Clustered clinical findings for diagnosis of cervical spine myelopathy

Chad Cook¹,², Christopher Brown², Robert Isaacs², Matthew Roman³, Samuel Davis², William Richardson⁴

¹Walsh University, North Canton, OH, USA, ²Department of Surgery, Duke University, Durham, NC, USA, ³Department of Physical and Occupational Therapy, Duke University, Durham, NC, USA, ⁴Department of Surgery, Duke University, Durham, NC, USA

Cervical spine myelopathy (CSM) is a clinical diagnosis made with imaging confirmation. At present, most clinical tests used to identify CSM are specific and no clusters of tests have proven more beneficial than stand alone tests in guiding treatment decision making. This study endeavored to produce a cluster of predictive clinical findings for a sample of patients using a clinical diagnosis/imaging confirmation as the reference standard for cervical spine myelopathy. Data from 249 patients with various conditions associated with cervical spine dysfunction were analyzed to determine which clinical tests and measures, when clustered together, were most diagnostic for CSM. Using multivariate regression analyses and calculations for sensitivity, specificity, and positive and negative likelihood ratios, a definitive cluster was identified. Thirteen clinical findings were investigated for capacity to diagnosis CSM. Five clinical: (1) gait deviation; (2) Hoffmann’s test; (3) inverted supinator sign; (4) Babinski test; and (5) age >45 years, were demonstrated the capacity when clustered into one of five positive tests to rule out CSM (negative likelihood ratio = 0.18; 95% CI = 0.12–0.42), and when clustered into three of five positive findings to rule in CSM (positive likelihood ratio = 30.9; 95% CI = 5.5–181.8). This study found clustered combinations of clinical findings that could rule in and rule out CSM. These clusters may be useful in identifying patients with this complex diagnosis in similar patient populations.

Keywords: Cervical spine myelopathy, Clinical prediction rule, Diagnostic accuracy, Sensitivity, Specificity

Introduction

Cervical spine myelopathy (CSM) has no single ‘pathognomonic’ sign or symptom, the onset is often insidious with long periods of episodic, stepwise progression, and may present with a vast array of clinical findings from patient to patient.¹⁻⁵ Cervical spine myelopathy is a clinical diagnosis that may involve lower extremities first (with subsequent gait related changes), weakness of the legs, and spasticity.⁶⁻⁷ As spinal cord degeneration progresses, lower motor neuron findings in the upper extremities such as loss of strength, atrophy, and difficulty in fine finger movements, may present.⁷ Additional clinical findings may include: neck stiffness, shoulder pain, paresthesia in one or both arms or hands,⁸ or radiculopathic signs.⁹ A magnetic resonance image (MRI) is considered the best imaging method for confirming the presence of cervical stenosis, cord compression, or myelomalacia, elements germane to CSM. An MRI is most useful because the tool expresses the amount of compression placed on the spinal cord, and demonstrates relatively high levels of sensitivity (79–95%) and specificity (82–88%) (LR+ = 4.39–7.92; LR− = 0.06–0.27).¹⁰

Past studies have studied the diagnostic accuracy and screening capacity of a number of active and passive clinical tests,⁸⁻¹³ including Hoffmann’s test,⁸⁻¹¹,¹⁴ deep tendon reflex testing,⁸⁻¹¹,¹⁵ inverted supinator sign,¹¹,¹⁶ Babinski sign,¹⁷ and clonus.¹⁸ Nearly all of these tests are specific, versus sensitive, and are useful to rule in a suspected condition versus ruling out the condition. In addition, the tests, when used alone, are not overtly diagnostic and may lead to a number of false negatives and in rare occasions, false positives.⁸

One method used to improve the diagnostic accuracy of clinical testing is combining tests into clusters that are more reflective of comprehensive examination findings. Clustering tests often overcome the inherent weaknesses of stand alone tests and can mimic actual clinical decision making processes by taking into account a larger assemblage of pertinent information. A past study that involved a small sample size attempted to group findings to improve
the diagnosis of myelopathy but failed to identify grouped findings that substantially modified the post-test probability of a diagnosis. One reason the study may have failed was the use of signal intensity changes on the MRI as the reference standard for CSM. In reality, the diagnosis of CSM involves MRI findings and clinical findings, with equal weighting of both results. Consequently, the purpose of this study is to produce a cluster of predictive clinical test findings for a sample of patient using a clinical diagnosis as the reference standard for CSM. Findings may improve our ability to confirm this disease earlier in stage of the condition or rule out the condition during screening.

