Correlation of quality of life and functional outcome measures for cervical spondylotic myelopathy

Daniel Lubelski, MD, Matthew D. Alvin, MD, MBA, MS, MA, Sergiy Nesterenko, MD, Swetha J. Sundar, BS, Nicolas R. Thompson, MS, Edward C. Benzel, MD, and Thomas E. Mroz, MD

OBJECT  Studies comparing surgical treatments for cervical spondylotic myelopathy (CSM) are heterogeneous, using a variety of different quality of life (QOL) outcomes and myelopathy-specific measures. This study sought to evaluate the relationship of these measures to each other, and to better understand their use in evaluating patients with CSM.

METHODS  A retrospective study was performed in all patients with CSM who underwent either ventral or dorsal cervical spine surgery at a single tertiary-care institution between January 2008 and July 2013. Severity of myelopathy was assessed pre- and postoperatively using both the Nurick scale and the modified Japanese Orthopaedic Association (mJOA) classification of disability. Prospectively collected QOL outcomes data included Pain Disability Questionnaire (PDQ), Patient Health Questionnaire–9 (PHQ-9), and EQ-5D. Spearman rank correlations were calculated to assess the construct convergent validity for each pair of health status measures (HSMs). To assess each HSM’s ability to discriminate favorable EQ-5D index, we performed receiver operating characteristic (ROC) curve analysis and assessed the area under the curve (AUC).

RESULTS  A total of 119 patients were included. The PDQ total score had the highest correlation with EQ-5D index (Spearman’s rho = −0.82). Neither of the myelopathy scales (mJOA or Nurick) had strong correlations between themselves (0.41) or with the other QOL measures (absolute value range 0.13–0.49). In contrast, the QOL measures correlated relatively well with each other (absolute value range 0.68–0.97). For predicting favorable EQ-5D outcomes, PDQ total score had an AUC of 0.909. The AUCs were significantly greater for the QOL measures in comparison with the myelopathy measures (AUCs were 0.677 and 0.607 for mJOA and Nurick scale scores, respectively).

CONCLUSIONS  The authors found that all included measures of QOL and CSM-specific (mJOA or Nurick scale) measures were valid and responsive. The PDQ was the most predictive of positive QOL after surgery (as measured by the EQ-5D index) for patients with CSM. The substantially lower correlation between myelopathy and QOL outcomes, compared with the various QOL measures themselves, suggests that these questionnaires are measuring different aspects of the patient experience. Solely assessing the myelopathy or disease-specific signs and symptoms is likely insufficient to fully understand and appreciate clinical outcome in its totality. These questionnaire types should be used together to best evaluate patients pre- and postoperatively.

http://thejns.org/doi/abs/10.3171/2015.6.SPINE159

KEY WORDS  quality of life; EQ-5D; Patient Health Questionnaire; PHQ-9; Pain Disability Questionnaire; PDQ; mJOA; Nurick; cervical spondylotic myelopathy

ABBREVIATIONS  AUC = area under the curve; CSM = cervical spondylotic myelopathy; HSM = health status measure; MDI = Myelopathy Disability Index; mJOA = modified Japanese Orthopaedic Association; ODI = Oswestry Disability Index; PDQ = Pain Disability Questionnaire; PHQ-9 = Patient Health Questionnaire–9; QOL = quality of life; ROC = receiver operating characteristic.


INCLUDE WHEN CITING  Published online November 27, 2015; DOI: 10.3171/2015.6.SPINE159.
Cervical spondylotic myelopathy (CSM) is the most common cause of spinal cord pathology; it can be treated using multiple approaches. Clinical presentation is variable and includes numbness, weakness, neck pain, loss of dexterity, and gait problems. Depending on the patient’s presentation, as well as the extent and nature of the pathological process, treatment options can range from conservative measures to decompression of the spinal cord with or without the addition of fusion. Furthermore, there are a number of operative approaches that are commonly considered to be within the standard of care. Much discussion has revolved around the comparative advantages and disadvantages.

