

Functional outcomes assessment for cervical degenerative disease

LANGSTON T. HOLLY, M.D.,¹ PAUL G. MATZ, M.D.,² PAUL A. ANDERSON, M.D.,³ MICHAEL W. GROFF, M.D.,⁴ ROBERT F. HEARY, M.D.,⁵ MICHAEL G. KAISER, M.D.,⁶ PRAVEEN V. MUMMANENI, M.D.,⁷ TIMOTHY C. RYKEN, M.D.,⁸ TANVIR F. CHOUDHRI, M.D.,⁹ EDWARD J. VRESILOVIC, M.D., PH.D.,¹⁰ AND DANIEL K. RESNICK, M.D.¹¹

¹Division of Neurosurgery, David Geffen School of Medicine, University of California at Los Angeles, California; ²Division of Neurological Surgery, University of Alabama, Birmingham, Alabama; Departments of ³Orthopaedic Surgery and ¹¹Neurological Surgery, University of Wisconsin, Madison, Wisconsin; ⁴Department of Neurosurgery, Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, Massachusetts; ⁵Department of Neurosurgery, University of Medicine and Dentistry of New Jersey—New Jersey Medical School, Newark, New Jersey; ⁶Department of Neurological Surgery, Neurological Institute, Columbia University, New York, New York; ⁷Department of Neurosurgery, University of California at San Francisco, California; ⁸Department of Neurosurgery, University of Iowa Hospitals and Clinics, Iowa City, Iowa; ⁹Department of Neurosurgery, Mount Sinai School of Medicine, New York, New York; and ¹⁰Department of Orthopaedic Surgery, Milton S. Hershey Medical Center, Pennsylvania State College of Medicine, Hershey, Pennsylvania

Object. The objective of this systematic review was to use evidence-based medicine to identify valid, reliable, and responsive measures of functional outcome after treatment for cervical degenerative disease.

Methods. The National Library of Medicine and Cochrane Database were queried using MeSH headings and key words relevant to functional outcomes. Abstracts were reviewed after which studies meeting inclusion criteria were selected. The guidelines group assembled an evidentiary table summarizing the quality of evidence (Classes I–III). Disagreements regarding the level of evidence were resolved through an expert consensus conference. The group formulated recommendations that contained the degree of strength based on the Scottish Intercollegiate Guidelines network. Validation was done through peer review by the Joint Guidelines Committee of the American Association of Neurological Surgeons/Congress of Neurological Surgeons.

Results. Myelopathy Disability Index, Japanese Orthopaedic Association scale, 36-Item Short Form Health Survey, and gait analysis were found to be valid and reliable measures (Class II) for assessing cervical spondylotic myelopathy. The Patient-Specific Functional Scale, the North American Spine Society scale, and the Neck Disability Index were found to be reliable, valid, and responsive (Class II) for assessing radiculopathy for nonoperative therapy. The Cervical Spine Outcomes Questionnaire was a reliable and valid method (Class II) to assess operative therapy for cervical radiculopathy.

Conclusions. Several functional outcome measures are available to assess cervical spondylotic myelopathy and cervical radiculopathy. (DOI: 10.3171/2009.2.SPINE08715)

KEY WORDS • cervical spine • functional outcome • myelopathy • practice guidelines • radiculopathy

Recommendations

IT is recommended that functional outcome measures—such as the MDI, JOA, and SF-36 scales, and gait analysis—be used in the assessment of patients undergoing surgery for CSM because they have proven

Abbreviations used in this paper: CSM = cervical spondylotic myelopathy; CSOQ = Cervical Spine Outcomes Questionnaire; EMS = European Myelopathy Score; ICC = intraclass correlation coefficient; JOA = Japanese Orthopaedic Association; MDI = Myelopathy Disability Index; NASS = North American Spine Society; NDI = Neck Disability Index; ODI = Oswestry Disability Index; PSFS = Patient-Specific Functional Scale; SF-12 = 12-Item Short Form Health Survey; SF-36 = 36-Item Short Form Health Survey.

to be valid and reliable in this setting (quality of evidence, Class II; strength of recommendation, B).