Materials and Methods

Study guidelines

Procedural guidelines for this study followed the STARD standards for reporting of diagnostic accuracy set forth by Bossuyt et al. Briefly, the STARD standards are used to improve reporting processes for diagnostic accuracy studies and involve 25 items associated with topics germane in a typical case control design. Topics are oriented toward description of participant, statistical analysis, results, and conclusions of findings.

Participants

The study protocol was approved by the Institutional Review Board of Duke University Health System. The study included 249 consecutive patients seen at a spine surgery center at Duke University from 2006 to 2009. All patients with cervical pain or dysfunction were eligible for the study if a clinical diagnosis was made and if imaging (MRI) was performed or available and if data were prospectively recorded in accordance to the standardized screening examination performed at the facility. Patients were seen for a variety of signs and symptoms but in all cases, cervical spine pain or dysfunction was the patient’s primary complaint.

Standardized screening examination

All patients received a standardized screening examination that consisted of self-report and physical examination methods. The screening process was facilitated by a number of orthopedic surgeons or a physician-extender (physical therapist or physician assistant) in accordance to the policy of the orthopedic clinic. All screening examination findings were input into a structured Excel compatible database that allows real-time Internet-based interface. The targeted variables for the study included descriptive and predictive variables.

Descriptive variables

Descriptive variables include the self-report findings of age, gender, race, marital status, employment status, exercise status, educational status, workman’s compensation status, use of physical therapy for current problem, duration of current complaints, neck disability index (NDI, scored as percentage of disability), SF12, pain score at its highest typical amount, and the physical finding of body mass index (BMI) calculated from height and weight.

The SF12 is a generic measure and does not target a specific age or disease group. The SF12 is a shortened version of the SF36 and is weighted and summed to provide an interpretable measure of quality of life. The NDI was developed in 1989 as a modification of the Oswestry low back pain disability index. The NDI is a frequently used instrument for measuring self-rated disability due to neck pain. The scale consists of 10 items and is scored from 0 to 5; the maximum score is 50 with higher scores indicating higher disability. For this study, the obtained score was multiplied by 2 to produce a percentage score. Pain at its highest point was calculated using a numeric analog scale. Patients were asked to report their current point at its highest level on an eleven point scale from 0 to 10. A score of 0 was associated with no pain, whereas a score of 10 was associated with maximum pain.

Predictive variables

Predictive variables were targeted after consultation of published literature. Predictive variables included age categorization after receiver operator characteristics (ROC analysis), the tests of Spurling, Hoffmann, and Babinski, the distraction test, presence of Clonus, presence of a Gait abnormality (abnormally wide based gait, ataxia, or spastic gait), presence of hyper-reflexia in the biceps, the inverted supinator sign, quadriceps, or Achilles, pain score (after ROC analysis), and pain constancy score.

Diagnosis of cervical spine myelopathy

For all participants, a diagnosis of CSM was made by orthopedic surgeons after careful consideration of presenting symptoms in the patient’s history (e.g. neck stiffness, dexterity loss, unilateral or bilateral deep, aching neck, arm and shoulder pain, gait dysfunction, and stiffness and clumsiness) and physical examination (e.g. multisegmental weakness, losses during coordination testing, and variable losses of sensation and proprioception). When suspected after pertinent clinical examination findings, a diagnosis of myelopathy was confirmed or denied by imaging methods (MRI) in all 249 cases. Anterior–posterior width reduction, cross-sectional evidence of cord compression, obliteration of the subarachnoid space, and signal intensity changes to the cord found on MR imaging are considered the most appropriate parameters for confirmation of a spinal cord compression myelopathy. Patients who met the
criteria for CSM were coded appropriately (‘yes’ or ‘no’) and input into the database.

Data analysis
Data downloaded from the Internet-based Excel dataset were transferred to SPSS 13.0 for Windows (SPSS Inc., Chicago, IL, USA). Descriptive statistics were captured and sequestered into groups of patients with and without a diagnosis of CSM. Bivariate analyses were carried out between patients with and without a clinical diagnosis of CSM. A P value of \(<0.05\) was considered significant.

All 13 predictor variables were individually examined for diagnostic accuracy. Contingency tables \((2 \times 2)\) were used to calculate sensitivity and specificity, and likelihood ratios (positive likelihood ratio = LR\(+\); negative likelihood ratio = LR\(–\)) for each predictive test item. ROC curves were used to determine all possible cutoff values for age and highest reported pain score.