Very few retrospective and small prospective cohort studies have been published evaluating the different surgical approaches for CSM. Most have shown no statistically significant differences in quality of life (QOL) outcomes before and after surgery. One of the limitations of these studies, however, is their heterogeneous study design, including various questionnaires used at differing time points. Outcomes questionnaires have included disease-specific measures that attempt to quantify the extent of myelopathic symptoms (e.g., modified Japanese Orthopaedic Association [mJOA], and Nurick scale), as well as instruments that capture general disability (Pain Disability Questionnaire [PDQ]), depression (Patient Health Questionnaire—9 [PHQ-9]), and overall quality of life (EQ-5D).

In the present study we sought to 1) examine the construct convergent validity of outcome measures (Nurick scale, mJOA, EQ-5D, PHQ-9, PDQ functional status, PDQ psychosocial status, PDQ total), 2) investigate the responsiveness of the outcome measures, and then 3) further understand the ability of the various health status measures (HSMs) to predict favorable quality of life (EQ-5D) in patients undergoing treatment for CSM.

Methods

Using the institutional electronic medical records, we performed a retrospective study in all patients with CSM who underwent either ventral or dorsal cervical spine surgery at a single tertiary-care institution between January 2008 and July 2013. CSM was defined as cervical spinal cord compression from degenerative spondylosis with corresponding myelopathic symptoms including extremity weakness, hand clumsiness, gait problems, bowel/bladder dysfunction, hyperreflexia, and/or long-tract signs. While not all patients had all questionnaires completed at each time point, all included patients had at least one of the included questionnaires completed at both pre- and postoperative periods. Preoperative scores were defined as those obtained within 30 days prior to surgery, and 1-year postoperative scores were based on the office visit closest to 1 year following surgery (mean 419 days after surgery).

Myelopathy Measures

Severity of myelopathy was assessed pre- and postoperatively using both the Nurick scale and the mJOA classification of disability. Both of these measures are commonly used myelopathy scales. A lower mJOA score is associated with a greater neurological deficit; this scale ranges from 1 to 17, with normal function associated with a score of 16 or 17. For the Nurick scale scores, a higher score indicates greater severity (range 1–5). The Nurick scale has been validated in a single study among CSM patients. Scores were obtained by chart review based on completed questionnaires at the time of the outpatient visit. Two independent reviewers (S.N. and S.J.S.) reviewed the notes, and discrepancies were resolved by consensus.

Quality of Life Outcomes

Prospectively collected QOL outcomes data included PDQ, PHQ-9, and EQ-5D. These were acquired via the internal institutional Knowledge Program. The Knowledge Program is a patient-derived outcome assessment tool that is embedded in our electronic medical record. These data have been collected since 2008, in a prospective fashion.

The PHQ-9 is a self-administered assessment for depression that evaluates the 9 DSM-IV criteria for major depressive disorder. Each of the 9 questions is scored from 0 (“not at all”) to 3 (“nearly every day”), making the total score range from 0 to 27. The total score is commonly categorized into one of 5 groups: minimal (score 0–4), mild (score 5–9), moderate (score 10–14), moderate-severe (score 15–19), and severe (score 20–27). The diagnostic validity of the PHQ-9 has been established in large multicenter analyses.

The PDQ evaluates chronic disabling musculoskeletal and spinal disorders, focusing on disability and patients’ ability to perform activities of daily living. The score ranges from 0 (optimal function) to 150 (total disability). The EQ-5D contains 5 dimensions of health state: Mobility, Self-Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression. Each dimension is scored from 1 to 3. The minimum clinically important difference used for PDQ was 16, for PHQ-9 it was 5, and for EQ-5D it was 0.1.