It is recommended that functional assessment of cervical radiculopathy in patients undergoing nonoperative therapy be undertaken through the PSFS, which has been shown to be reliable, valid, and responsive in this setting (quality of evidence, Class II; strength of recommendation, C). Other options to follow functional improvement for nonoperative therapy of cervical radiculopathy are the NASS scale, the NDI (quality of evidence, Class II; strength of recommendation, C).

It is recommended that functional assessment of cervical radiculopathy in patients undergoing operative therapy be undertaken using the CSOQ (quality of evidence, Class II; strength of recommendation, C).

Rationale

Cervical spine surgery is frequently advocated in the management of common spinal disorders such as CSM and radiculopathy. A variety of different surgical treatment options exist for these conditions including anterior cervical discectomy and fusion, anterior corpectomy, posterior foraminotomy, laminectomy, laminectomy and fusion, and laminoplasty. Our review of the medical literature yielded numerous citations supporting the advantages of each of these individual techniques; not surprisingly, controversy exists regarding the selection of the optimal surgical treatment. One of the challenges in defining surgical treatment strategies for cervical spine disease is the prior use of subjective outcome measures based largely on the surgeon's judgment or impression of patient outcome. Studies have shown a potential disconnect between physician-expected outcomes and actual patient-reported functional outcomes such as pain, work-related activities, and social/recreational activities.²

The objective of this review was to identify valid, reliable, and responsive measures of functional outcome after treatment for cervical degenerative disease. The prevalence of cervical spine disease, the variety of treatment options available, and the economic impact of treatment in these patients make the implementation of suitable functional outcome measures a high priority. The advent of novel surgical techniques, advances in spinal instrumentation, and development of osteobiologics further necessitate the rigorous analysis of surgical outcomes.

Search Criteria

The group completed a computerized search of the Cochrane Database and the National Library of Medicine Database of the literature published between 1966 and 2007 using keywords and MeSH headings. A search using the subject heading "cervical spine surgery" yielded 9537 citations. The following subject headings were combined: "cervical spine surgery and outcomes" and 324 citations were obtained. A search using the headings "cervical spondylotic myelopathy and outcomes" provided 42 citations, and "cervical radiculopathy and outcomes" yielded 106 citations. Alternative searches included each outcome measurement scale by name. We evaluated abstracts and titles of the aforementioned citations and selected articles that focused on cervical spine surgery outcome measurements for detailed review. We also chose additional manuscripts from the reference lists of selected articles. Among the articles reviewed, we found 11 studies of cervical degenerative disease and functional outcome. These studies formed the basis of the evidentiary table (Table 1).

Scientific Foundation

To assess outcomes accurately following an intervention, a functional instrument must have 3 important characteristics: validity, responsiveness, and reliability.^{9,18} Validity is the ability of the instrument to measure the specific function or property that it was designed to as-

sess. There are 3 key components to valid outcome measures. Content validity ensures that the instrument's questions will accurately portray the concepts that they are designed to examine. Criterion validity is the correlation between the instrument's measurements and other accepted criteria. Lastly, construct validity is the correlation between the instrument's measurements based on well-developed theories or hypotheses. Responsiveness is the ability of the instrument to detect clinically significant changes in the function being measured. It is also desirable for an instrument to show a large sensitivity to change (such as the magnitude of the change) as well as to distinguish between differences in function severity among populations.

Reliability refers to the ability of the outcome tool to yield reproducible measurements over time or across methods of obtaining data. Test-retest (external) reliability is the stability of responses or outcomes after testing at 2 different time points (provided the clinical condition has not significantly changed). Interrater reliability is the ability of an instrument to yield similar results if different testers apply the measurement to the same or comparable populations. Internal reliability is important for a multidomain instrument because each component of a multicomponent outcome measure should correlate with the final result. Cronbach's alpha test is a widely accepted method of determining internal consistency, and a score of 1 indicates perfect correlation and high reliability between different components of the same scale.⁷ Based on the criteria described by Nunnally and Bernstein,¹⁸ alpha scores > 0.7 demonstrate acceptable consistency. The kappa value corresponds to the degree of agreement of interrater observations, and in patient-based outcome measurements denotes consistency in response at a given time point.¹⁵ In keeping with prior guidelines work, a kappa value > 0.8 is ideal, while a kappa value of 0.6 is very good.²²

