Psychometric analysis and critical appraisal of the original, revised, and modified versions of the Japanese Orthopaedic Association score in the assessment of patients with cervical spondylotic myelopathy

Julio C. Furlan, MD, LLB, MBA, MSc, PhD, FRCPC,1–3 and B. Catharine Craven, BA, MD, MSc, FRCPC2,3

1Division of Neurology; 2Division of Physical Medicine and Rehabilitation, Department of Medicine, University of Toronto; and 3Lyndhurst Centre, Toronto Rehabilitation Institute, University Health Network, Toronto, Ontario, Canada

OBJECTIVE Cervical spondylotic myelopathy (CSM) is the most common cause of nontraumatic spinal cord impairment and disability in the world. Given that the Japanese Orthopaedic Association (JOA) score is the most frequently used outcome measure in clinical research and practice for treating patients with CSM, this review was undertaken to comprehensively and critically evaluate the psychometric properties of the JOA score.

METHODS The authors identified studies (published in the period of January 1975 to November 2015) on the psychometric properties of the original, revised, and modified versions of the JOA score in Medline, PsycINFO, Excerpta Medica dataBASE (EMBASE), American College of Physicians Journal Club, and Cochrane Database of Systematic Reviews. Additional publications were captured in a secondary search of the bibliographies in both original research articles and literature reviews identified in the original search. The JOA scores were evaluated for item generation and reduction, internal consistency, reliability, validity, and responsiveness. This review included all those versions of the JOA score whose psychometric properties had been reported in at least 2 published studies.

RESULTS The primary search strategy identified 59 studies, of which 9 fulfilled the inclusion and exclusion criteria. An additional 18 publications were captured in the secondary search and included in the analysis. The key findings from the 27 studies analyzed indicated the following: 1) the original JOA score (1975) was the source for the revised JOA score (1994) and 3 modified versions (1991, 1993, and 1999 JOA scores) reported or used in at least 2 published studies; 2) the revised and modified versions of the JOA score are markedly different from each other; 3) only the revised JOA score (1994) was validated with the original JOA score; and 4) the 1975 JOA score is the most appropriate instrument for assessing patients in Asian populations (especially from Japan) because of its psychometric attributes, and the 1991 JOA score is the most appropriate version for use in Western populations.

CONCLUSIONS The authors’ results indicate that the original (1975), revised (1994), and modified (1991, 1993, and 1999) versions of the JOA score are substantially different from each other in terms of their content and have been incompletely examined for their psychometric properties and cultural sensitivity. Whereas the 1975 JOA score is the most appropriate version for assessing individuals from Asian populations (particularly those eating with chopsticks), the 1991 JOA score is most suitable for evaluating patients in Western populations. Nonetheless, further investigation of the psychometric properties of the 1975 and 1991 JOA scores is recommended because of a paucity of studies reporting on the responsiveness of these 2 scoring instruments.

http://thejns.org/doi/abs/10.3171/2016.3.FOCUS1648

KEY WORDS Japanese Orthopaedic Association score; cervical spondylotic myelopathy; psychometric properties; reliability; validity; responsiveness; spinal disorders

ABBREVIATIONS ADL = activities of daily living; CSM = cervical spondylotic myelopathy; DTI = diffusion tensor imaging; ICC = intraclass correlation coefficient; JOA = Japanese Orthopaedic Association; MCID = minimum clinically important difference; MCS = Mental Component Summary; MDI = Myelopathy Disability Index; NDI = Neck Disability Index; PCS = Physical Component Summary; QuickDASH = Quick Disabilities of the Arm, Shoulder, and Hand; ROC = receiver operating characteristic; SSEP = somatosensory evoked potential.

SUBMITTED January 30, 2016. Accepted March 7, 2016.

INCLUDE WHEN CITING DOI: 10.3171/2016.3.FOCUS1648.
CERVICAL spondylotic myelopathy (CSM) is the most common cause of nontraumatic spinal cord impairment, having a significant worldwide health, social, and economic impact, but also a paucity of accurate incidence and prevalence data. According to a recent Dutch study, 1.6 per 100,000 inhabitants per annum required surgical treatment for symptomatic CSM. Although the precise economic impact of the treatment of patients with CSM remains unknown, the burden of this disorder likely represents a significant portion of the overall hospital costs for spine surgeries (excluding thoracic surgeries and spinal trauma), with an estimated total cost of $150 million US in Canada in 2012 (https://www.cihi.ca/en/spending-and-health-workforce).