Statistical Methods

Summary statistics (e.g., means, standard deviations, frequencies, percentages) were calculated for all eligible patients using R version 2.15.13 (R Core Team, 2013). For the first goal of assessing the construct convergent validity of the various measures, we combined scores taken preoperatively and 1 year postoperatively to increase the sample size. For each pair of the HSMs, we computed the Spearman rank correlation to assess the construct convergent validity. For the second objective of investigating the responsiveness of each scale, a paired t-test on each scale’s preoperative and 1-year postoperative scores was performed. Lastly, to assess each HSM’s ability to discriminate favorable EQ-5D index, we performed receiver operating characteristic (ROC) curve analysis using the pROC package.

Favorable EQ-5D index was defined as a value greater than 0.689 (the median value for all EQ-5D index values in our cohort). For each ROC curve, we computed the area under the curve (AUC). Pairwise tests to compare the AUC for each HSM were conducted.
Results

There were 119 patients that met inclusion criteria and had complete pre- and postoperative HSM data sets. Summary statistics for preoperative and operative data are displayed in Tables 1 and 2. Average age of the included patients was 62.5 ± 12.7 years (range 31–88 years), 56.3% were male, 77.3% were white, and 60.5% were married.

Correlation of HSMs

Table 3 shows Spearman rank-order correlations among the various HSMs of interest. The PDQ total score had the highest correlation with EQ-5D index (Spearman’s rho = –0.82). The PHQ-9 scores had a moderate correlation with EQ-5D (Spearman’s rho = –0.68). The Nurick grade had weak correlations with other scales. The mJOA correlated better than the Nurick scale with all QOL measures analyzed. Neither of the myelopathy scales (mJOA or Nurick scale), however, had strong correlations between themselves (–0.41) or with the other QOL measures (absolute value range 0.13–0.49). In contrast, the QOL measures correlated relatively well with each other (absolute value range 0.68–0.97).

Responsiveness of the Outcome Measures

Table 4 displays mean changes (postoperative score – preoperative score) along with 95% confidence intervals and p values for each of the HSMs. A statistically significant improvement was seen for each scale.

Predictors of Favorable EQ-5D Value

Figure 1 displays ROC curves showing how well each of the PHQ-9, PDQ functional status domain, PDQ psychosocial domain, PDQ total score, Nurick scale, and mJOA predicts an EQ-5D index greater than 0.689 (median value). Table 5 displays p values for each of the pairwise comparisons of the AUC for this scenario. In the present study, the AUC measures how well each HSM discriminates between favorable and unfavorable EQ-5D index. An AUC of 0.5 indicates no discriminative ability, and a value of 1 indicates perfect discrimination. For example, PDQ total score had an area under the curve of 0.909. The next highest AUC was for the PDQ functional status domain (AUC = 0.902). However, these 2 AUCs were not significantly different (p = 0.38). The AUCs were significantly greater (superior ability to predict favorable outcomes) for the QOL measures as compared with the myelopathy measures, for almost all QOL scores analyzed. The mJOA had an AUC of 0.677, whereas the AUC for the Nurick scale was 0.607; this difference was not significantly different (p = 0.4).

Discussion

Quality of life outcomes data are commonly used to compare treatment strategies for various pathologies, including CSM. Rather than rely solely on clinical signs/symptoms (e.g., gait, fine motor skill, or urinary difficulties), it is becoming increasingly recognized that providers should use quantitative assessments of quality of life to better understand patient-specific concerns and subjective value system. However, with numerous questionnaires available for various pathologies, it remains unclear which measures are optimal for clinical decision making. In addition, it is difficult to interpret the implications of studies that use different QOL measures. Above all, the selected measure must be responsive, reliable, and valid. As our health care system adjusts to making decisions based on value and cost-effectiveness of interventions, providers must be mindful of disease-specific validated and reliable outcome measures with a high correlation with the EQ-5D. Such outcome measures could then be applied to the decision-making process, thus facilitating the optimizing of patient care.