Cervical Myelopathy Outcome Measures

Singh and Crockard²³ evaluated 100 patients with CSM who underwent functional assessment both preoperatively and 6 months after decompressive surgery. The study used 7 different scales, including the MDI,⁴ JOA,³⁰ EMS,¹² Nurick score,¹⁹ Ranawat score,²¹ Odom's criteria,²⁰ and SF-36.³ They analyzed outcome measures with respect to internal consistency, sensitivity, validity, and responsiveness. All of the scales demonstrated responsiveness, as each showed a statistically significant clinical improvement following surgery ($p < 0.001$). Sensitivity to change was quantified by the normalized change (difference in preoperative score and postoperative score divided by the median of all scores). The MDI was the most sensitive to change and therefore the best scale to demonstrate the magnitude of clinical change. The EMS was the least sensitive, and the remaining scales were evenly distributed in between. Cronbach's alpha test confirmed internal consistency for each of the multidomain outcome measures (preoperative, postoperative alpha)—MDI (0.92, 0.95), SF-36 (0.82, 0.86), JOA (0.72, 0.77), and EMS (0.68, 0.77). The Ranawat, Nurick, and Odom's

TABLE 1: Evidentiary summary of studies regarding functional outcome measures and cervical degenerative disease*

Authors & Year	Description	Results	Class	Conclusions
Singh & Crockard, 2001	100 patients w/ CSM underwent functional assessment preoperatively & 6 mos after decompressive surgery. 7 different scales were used including MDI, JOA, EMS, Nurick Score, Ranawat score, Odom's criteria, & the SF-36. Outcome measures were analyzed w/ respect to internal consistency, sensitivity, validity, & responsiveness.	All scales demonstrated responsiveness, as each showed statistically significant postop clinical improvement ($p < 0.001$). However, each scale demonstrated varying degrees of reliability, validity, & responsiveness that made it more or less appropriate to evaluate different functional parameters.	II	The MDI showed the greatest sensitivity between severity levels, sensitivity to operative change, & reliability.
Yonenobu et al., 2001	The inter- & intraobserver reliabilities of the JOA scoring system for CSM were evaluated in a group of 29 patients w/ myelopathy. Several different surgeons twice evaluated the patients at intervals of 1–6 weeks.	The inter- & intraobserver reliability of the total score were high. The level of experience & the hospital slightly affected the reliability of the scoring system.	II	The inter- & intraobserver reliabilities of the JOA scoring system were high, indicating that it is a useful functional outcome measure in CSM patients.
Casey et al., 1996	The reliability, validity, & responsiveness of the MDI were evaluated in 250 patients w/ RA & cervical spine disease. The scale was also evaluated by comparing pre- & postop MDI scores w/ the Steinbrocker & Ranawat scoring systems in a subgroup of 192 patients who eventually underwent surgery.	MDI demonstrated internal consistency w/ a Cronbach's alpha value of 0.95. The MDI was able to predict postop outcome as it correlated well w/ the comparative postop outcome measures ($p < 0.0001$).	II	The MDI provides a reliable method of assessing disability in patients w/ RA & cervical spine disease, & predicts postop outcome.
King & Roberts, 2002	The SF-36 was administered to 88 CSM patients over a 12-mo period. Construct validity was determined by observing the SF-36 score variance w/ the myelopathy scales of Nurick, Cooper, Harsh, & mJOA scales.	Construct validity was achieved by confirming the relationship between the SF-36, & the selected functional outcome measures for CSM.	II	The SF-36 is a valid & reliable functional outcome measure for CSM patients.
Singh et al., 2006	105 CSM patients undergoing cervical decompression were evaluated pre- & postoperatively using the SF-36 & the SF-12 forms. The validity, reliability, & sensitivity to change of these assessment scales were compared.	The reliability of the SF-36 was marginally better than the SF-12, but the sensitivity to change & the absolute sensitivity of both scales were comparable.	II	Both the SF-36 & SF-12 are valid & sensitive to changes in physical & mental health in CSM patients undergoing surgical decompression.
Singh & Crockard, 1999	41 CSM patients underwent a 30-m walking test before & after surgical decompression. Preop neurological status & functional outcome were also compared using the MDI & Nurick scales.	Gait analysis demonstrated a significant mean postop improvement in walking time & no. of steps taken. Pre- & postop walking scores were significantly correlated w/ the MDI & Nurick scores.	II	Timed walks are a reliable & valid method of assessing CSM patients & the effects of decompressive surgery.
Moorthy et al., 2005	6 CSM patients underwent pre- & postop gait analysis following anterior corpectomy for myelopathy. Changes in gait parameters were compared to the Nurick & mJOA scoring systems.	Significant postop neurological improvement was obtained in gait parameters such as walking speed, stride length, percentage of single length stance time, etc. Improvement in postop gait parameters correlated w/ improvements in mJOA & Nurick scores.	II	Gait analysis can be used as a quantitative functional tool in the preop & postop evaluation of CSM patients.
Kuhtz-Buschbeck et al., 1999	12 CSM patients underwent gait analysis before & after decompressive surgery. The patients were compared to a matched control group of healthy volunteers.	Compared to controls, CSM patients demonstrated abnormal gait parameters such as reduced gait velocity, step length, & prolonged double support. Postoperatively, the patients obtained a significant improvement in several of these parameters.	II	Gait analysis is an objective tool to document functional recovery after decompressive surgery in CSM.