In 1972, Nurick described CSM as “a generally benign disorder caused by compression of the cord in which disability develops in an initial phase of duration, the degree of incapacity reached being related to the severity of the compression. The disability later remains static except in older patients, in whom it may progress.” More recently, Matz et al. systematically reviewed the literature on the natural history of CSM. The results of their study indicated mixed natural history trajectories for individuals living with CSM, including slow manifestation, stepwise decline, or a long period of quiescence. Long periods of severe stenosis are associated with demyelination and may result in necrosis of both gray and white matter. Severe or long-lasting symptoms of CSM suggest a low probability of clinical improvement with nonoperative measures.

The presence of abnormal electrophysiological findings or clinical radiculopathy is associated with the development of symptomatic CSM in patients with cervical stenosis without myelopathy. These disease characteristics illustrate the challenges of evaluating outcomes in individuals with CSM. Interestingly, the Japanese Orthopaedic Association (JOA) score and Nurick grade were the primary outcome measures in those studies and became the most widely used outcome measures in both research studies and clinical practice on CSM.

The JOA score includes aspects of motor, sensory, and autonomic functions (e.g., “unable to move hands,” “complete loss of hand sensation,” and “unable to urinate voluntarily”) as well as features of activities of daily living (ADL) (e.g., “unable to button shirt” and “able to walk on flat floor with a walking aid”). As defined by the International Classification of Functioning, Disability, and Health of the World Health Organization (WHO), neurological impairment is related to the level of “body function and structures,” and difficulty to execute a task or action is related to limitations of activities.

For individuals with spinal cord disease, the neurological impairments as conceptualized by the WHO combine the domains of motor, sensory, and autonomic functions. A change in motor function herein was defined as a deterioration or recovery (partial or complete) of intentional movements from muscles caudal to the level of injury. A change in sensory function was defined in terms of extent and degree of somatic sensibility. A change in autonomic function was characterized by the degree of micturition dysfunction from normal voiding to urinary retention. According to the WHO concepts, activity limitations related to spinal cord dysfunction also include any difficulties an individual may have in executing tasks or actions with their upper or lower limbs.

The JOA score was created in 1975 and originally developed for the global assessment of body motor and sensory functions, urinary autonomic function, and ADL. This self-administered, disease-specific instrument contains 4 domains: upper-limb function (scale range 0–4), lower-limb function (scale range 0–4), sensory function (upper and lower limbs and trunk; scale range 0–2 at each segment), and bladder function (scale range 0–3). Overall, the original JOA score (called the “1975 JOA score” hereafter) has a range from 0 (worst) to 17 (normal). The JOA score was later revised (called “revised JOA score” hereafter) and adapted for the Western world in different versions (called “modified JOA scores” hereafter). According to the framework of Kirshner and Guyatt and for descriptive and evaluative purposes, this outcome measure has become the instrument of choice in the clinical practice and research areas of CSM.

The fact that the JOA score is one of the most commonly used outcome measures in the clinical practice and research of CSM underscores its good face validity. Nonetheless, the generic term “JOA score” has been in distinctively used in the literature when, in fact, it refers to different versions derived from the original JOA score established in 1975. Moreover, the real differences among each version of the JOA score in terms of their content and psychometric properties could have significant clinical and research implications but have not been properly studied. This systematic review was undertaken to comprehensively examine and critically appraise all studies that reported the psychometric properties (i.e., reliability, validity, and responsiveness) of the 1975 JOA score and its derived versions for use in the CSM population.

Methods

Literature Search Strategy

The primary literature search was carried out in Medline, Excerpta Medica dataBASE (EMBASE), PsycINFO, Cochrane Database of Systematic Reviews, and American College of Physicians Journal Club to identify studies reporting on the psychometric properties of the JOA score and published from January 1975 to November 2015. The search criteria included subject headings (“JOA score” OR “Japanese Orthopaedic Association score”) AND (“item generation” OR “item reduction” OR “reliability” OR “validity” OR “responsiveness” OR “psychometric”), limited to “humans.” Additional articles were captured in a secondary search using the reference lists from original articles and previous reviews that were captured in the primary search as well as from the articles subsequently captured until the references were thoroughly reviewed. Abstracts and, where necessary, actual articles were screened to identify those that have studied item generation, item reduction, reliability, validity, or responsiveness of any version of the JOA score applied to individuals with CSM. Only versions of the JOA score that were reported in at least 2 publications were included in the analysis. Case reports, case series, animal studies, and editorials were
excluded from this review. Given the paucity of studies describing some of the versions of the JOA score, no restriction for inclusion based on study quality was applied in this review.