The literature is sparse with regard to the comparison of outcome measures for patients with CSM. In 2001, Singh and Crockard conducted a prospective 6-month
analysis of QOL outcome measures for 100 patients who underwent surgery for CSM. The authors used Odom’s criteria, Nurick scale,16 Ranawat grade,19 Myelopathy Disability Index (MDI),3 JOA, European Myelopathy Score,27 and SF-36.4 All the measures showed statistically significant improvement over 6-month follow-up, as well as high yet variable correlation. The MDI was found to be most sensitive and reliable for detecting differences in symptom severity. In 2009, Holly et al.8 conducted a systematic review to identify valid measures of functional outcome after treatment for cervical degenerative disease. They provided Class II–III evidence in support of the MDI, JOA, and SF-36 scales as valid and reliable tools for assessing patients undergoing surgery for CSM. This evidence was based on 4 studies: 1) the aforementioned study by Singh and Crockard,22 2) a study by Yonenobu et al.30 (n = 29; 1- to 6-week follow-up) evaluating the JOA scale for CSM that found high inter- and intraobserver reliability of the scale, 3) a study by King et al.23 (n = 88; 1-year follow-up) evaluating the SF-36 scale for CSM that found high validity of the SF-36 when compared with the Nurick and JOA scales, and 4) a study by Singh et al.23 (n = 105, 6-month follow-up) that found high validity for both the SF-12 and SF-36.

Most recently, Whitmore et al.28 conducted a 1-year prospective analysis (n = 103) comparing various QOL outcome measures for patients undergoing surgery for CSM. They used Spearman rank correlations to analyze the convergent validity of their measures, which included mJOA, Oswestry Disability Index (ODI),6 Nurick scale, SF-36, and EQ-5D. They also analyzed the ability of each measure to predict EQ-5D index and used ROC analysis to compare each measure’s ability to discriminate positive from negative QOL outcomes. Correlations among all measures were highly significant with variable degrees of correlation (i.e., the lowest were Nurick and EQ-5D with Spearman’s rho of 0.5539, and the highest were ODI and EQ-5D with Spearman’s rho of 0.8306). All measures showed statistically significant improvements over the 1-year follow-up period. The ODI was the best discriminating measure among all the measures between positive and negative QOL outcomes (p = 0.02) with positive versus negative being defined as above or below the patient sample EQ-5D median index score of 0.778.

In the present study, we sought to achieve 3 goals: examine the convergent validity of QOL outcome measures for CSM, evaluate the responsiveness of each outcome measure, and assess the ability of each measure to predict positive or negative surgical outcomes via EQ-5D index scores. We measured QOL outcomes of patients diagnosed with CSM for a 1-year period using 5 different functional outcome measures: EQ-5D, mJOA, Nurick scale, PHQ-9, and PDQ (the latter including subscores for functional status, psychosocial status, and total). We found that all measures demonstrated statistically significant correlation with the EQ-5D and PDQ functional and total scores. In contrast, the Nurick scale did not show significant correlation with either the PHQ-9 or psychosocial component of the PDQ. Furthermore, even among those questionnaires with which the Nurick scale did have a statistically significant correlation, the extent of correlation was among the lowest.

Specifically, it is interesting to note the correlation of the depression index with the other measures of myelopathy and QOL. The PHQ-9 depression score correlated moderately well with the PDQ and EQ-5D scores (range 0.63–0.68), as there is a relationship between subjective disability and worse QOL among those who are most depressed. However, the mJOA and Nurick scales correlated worst with the PHQ-9 depression index (0.3 and 0.1, respectively), as compared with their correlations with the other measures of QOL (absolute value range 0.42–0.49 for mJOA, and 0.22–0.37 for Nurick scale). This suggests that the myelopathy-specific mJOA and Nurick scales assess factors that are almost completely distinct from those measured by the QOL indices.