(continued)

TABLE 1: Evidentiary summary of studies regarding functional outcome measures and cervical degenerative disease* (continued)

Authors & Year	Description	Results	Class	Conclusions
Cleland et al., 2006	The reliability, validity, & responsiveness of the NDI & PSFS were evaluated in 38 patients w/ cervical radiculopathy undergoing physical therapy. The patients were evaluated at baseline & at the end of the Tx period.	The PSFS was more responsive to change than the NDI. Additionally, the test-retest reliability was also higher for the PSFS.	II	The PSFS demonstrated superior reliability, construct validity, & responsiveness in this group of cervical radiculopathy patients than the NDI.
BenDebba et al., 2002	216 patients who underwent surgery for a cervical spine disorder were evaluated using the CSOQ preoperatively, then at 3 & 6 mos postoperatively. The data were then used to assess the validity, reliability, & responsiveness of the questionnaire.	The CSOQ has good construct validity, high test-retest reliability, & responsiveness to change after Tx.	II	The CSOQ is a comprehensive disease specific instrument for evaluating complaints of neck pain & assessing outcomes of Tx. The outcome measure is highly reliable, valid, & responsive.
Stoll et al., 2004	The validity & sensitivity to change of the cervical NASS questionnaire was evaluated in a cohort of 140 patients w/ cervical spine disorders referred either to an inpatient rehabilitation or an outpatient physical therapy program. The NASS questionnaire & SF-36 were administered at baseline & at the conclusion of Tx.	Criterion validity was established by the significant correlation between the NASS & SF-36 subscores. The NASS subscores also demonstrated good sensitivity to change.	II	The NASS patient questionnaire is a valid outcome instrument in patients w/ disorders of the cervical spine.

* The criteria for scoring each manuscript into a class are described in *Introduction and Methodology: Guidelines for the Surgical Management of Cervical Degenerative Disease*, which appears in this issue of the *Journal of Neurosurgery: Spine*. Abbreviations: mJOA = modified JOA; RA = rheumatoid arthritis.

scales are simple, 1-dimensional instruments and do not receive an alpha value. The validity of the different scales was evaluated by calculating the correlation coefficient for pre- and postoperative scores, and postoperative changes. The best correlation was between the MDI and EMS scales ($r = 0.82$).