A backward stepwise binary logistic regression model was used to determine a clinical prediction rule for the diagnosis of CSM.\(^\text{26}\) The conditionally independent variables from the individual \(2 \times 2\) analyses that resulted in LR\(+\) values of \(>2.0\) and/or LR\(–\) values \(<0.5\) were entered into the model. A backward stepwise selection procedure was used to select variables, with \(P\) values of 0.15 to exit the model and 0.10 to enter it. Variables retained by the regression model were used to develop a clinical prediction rule and were then input into \(2 \times 2\) contingency tables that involved the conditions of one of five, two of five, three of five, four of three, and five of five positive findings. For each condition, sensitivity, specificity, and likelihood ratios were analyzed. In addition, in each condition, post-test probability measures were calculated using a pre-test probability of 35\% (the prevalence of CSM in this sample).

Most diagnostic accuracy studies are powered on tabulated probabilities of values being lower than acceptable confidence interval estimates (lower levels) for sensitivity values.\(^\text{27}\) whereas studies involving clustered findings often drives up specificity values at the expense of sensitivity values. Consequently, the study was powered using the regression values and the 13 predictor variables. Using Monte Carlo simulations, Peduzzi et al.\(^\text{28}\) reported that an \(n\) of 10–20 per predictor is appropriate for a boundary level per variable for regression analyses. This finding suggests that 130–260 patients would provide adequate values for the regression analysis, whereas larger numbers may be necessary to further reduce confidence intervals for the diagnostic accuracy statistics.

Results
Bivariate analyses found significant differences in age (\(P=0.05\)), duration of symptoms (\(P<0.01\)), SF12 scores (\(P=0.01\)), and BMI (\(P=0.05\)) (Table 1). Patients with CSM were typically older, had a longer duration of symptoms, had lower SF12 scores, and had higher BMI scores.

Individual predictor variable scores are reported in Table 2. There were 10 variables that met the criteria of LR\(+\) greater than 2 and/or LR\(–\) less than 0.50. The inverted supinator sign demonstrated the single highest LR\(+\) (29.1; 95\% CI = 5.1–171.5) followed by hyperreflexia of the Achilles (7.8; 95\% CI = 2.5–25.4) and the quadriceps (6.9; 95\% CI = 2.8–17.5). Only age \(>45\) years demonstrated a LR\(–\) of less than 0.50 (0.48; 95\% CI = 0.26–0.85).

All 249 patient values were included in the regression analysis. Of the 10 variables included in the regression modeling, the tests of Babinski’s and Hoffmann’s signs, the inverted supinator sign, gait abnormality, and age \(>45\) years were retained. The Hosmer–Lemeshow test indicated that the model fit the data (\(P=0.59\)), and the Nagelkerke \(R^2\) equaled 0.36.

The five test variables and their diagnostic properties according to the number of abnormalities required for a positive test are listed in Table 3. Diagnostically, one of five positive tests resulted in the strong screening combination yielding a sensitivity of 0.94 and a negative likelihood ratio of 0.18. A finding that included three of five positive tests yielded a positive likelihood ratio of 30.9 (95\% CI = 5.5–181.8) and a post-test probability of 94\%. There were only eight instances in which four of five tests were positive and no instances in which all five of five tests were positive.

Discussion
This study endeavored to produce identify a useful cluster of clinical tests that was indicative of a clinical diagnosis (with imaging confirmation) of cervical spine myelopathy. The findings are unique as this is the first that has captured a cluster of findings that not only function as a screening tool, useful in ruling out the condition of myelopathy, but also provide combinations that are confirmatory, ruling in conditions of myelopathy. In addition, the use of a clinical diagnosis as the reference standard (with confirmation with an MRI) is more consistent with an acceptable diagnosis of CSM and improves upon previous works that used imaging findings only.\(^\text{11}\)

Many healthcare clinicians look for negative findings during testing of Hoffmann’s sign, Babinski’s sign, Clonus, and hyper-reflexia to rule out myelopathy. As identified by others,\(^\text{8,12,13}\) these tests demonstrate low sensitivity and are not appropriate for ruling out myelopathy. Although previous authors have suggested the use of clusters of test findings to improve the sensitivity of the examination,\(^\text{8}\)
to this point, no study has demonstrated a sensitive combination of clusters. In this study, any positive one of five tests yielded a specificity of 0.94 (0.89–0.97) and a negative likelihood ratio of 0.18 (0.12–0.42). Values within this range provide moderate values in one’s ability to rule out a condition and have been identified as useful during clinical decision making. This suggests that clinicians who identify only ≤1 of 5 positive test findings should be confident that the patients do not have CSM.