To evaluate the ability of each measure to predict positive and negative surgical outcomes based on the EQ-5D index, an ROC analysis was performed. Using previously described methodology,26 we used our patient sample median EQ-5D index score of 0.689 as the cutoff between good and bad outcomes. The PDQ scores (including total,

### Table 2. Operative summary statistics

<table>
<thead>
<tr>
<th>Surgical Approach</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posterior cervical fusion</td>
<td>44 (45)</td>
</tr>
<tr>
<td>Anterior cervical discectomy &amp; fusion</td>
<td>46 (47)</td>
</tr>
<tr>
<td>Laminoplasty</td>
<td>5 (5)</td>
</tr>
<tr>
<td>Laminectomy/foraminotomy</td>
<td>2 (2)</td>
</tr>
<tr>
<td>No. of surgical levels</td>
<td>2.98 ± 1.2</td>
</tr>
</tbody>
</table>

* Continuous data are presented as mean ± SD. Categorical data are presented as number (%).

### Table 3. Spearman rank-order correlations among scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>EQ-5D</th>
<th>PHQ-9</th>
<th>PDQ Functional Status</th>
<th>PDQ Psychosocial</th>
<th>PDQ Total</th>
<th>Nurick Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ-5D</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>-0.68*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDQ functional status</td>
<td>-0.80*</td>
<td>0.63*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDQ psychosocial</td>
<td>-0.78*</td>
<td>0.67*</td>
<td>0.84*</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDQ total</td>
<td>-0.82*</td>
<td>0.68*</td>
<td>0.97*</td>
<td>0.94*</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Nurick grade</td>
<td>-0.28†</td>
<td>0.13</td>
<td>0.37†</td>
<td>0.22</td>
<td>0.31†</td>
<td>1.00</td>
</tr>
<tr>
<td>mJOA</td>
<td>0.42*</td>
<td>-0.30†</td>
<td>-0.49*</td>
<td>-0.43*</td>
<td>-0.48*</td>
<td>-0.41*</td>
</tr>
</tbody>
</table>

* p < 0.001.
† p < 0.05.
TABLE 4. Mean changes in score (postoperative score – preoperative score)

<table>
<thead>
<tr>
<th>Scale</th>
<th>No.*</th>
<th>Difference (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ-5D</td>
<td>108</td>
<td>0.06 (0.02 to 0.1)</td>
<td>0.0022</td>
</tr>
<tr>
<td>PHQ-9</td>
<td>57</td>
<td>-2.19 (-3.39 to -0.99)</td>
<td>0.0006</td>
</tr>
<tr>
<td>PDQ functional status</td>
<td>59</td>
<td>-11.07 (-16.5 to -5.64)</td>
<td>0.0001</td>
</tr>
<tr>
<td>PDQ psychosocial</td>
<td>59</td>
<td>-4.39 (-7.71 to -1.07)</td>
<td>0.0105</td>
</tr>
<tr>
<td>PDQ total</td>
<td>59</td>
<td>-15.46 (-23.51 to -7.41)</td>
<td>0.0004</td>
</tr>
<tr>
<td>Nurick grade</td>
<td>42</td>
<td>-0.76 (-1.12 to -0.41)</td>
<td>0.0004</td>
</tr>
<tr>
<td>mJOA</td>
<td>42</td>
<td>1.47 (0.97 to 1.96)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

* Sample size (no.) represents the number of patients from the total cohort who completed both the pre- and postoperative respective questionnaire.

functional, and psychosocial subscores) had the highest AUC, while the mJOA and Nurick scale had the lowest AUC. Accordingly, the PDQ showed significantly greater ability to predict positive versus negative EQ-5D index outcomes, compared with the mJOA and Nurick scales, as well as when compared with the PHQ-9 score. The PHQ-9 depression score, in turn, was also significantly better at predicting positive outcomes compared with the myelopathy scales.