Yonenobu et al.³⁰ evaluated the inter- and intraobserver reliabilities of the JOA scale for CSM in a cohort of 29 patients with stable myelopathy. Thirteen patients had severe myelopathy (JOA scale score < 9), and moderate (JOA scale Score 9–13) and mild (JOA scale score > 13) myelopathies were present in 8 patients each. These patients had not noticed any neurological changes over the previous month. The authors considered them unlikely to suffer any neurological worsening in the immediate future. Three groups of surgeons with varying levels of experience (high, moderate, and low) from 5 different hospitals participated in the study. A minimum of 3 observers of the hospital independently interviewed each patient on the same day. The same patients were then re-interviewed by the same surgeons in the same fashion at intervals of 1–6 weeks. The reliability of the JOA scale was assessed by determining the intraclass correlation coefficient. The interobserver reliability for the initial interview was high (0.813). The inter- and intraobserver reliability calculated by combining the data from the 2 visits was also high (0.826). The effect of observer experience or practice location was not significant to the JOA score. The interobserver reliability of each independent JOA category was also assessed using the kappa value. Agreement between the 2 interviews was high for motor function of the fingers, sensory function of the trunk, and bladder function. The kappa value was lower for motor function of the shoulder and elbow, as well as lower extremity sensation. One limitation of the study was that no other outcome measure was compared with the JOA scale. However, nearly 80% of the data pairs between the 2 interviews were within ± 1 point, further supporting the reliability of the JOA scale.

Casey et al.⁴ evaluated the reliability, validity, and responsiveness of the MDI in 250 patients with rheumatoid arthritis and cervical spine disease who had been referred for surgical intervention. The authors developed the MDI by making disease-specific modifications to the 20-question Stanford Health Assessment Questionnaire, a functional assessment tool that has been validated in the evaluation of patients with rheumatoid arthritis.¹⁰ The MDI proved to be reliable, as internal consistency was demonstrated with a Cronbach's alpha value of 0.95. The correlation between the MDI and the health assessment questionnaire was $r = 0.98$, indicating conservation of information and the presence of criterion validity. Criterion validity was further evaluated by comparing pre- and postoperative MDI scores with the Steinbrocker et al.²⁶ and Ranawat et al.²¹ scoring systems in a subgroup of 192 of these patients who eventually underwent surgery. The MDI was able to predict postoperative outcome as it correlated well with the aforementioned postsurgical outcome measures ($p < 0.0001$). The MDI also demonstrated suitable responsiveness with a standardized response mean value classified within the moderately sensitive range.

King and Roberts¹⁴ administered the SF-36 questionnaire to a group of 88 patients with CSM at a Veterans Association Medical Center over a 12-month period. The patients underwent a detailed medical history including demographics, personal habits, and CSM symptomatology, as well as review of radiological imaging and a neurological examination. Based on their symptoms and examination findings, the authors compiled individual scores for the Nurick,¹⁹ Cooper,⁶ Harsh,¹¹ and JOA myelopathy scales. Construct validity was determined by assessing whether higher scores on the mobility-related SF-36 domains (physical functioning, role functioning-physical, general health, social functioning, and physical component scores) correlated with higher functioning on the myelopathy scales. Analysis using the Cuzick nonparametric test for trend⁸ demonstrated that higher scores on the relevant SF-36 domains corresponded to better functioning on the myelopathy scales of Nurick ($p \leq 0.003$), Cooper leg subscale ($p \leq 0.012$; except the general health perceptions domain [$p = 0.091$]), Harsh ($p \leq 0.016$), and the motor component of the modified JOA scale ($p \leq 0.006$). Cronbach alpha values ranging from 0.79 to 0.92 confirmed reliability for each of the SF-36 subscales.

Singh and colleagues²⁵ evaluated the validity, reliability, and sensitivity to change of the SF-12, an abbreviated version of the SF-36, in a group of 105 patients with CSM who underwent decompression surgery. Patients prospectively completed the SF-36 questionnaire before and again 6 months after surgery. The SF-12 responses were culled from the SF-36 form, and the data from the physical component and mental component summary were compared. There were significant postoperative improvements in the physical and mental components of both the SF-12 and the SF-36 ($p < 0.001$). The internal consistency was marginally higher for the SF-36 than the SF-12, yet the SF-12 alpha value of 0.77 still demonstrated suitable reliability. The sensitivity to change and the absolute sensitivity for both scales were comparable. There were close and linear correlations between the pre- and postoperative physical and mental components as well ($r = 0.86-0.93$; $p < 0.0001$).