All subjects in this study received MR imaging to confirm the presence of the suspicion of CSM. MRI is

<table>
<thead>
<tr>
<th>Test item</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Positive likelihood ratio (95% CI)</th>
<th>Negative likelihood ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &gt;45 years</td>
<td>0.86 (0.81–0.93)</td>
<td>0.26 (0.23–0.29)</td>
<td>1.2 (1.0–1.3)</td>
<td>0.48 (0.26–0.85)</td>
</tr>
<tr>
<td>Spurling’s test</td>
<td>0.15 (0.09–0.19)</td>
<td>0.94 (0.92–0.97)</td>
<td>2.6 (1.2–5.8)</td>
<td>0.90 (0.84–0.98)</td>
</tr>
<tr>
<td>Distraction test</td>
<td>0.40 (0.02–0.07)</td>
<td>0.97 (0.96–0.99)</td>
<td>1.8 (0.51–6.5)</td>
<td>0.97 (0.94–1.0)</td>
</tr>
<tr>
<td>Hoffmann’s test</td>
<td>0.31 (0.25–0.35)</td>
<td>0.73 (0.59–0.84)</td>
<td>4.9 (2.6–9.6)</td>
<td>0.74 (0.67–0.83)</td>
</tr>
<tr>
<td>Clonus</td>
<td>0.07 (0.04–0.08)</td>
<td>0.99 (0.97–0.99)</td>
<td>5.4 (1.2–23.4)</td>
<td>0.94 (0.99–0.99)</td>
</tr>
<tr>
<td>Babinski test</td>
<td>0.07 (0.42–0.68)</td>
<td>1.0 (0.98–1.0)</td>
<td>Inf (2.9–Inf)</td>
<td>0.93 (0.93–0.97)</td>
</tr>
<tr>
<td>Gait deviation</td>
<td>0.19 (0.14–0.24)</td>
<td>0.94 (0.91–0.97)</td>
<td>3.4 (1.6–7.3)</td>
<td>0.85 (0.78–0.94)</td>
</tr>
<tr>
<td>Hyper-reflexia biceps</td>
<td>0.18 (0.13–0.22)</td>
<td>0.96 (0.93–0.98)</td>
<td>4.8 (2.0–11.7)</td>
<td>0.85 (0.79–0.93)</td>
</tr>
<tr>
<td>Inverted supinator sign</td>
<td>0.18 (0.14–0.19)</td>
<td>0.99 (0.97–0.99)</td>
<td>29.1 (5.1–171.5)</td>
<td>0.82 (0.81–0.84)</td>
</tr>
<tr>
<td>Hyper-reflexia quadriceps</td>
<td>0.22 (0.17–0.25)</td>
<td>0.97 (0.94–0.99)</td>
<td>6.9 (2.8–17.5)</td>
<td>0.81 (0.76–0.89)</td>
</tr>
<tr>
<td>Hyper-reflexia Achilles</td>
<td>0.15 (0.11–0.17)</td>
<td>0.98 (0.96–0.99)</td>
<td>7.8 (2.5–25.4)</td>
<td>0.87 (0.84–0.93)</td>
</tr>
<tr>
<td>Pain score</td>
<td>0.36 (0.28–0.45)</td>
<td>0.52 (0.48–0.57)</td>
<td>0.8 (0.52–1.1)</td>
<td>1.2 (0.95–1.5)</td>
</tr>
<tr>
<td>Pain constancy</td>
<td>0.62 (0.53–0.70)</td>
<td>0.40 (0.35–0.45)</td>
<td>1.03 (0.82–1.3)</td>
<td>0.95 (0.67–1.3)</td>
</tr>
</tbody>
</table>

Note: Useful likelihood ratios appear in bold.
Table 3 Clustered findings for diagnosis of cervical spine myelopathy

<table>
<thead>
<tr>
<th>Clustered results</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Positive likelihood ratio (95% CI)</th>
<th>Negative likelihood ratio (95% CI)</th>
<th>Post-test probability of CTS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 of 5 positive tests</td>
<td>0.94 (0.89-0.97)</td>
<td>0.31 (0.27-0.32)</td>
<td>1.4 (1.2-1.4)</td>
<td>0.18 (0.12-0.42)</td>
<td>43</td>
</tr>
<tr>
<td>2 of 5 positive tests</td>
<td>0.39 (0.33-0.46)</td>
<td>0.88 (0.84-0.92)</td>
<td>3.3 (2.1-5.5)</td>
<td>0.63 (0.59-0.79)</td>
<td>64</td>
</tr>
<tr>
<td>3 of 5 positive tests</td>
<td>0.19 (0.15-0.20)</td>
<td>0.99 (0.97-0.99)</td>
<td>30.9 (5.5-181.8)</td>
<td>0.81 (0.73-0.87)</td>
<td>94</td>
</tr>
<tr>
<td>4 of 5 positive tests</td>
<td>0.09 (0.06-0.09)</td>
<td>1.0 (0.96-1.0)</td>
<td>Inf (3.9-Inf)</td>
<td>0.91 (0.90-0.95)</td>
<td>99+</td>
</tr>
</tbody>
</table>