Similar to the study published by Whitmore and colleagues, our results demonstrated that the Nurick scale correlated the least with other measures and was the worst regarding predictive ability to discriminate good outcomes. Our study provides further validation of the limited utility of the Nurick scale for this purpose. While the study by Whitmore et al. used mostly functional measures including ODI and SF-36, we included the additional measures of function/disability (PDQ) as well as psychosocial measures (i.e., the PHQ-9 and PDQ psychosocial score). Of note, the most predictive score in both our study and that of Whitmore et al. was the measure of functional disability (i.e., the PDQ in our study and the ODI in the previous study).

Overall, among the myelopathy scores, the mJOA scale was associated with a better correlation with the QOL measures and better predictive ability compared with the Nurick scale. The substantially lower correlation between myelopathy and QOL outcomes, as compared with the various QOL measures themselves, suggests that these questionnaires are measuring distinct aspects of the patient experience. Solely assessing the myelopathy or diseasespecific signs and symptoms is likely insufficient to fully understand and appreciate clinical outcome in its totality. Such, therefore, cannot be reliably used to determine op-
timal treatment options for this patient population. As we move forward toward value-based care, it is important for clinicians to use both the overall QOL outcome measures as well as the disease-specific measures (such as the mJOA scale) to best evaluate patients pre- and postoperatively.

The present study has limitations that must be acknowledged. This was a retrospective analysis with the associated inherent biases. The 1-year follow-up period allowed for the relatively large sample size, but future studies are needed to evaluate longer-term QOL outcomes. To date, none of the relevant studies on this topic have included follow-up for the QOL outcomes for greater than 1 year. Nonetheless, the present study is the first to include psychosocial measures including the PHQ-9 depression index and the PDQ psychosocial score, as well as the total PDQ measure of overall disability. Using appropriate statistical analysis, we were able to present the relevant data on the various QOL instruments and their utility regarding the evaluation of patients undergoing surgery for CSM.

Conclusions

We found that all included measures of QOL (EQ-5D, PHQ-9, and PDQ) as well as the CSM-specific (mJOA, Nurick scale) measures were valid and responsive. Neither the Nurick scale nor mJOA was associated with a good correlation with the other QOL measures. The PDQ was the most predictive of positive QOL after surgery (as measured by the EQ-5D index) for patients with CSM. Based on our data, mJOA is best used with the PDQ questionnaire to accurately capture the patient’s experience following surgery for CSM. Future studies should evaluate longer-term durability of these results.

References


TABLE 5. Probability values for pairwise comparisons of the AUCs for HSMs predicting whether or not the EQ-5D index was greater than 0.689

<table>
<thead>
<tr>
<th>Scale</th>
<th>PDQ Functional Status</th>
<th>PDQ Psychosocial</th>
<th>PDQ Total</th>
<th>Nurick Grade</th>
<th>mJOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHQ-9</td>
<td>0.0378</td>
<td>0.0940</td>
<td>0.0204</td>
<td>0.0008</td>
<td>0.0293</td>
</tr>
<tr>
<td>PDQ functional status</td>
<td>0.5038</td>
<td>0.3993</td>
<td>&lt;0.0001</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>PDQ psychosocial</td>
<td>0.1215</td>
<td>&lt;0.0001</td>
<td>0.0006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDQ total</td>
<td>&lt;0.0001</td>
<td>0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurick grade</td>
<td>0.3607</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
QOL versus myelopathy measures for CSM


Disclosures
The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

Author Contributions
Conception and design: Mroz, Lubelski. Acquisition of data: Lubelski, Nesterenko, Sundar. Analysis and interpretation of data: all authors. Drafting the article: Lubelski, Alvin, Thompson. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Statistical analysis: Lubelski, Thompson.

Supplemental Information
Current Affiliation
Dr. Lubelski: Department of Neurosurgery, Johns Hopkins Hospital, Baltimore, Maryland.

Correspondence
Thomas E. Mroz, Departments of Orthopaedic and Neurological Surgery, Cleveland Clinic Center for Spine Health, The Cleveland Clinic, 9500 Euclid Ave., S-80, Cleveland, OH 44195. email: mrozt@ccf.org.