Gait Analysis

In addition to the aforementioned scoring scales, gait analysis has also been proven to be a valid and reliable outcome measurement tool in patients undergoing surgery for CSM. Singh and Crockard²⁴ used a walking test, MDI, and Nurick grades to measure severity of CSM and functional outcome after surgical decompression. Forty-one patients with CSM underwent gait analysis examining walking time and number of steps taken over 30 m preoperatively and again 2 months postoperatively. Each patient performed 3 trials of ambulation at both time points, and the mean values were calculated. The walking time data were highly reproducible and external reliability (test-retest) was proven as there were no statistically significant changes between trials ($p = 0.995$). As expected, there were significant variations in pre- and postoperative walking times ($p < 0.001$), suggesting that the detected changes probably represented actual alterations in func-

tional status after surgery. The authors observed similar results when comparing number of steps taken between trials ($p = 0.981$) and pre- and postoperative values ($p = 0.003$). Mean MDI and Nurick scores showed significant postoperative improvement ($p < 0.0001$). Preoperative and postoperative walking scores significantly and equally correlated with the MDI and Nurick scores.

Moorthy et al.¹⁷ performed pre- and postoperative quantitative gait analysis in 6 patients with CSM who underwent anterior corpectomy. They found that all patients had significant postoperative improvement in ambulation parameters such as walking speed, stride length, and percentage of single-limb stance time. These changes correlated with functional improvement as determined both by mean Nurick ($p = 0.02$) and JOA lower limb ($p = 0.02$) scores. Kuhtz-Buschbeck et al.¹⁶ similarly used gait analysis to evaluate patients with CSM, and concluded that this technique is an effective tool for measurement of functional recovery after decompression surgery.

Cervical Radiculopathy

The NDI and PSFS have been shown to be valid and reliable in the evaluation of patients with neck pain.^{28,29} Cleland et al.⁵ assessed the reliability, validity, and responsiveness of the NDI and PSFS in 38 patients with cervical radiculopathy undergoing physical therapy. The participants completed the NDI and PSFS at baseline and at the conclusion of treatment. The patients also performed a 15-point global rating of change¹³ at the last follow-up examination. This instrument was used to stratify the patients as either improved or stable. The PSFS demonstrated excellent test-retest reliability (ICC = 0.82), and the NDI also manifested adequate reliability (ICC = 0.68). Construct validity was determined by comparing the baseline and follow-up scores for both the stable and improved groups. The PSFS showed construct validity as there was a significant difference in scores between stable and improved patients based on the global rating of change ($p < 0.001$). However, the NDI failed to demonstrate construct validity as there was no statistical difference in scores between stable and improved patients. Lastly, the PSFS showed superior responsiveness to change than the NDI: the minimal detectable change for the PSFS was 2.1, compared with 10.2 for the NDI.

In their large multicenter study, BenDebba et al.¹ used the CSOQ in the evaluation of 216 patients who underwent surgery for cervical spine disorders. Approximately 60% of patients presented with radiculopathy, 21% with myelopathy, and the remainder with neck pain. The patients completed the CSOQ, ODI questionnaire, and the SF-36 preoperatively, and at 3 and 6 months postoperatively. The test-retest reliability of the CSOQ was demonstrated by ICCs ranging from 0.75 to 0.85 for the 6 component measures. Construct validity was ascertained as component subscores correlated with the corresponding components of the ODI and SF-36 (that is, the pain severity scores of the CSOQ and the bodily pain scores of the ODI and SF-36). Responsiveness was demonstrated as the mean score change between improved and unimproved patients, and was statistically significant in all 6 categories ($p < 0.0001$) except for healthcare utilization.

Stoll et al.²⁷ evaluated the validity and sensitivity to change of the cervical NASS questionnaire in a group of 140 patients with cervical spine disorders (including radiculopathy and neck pain) that were referred to either an inpatient rehabilitation or outpatient physical therapy program. The patients completed the NASS questionnaire and the SF-36 immediately before commencing the inpatient rehabilitation or physical therapy. The patients completed the same questionnaires again after completing the treatment regimen. Criterion validity for the cervical NASS questionnaire was established by the strong correlations between the NASS subscores and SF-36. Not surprisingly, the NASS subscore Pain and Disability showed the most correlation with the SF-36 subscores Physical Function and Pain (Spearman rho = 0.75 and 0.65). The discriminative validity of the cervical NASS questionnaire was demonstrated by the fact that patients referred for outpatient treatment had significantly higher functional and health status scores than those referred for inpatient rehabilitation. Moreover, the NASS questionnaire documented statistically significant clinical improvement after treatment. Improvement was manifested in both cohorts, and was in agreement with the SF-36 subscores. Lastly, the NASS Pain and Disability subscore demonstrated satisfactory responsiveness and sensitivity to change (standard response mean 0.64–1.24).

