate or ''sensitivity'' against the false positive rate (1-sensitivity) at varying thresholds.[1,2] In terms of medical application, ROC curves can be used to evaluate the predictive ability of various screening and diagnostic tests, to assess the performance of models developed to identify high risk patients, to compare the accuracy of competing prediction models and to determine a minimal clinically important difference of a measurement tool.[3]

The AUC is a measure of diagnostic or predictive accuracy and ranges from 0.5 to 1. With respect to patient screening and diagnosis, this area measures the ability of a specific test to discriminate between patients with and without a disease. AUC can also evaluate the performance of a prediction model by assessing how accurately it can distinguish between patients who achieved a specific outcome and those who did not. An AUC of 1 reflects a perfect test and an AUC of 0.5 indicates that the test or model is equivalent to random guessing. It is evident that ROC and AUC analysis are extremely valuable in diagnostic testing and predictive modeling; however, there is ongoing development of AUC analysis and interest in expanding these concepts to compare outcomes.

In this study, we normalized each patient's mJOA scores to their baseline value in order to compare recovery indices between approach groups instead of final postoperative scores. This concept is similar to how a ROC curve adjusts for variations in sensitivity and specificity that occur when changing a threshold value. We plotted these values at each time point and computed the AUC. In the context of mJOA, AUC represents the postoperative recovery state or the time a patient spends suffering (or not suffering) after surgical intervention.

In this study, a simple, static plot of the mJOA demonstrated that patients treated anteriorly had significantly better outcomes at each follow-up visit than those treated posteriorly. However, when the mJOA scores were normalized to baseline values, the posterior approach group had superior recovery indices at 12- and 24-months following surgery. These results suggest that, although patients treated anteriorly achieve higher postoperative scores, those treated posteriorly have greater potential for improvement and, as such, exhibit higher recovery rates. This study has provided an alternate method for comparing outcomes between 2 cohorts.

Clinical heterogeneity must be controlled for in order to effectively evaluate outcome differences across groups. In CSM, this heterogeneity arises from variations in causative pathologies, clinical presentations, natural histories, and even perception of disability.[8–10] In this analysis, we were unable to control for all of these factors and could only focus on adjusting for the baseline severity score.

This study represents the first to use AUC to assess postoperative recovery in patients with CSM and the applications of this novel method warrant further study. Future studies should examine mJOA recovery in a cohort of patients followed for more than 2 years. These methods could reveal how surgery impacts the natural course of the disease over a decade or how the disease progresses in patients with mild myelopathy that opt for conservative management. Further work should also apply AUC to other health-related quality of life instruments and assess the suffering index (the AUC for HRQOLs where higher scores represent worse outcomes) in CSM and other spinal pathologies.

CONCLUSION

This novel method of AUC analysis allows for comparison across dissimilar patients, which is especially important in CSM since patients exhibit wide variations in presentation and symptomatology. There were no significant differences in AUC between the anterior and posterior approach groups at 6-months; however at 12- and 24-months, patients in the posterior group exhibited superior recovery indices.

➢ Key Points

- ❑ This novel method of area under the curve (AUC) analysis allows for comparison across dissimilar patients, which is especially important in cervical spondylotic myelopathy since patients exhibit wide variations in presentation and symptomatology.
- ❑ There were no significant differences in AUC between the anterior and posterior approach groups at 6-months; however at 12 and 24 months, patients in the posterior group exhibited superior recovery indices.

Copyright © 2015 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

References

1. Bewick V, Cheek L, Ball J. Statistics review 13: receiver operating characteristic curves. *Critical Care* 2004;8:508–12.
2. McNeil BJ, Hanley JA. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. *Med Decis Making* 1984;4:137–50.
3. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 2007;115:654–7.
4. Fehlings MG, Barry S, Kopjar B, et al. Anterior versus posterior surgical approaches to treat cervical spondylotic myelopathy: outcomes of the prospective multicenter AOSpine North America CSM study in 264 patients. *Spine* 2013;38:2247–52.
5. Fehlings MG, Smith JS, Kopjar B, et al. Perioperative and delayed complications associated with the surgical treatment of cervical spondylotic myelopathy based on 302 patients from the AOSpine North America Cervical Spondylotic Myelopathy Study. *J Neurosurg Spine* 2012;16:425–32.
6. Fehlings MG, Wilson JR, Kopjar B, et al. Efficacy and safety of surgical decompression in patients with cervical spondylotic myelopathy: results of the AOSpine North America prospective multi-center study. *J Bone Joint Surg Am* 2013;95:1651–8.
7. Kalsi-Ryan S, Singh A, Massicotte EM, et al. Ancillary outcome measures for assessment of individuals with cervical spondylotic myelopathy. *Spine* 2013;38:S111–22.
8. Young WF. Cervical spondylotic myelopathy: a common cause of spinal cord dysfunction in older persons. *Am Fam Physician* 2000;62:1064–70; 73.
9. Klineberg E. Cervical spondylotic myelopathy: a review of the evidence. *Orthop Clin N Am* 2010;41:193–202.
10. Kalsi-Ryan S, Karadimas SK, Fehlings MG. Cervical spondylotic myelopathy: the clinical phenomenon and the current pathobiology of an increasingly prevalent and devastating disorder. *Neuroscientist* 2013;19:409–21.

Copyright © 2015 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.