Note: Five tests are included in the rule: (1) gait deviation; (2) + Hoffmann’s test; (3) inverted supinator sign; (4) + Babinski test; and (5) age >45 years. The associated post-test probability values are based on a pre-test probability of 35%.

considered the best confirmatory imaging method for CSM and has yielded acceptable sensitivity and specificity values, with notable capacities in identifying selected abnormalities such as space occupying tumors, disc herniation, and ligamentous ossification. The MRI provides the ability to rule out a tumor or syrinx (fluid-filled cavity that develops in the spinal cord), and provides detailed views of the spinal cord, intervertebral disc, vertebral osteophytes, and ligaments, all structures that potentially compress the spinal cord. Furthermore, MRI findings have been shown to correlate with pre-operative severity of cervical compressive myelopathy and prognosis after surgery. Nonetheless, exclusive use of an MRI for diagnosis of CSM is unwarranted and inappropriate. At present, with the exception of myelomalacia (a chronic and specific condition identified through signal intensity changes to the cord), there are no definitive objective findings on MRI consistently described by radiologists that are reflective of myelopathy.

One possible reason why past studies have failed to outline clusters of findings is the concept of conditional dependence. Conditional dependence occurs when a subsequent test finding is not dissimilar to the first test finding or when a series of tests actually measure the same thing and are positive together in clusters or negative together in clusters. When this phenomenon occurs, additional test results yield no further value and the finding of one test has the same diagnostic accuracy as the findings of four conditionally dependent tests. In our study, the hyperreflexia measures captured at the biceps, quadriceps, and Achilles, and the inverted supinator sign were always positive in clusters. Because the inverted supinator sign finding was the most accurate, this measure alone was left in the stepwise regression.

One of the clustered findings in our study was assessment of abnormal gait. Abnormal gait was visually assessed as any incidence of abnormally wide based gait, ataxia, or spastic gait. These descriptors were selected because they are common identifiers associated with upper motor neuron changes involving corticospinal tracts and spinocerebellar tracts dysfunction; problems notable in CSM. Past studies have identified selected quantifiable gait variations using tools such as the 10-second step test, the tandem walk test, or the parallel walk test. Although gait problems are the first symptoms associated with myelopathy, our study found low sensitivity values for this assessment (0.19; 95% CI: 0.14–0.24). Furthermore, specific gait-related abnormalities associated with CSM are poorly defined in the literature. A more quantifiable method of gait assessment may have improved the diagnostic accuracy of our clustered model.

Limitations

This study has a number of notable limitations. The first limitation is incorporation bias, which is present in any situation in which the diagnostic standard is a clinical diagnosis. Consequently, because in many cases (not all) the same clinicians who made the diagnosis were also those that analyzed the 13 predictor tests, the chance that the tests influenced the outcome of the diagnosis is present. Additionally, the variables for the clustered analysis were examined retrospectively and retrospective analysis has been identified as a potential for bias by the STARD initiative. Nonetheless, two recent meta-analyses that examined potential biasing factors for diagnostic accuracy studies both indicated minimal bias occurs from retrospective data, one indicating no difference than data that were captured prospectively and analyzed as such. The reliability of the visual assessment of gait was not examined in this study and deserves further exploration. Lastly, the patients in the study were attendees of clinical appointments for orthopedic surgeons and involved patients with a wide degree of neck problems, often quite severe. Consequently, this form of spectrum bias may exaggerate the sensitivity and specificity values and could overestimate the accuracy of this cluster in a population of subjects dissimilar to our sample.

Conclusion

This study found that selected combinations of clinical findings that consisted of (1) gait deviation; (2) + Hoffmann’s test; (3) inverted supinator sign; (4) + Babinski test; and (5) age >45 years were affective in ruling out and ruling in cervical spine myelopathy. Combinations of three of five or four of five tests enabled adjustments of post-test probability of the
condition to 94.99%. These clusters may be useful in identifying patients with this complex diagnosis in similar patient populations.

References