Summary

Because the operative and nonoperative management of cervical spine disorders has become increasingly prevalent, it is essential that appropriate functional outcomes measures are used to assess efficacy of treatment. Valid and reliable outcome measures must demonstrate validity, reliability, and responsiveness. Outcome measurements supported by Class II medical data for the evaluation of CSM include the MDI, JOA, SF-36, SF-12, and gait analysis. The CSOQ was valid and reliable in measuring functional outcomes following cervical spine surgery in a mixed group of patients who presented with radiculopathy, myelopathy, and neck pain. The PSFS and cervical NASS were valid and reliable for assessing outcomes in patients with cervical radiculopathy undergoing physical therapy. Patient satisfaction surveys provide important information regarding the treatment experience; however, they appear unable to provide the necessary reliability, and cannot measure responsiveness.

Key Issues for Future Investigation

Although a variety of functional outcome instruments have been validated for assessing patients with CSM, there has been a relative paucity of instruments for evaluating surgical patients with radiculopathy. The existing literature has been validated for nonoperative treatment of radiculopathy, and surgical treatment of a mixed group of patients with cervical spine disorders. Future studies should identify valid outcome measurements for patients undergoing surgical treatment of cervical radiculopathy.

Disclosure

Administrative costs of this project were funded by the Joint Section on Disorders of the Spine and Peripheral Nerves of the American Association of Neurological Surgeons and Congress of Neurological Surgeons. No author received payment or honorarium for time devoted to this project. Dr. Resnick owns stock in Orthovita. Dr. Matz receives support from the Kyphon Grant for Thoracolumbar Fracture Study, and an advisory honorarium from Synthes for the cadaver laboratory. Dr. Heary receives support from DePuy Spine and Biomet Spine, and receives royalties from DePuy Spine and Zimmer Spine. Dr. Groff is a consultant for DePuy Spine. Dr. Mummaneni is a consultant for and receives university grants from DePuy Spine and Medtronic, Inc., and is a patent holder in DePuy Spine. Dr. Anderson is an owner of, consultant for, and stockholder of Pioneer Surgical Technology; a consultant for and receives non-study related support from Medtronic, Inc.; and is a patent holder in Stryker. The authors report no other conflicts of interest concerning the materials or methods used in this study or the findings specified in this paper.

References

1. BenDebba M, Heller J, Ducker TB, Eisenger JM: Cervical spine outcomes questionnaire. **Spine** 27:2116–2124, 2002
2. BenDebba M, Long DL, Ducker T, Torgerson W: Physician-expected treatment outcome and actual outcome in patients treated for neck pain. **Orthop Trans** 20:446–447, 1996
3. Brazier JE, Harper R, Jones NM, O’Cathain A, Thomas KJ, Usherwood T, et al: Validating the SF-36 health survey questionnaire: new outcome measure for primary care. **BMJ** 305:160–164, 1992
4. Casey ATH, Bland JM, Crockard HA: Development of a functional scoring system for rheumatoid arthritis patients with cervical myelopathy. **Ann Rheum Dis** 55:901–906, 1996
5. Cleland JA, Fritz M, Whitman JM, Palmer JA: The reliability and construct validity of the neck disability index and patient specific functional scale in patients with cervical radiculopathy. **Spine** 31:598–602, 2006
6. Cooper PR, Epstein F: Radical resection of intramedullary spinal cord tumors in adults. Recent experience in 29 patients. **J Neurosurg** 63:492–499, 1985
7. Cronbach LJ: Coefficient alpha and the internal structure of tests. **Psychometrika** 16:297–334, 1951
8. Cuzick J: A Wilcoxon-type test for trend. **Stat Med** 4:87–90, 1985
9. Deyo RA, Diehr P, Patrick DL: Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. **Control Clin Trials** 12 (4 Suppl):142S–158S, 1991
10. Fries JF, Spitz P, Kraines RG, Holman HR: Measurement of patient outcome in arthritis. **Arthritis Rheum** 23:137–145, 1980
11. Harsh GR IV, Sybert GW, Weinstein PR: Cervical spine stenosis due to ossification of the posterior longitudinal ligament. **J Neurosurg** 67:349–357, 1987
12. Herdman J, Linzbach M, Krzan M: The European myelopathy score, in Bauer BL, Brock M, Klingner M (eds): **Advances in Neurosurgery**. Berlin: Springer, 1994, pp 266–268
13. Jaeschke R, Singer J, Guyatt GH: Measurement of health status. Ascertain the minimal clinically important difference. **Control Clin Trials** 10:407–415, 1989
14. King JT Jr, Roberts MS: Validity and reliability of the SF-36 in cervical spondylotic myelopathy. **J Neurosurg** 97 (2 Suppl): 180–185, 2002
15. Kraemer HC: Extension of the kappa coefficient. **Biometrics** 36:207–216, 1980
16. Kultz-Buschbeck JP, Johnk K, Mader S, Stolze H, Medhorn M: Analysis of gait in cervical myelopathy. **Gait Posture** 9: 184–189, 1999

17. Moorthy RK, Bhattacharji S, Thayumanasamy G, Rajshekhar V: Quantitative changes in gait parameters after central corpectomy for cervical spondylotic myelopathy. **J Neurosurg Spine** **2**:418–424, 2005
18. Nunnally JC, Bernstein IH: **Psychometric Theory**, ed 2. New York: McGraw-Hill, 1994
19. Nurick S: The pathogenesis of the spinal cord disorder as associated with cervical spondylosis. **Brain** **95**:87–100, 1972
20. Odom GL, Finney W, Woodall B: Cervical disc lesions. **JAMA** **166**:23–28, 1958
21. Ranawat CS, O'Leary P, Pellicci P, Tsairis P, Marchisello P, Dorr L: Cervical fusion in rheumatoid arthritis. **J Bone Joint Surg Am** **61**:1003–1010, 1979
22. Resnick DK, Choudhri TF, Dailey AT, Groff MW, Khoo L, Matz PG: Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 2: assessment of functional outcome. **J Neurosurg Spine** **2**:639–646, 2005
23. Singh A, Crockard HA: Comparison of seven different scales used to quantify severity of cervical spondylotic myelopathy and postoperative improvement. **J Outcome Meas** **5**:798–818, 2001
24. Singh A, Crockard HA: Quantitative assessment of cervical spondylotic myelopathy by a simple walking test. **Lancet** **354**:370–373, 1999
25. Singh A, Gnanalingham K, Casey A, Crockard A: Quality of life assessment using the Short Form-12 (SF-12) questionnaire in patients with cervical spondylotic myelopathy: comparison with SF-36. **Spine** **31**:639–643, 2006
26. Steinbrocker O, Traeger CH, Batterman RC: Therapeutic criteria in rheumatoid arthritis. **JAMA** **140**:659–662, 1949
27. Stoll T, Huber E, Bachmann S, Baumeler HR, Mariacher S, Rutz M, et al: Validity and sensitivity to change of the NASS questionnaire for patients with cervical spine disorders. **Spine** **29**:2851–2855, 2004
28. Vernon H, Mior S: The neck disability index: a study of reliability and validity. **J Manipulative Physiol Ther** **14**:409–415, 1991
29. Westaway MD, Stratford PW, Binkley JM: The patient-specific functional scale: validation of its use in persons with neck dysfunction. **J Orthop Sports Phys Ther** **27**:331–338, 1998
30. Yonenobu K, Abumi K, Nagata K, Taketomi E, Ueyama K: Interobserver and intraobserver reliability of the Japanese orthopaedic association scoring system for evaluation of cervical compression myelopathy. **Spine** **26**:1890–1895, 2001

Manuscript submitted October 18, 2008.

Accepted February 19, 2009.

Address correspondence to: Paul G. Matz, M.D., Neurosurgery and Neurology, LLC, 232 South Woods Mill Road, Chesterfield, Missouri 63017. email: matzpg@yahoo.com.