**Versions of the JOA Score**

This review included publications on the psychometric properties on the original JOA score (1975), its revised version (1994), and its modified versions (1991, 1993, and 1999). The items of each version of the JOA score are described and delineated in Table 1.

**Measurement Properties and Definitions**

The JOA score was examined for item generation, item reduction, reliability validity, and responsiveness. Item generation is the method used to choose the potential items for a measurement instrument. Item reduction follows an extensive selection of items that first samples the full range, and redundant and less relevant items are then eliminated.

Reliability reflects the degree of consistency or agreement across repeated assessments of different patients by the same rater (intrarater reliability) or across different raters for the same patient (intrarater reliability). Reliability studies that used the intraclass correlation coefficient (ICC) were interpreted with the criteria of Nunnally and Bernstein, where an ICC ≥ 0.7 is aimed for when groups of patients are compared, and an ICC ≥ 0.9 is recommended when individuals are compared. Reliability studies that used the kappa agreement coefficient were interpreted according to the original classification of Landis and Koch as follows: poor (kappa statistic < 0), slight (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (≥ 0.81). Many studies on psychometric properties have used Cronbach’s α coefficient to evaluate reliability, even though this coefficient was actually developed to assess internal consistency of an outcome measure. Of note, internal consistency describes the interrelatedness of the items within an outcome measure. Cronbach’s α coefficient varies from 0 (no internal consistency) to 1 (perfect internal consistency) with an acceptable value of 0.7 or greater for research purposes and of at least 0.9 for clinical purposes.

Responsiveness refers to the ability of a measurement instrument to accurately detect changes. An observed change is the variation in motor, sensory, or autonomic function that is measured by gathering data at 2 different time points. A minimum change that is potentially detectable is the smallest amount of change that the JOA score is capable of measuring at an individual level, and a minimal detectable change is defined by the error associated with the measurement. Of note, the minimal detectable difference could be estimated with the reliability coefficient and standard deviation of difference scores between test and retest. A minimum clinically important difference (MCID) indicates an observed change in the JOA score in the CSM population considered to represent a clinically important neurological improvement or deterioration.

Validity reflects the degree to which the instrument measures what it is supposed to measure. Criterion validity represents the degree to which the instrument measurement complies with the criterion or gold standard. Because of the lack of a gold standard for assessing motor, sensory, and autonomic functions in the CSM population, criterion validity was excluded from this review. Construct validity is generally classified into convergent or divergent categories. Convergent construct validity describes the degree of similarity between 2 constructs that theoretically should be related to each other. Divergent construct validity indicates how dissimilar 2 constructs are when, in theory, they actually should not be related to each other.

**Results**

**Literature Search**

The primary literature search identified 59 studies, of which 9 fulfilled the inclusion criteria (Fig. 1). In addition, 18 studies were retrieved in the secondary search of the lists of references in the studies that were captured in the initial search. This review therefore included in total 27 studies.

**Item Generation and Reduction: Development, Revision, and Modifications of the JOA Score**

The development of the original JOA score in 1975 was a landmark for the assessment of impairment and disability among patients with CSM. Although the exact process of item generation and reduction of the original JOA score is not available or was lost during English language translation of the original study, according to personal communications with Dr. Kawakami Mamoru, its items apparently were selected by orthopedic spine specialists including Professor Ono (Osaka University), Dr. Hirabayashi (Key University), and Professor Tsuyama (Tokyo University).

In 1994, after incorporation of constructive critiques into the JOA score, a revised version was published. This revised version refined the sensory and autonomic function scoring and enhanced the motor function scoring, including the assessment of elbow and shoulder function through testing of manual muscle function (Table 1). Of note, the JOA revised the assessment tools for CSM in the 2000s and developed a new instrument, the JOA Cervical Myelopathy Evaluation Questionnaire (JOACMEQ). Although the subcommittee on the clinical outcome of the JOA stated that the JOACMEQ is “more scientific from the standpoint of medical statistics and more patient-oriented,” the JOA score is still more often cited in the CSM literature.

In addition, several different investigators have modified the JOA score to make it more compatible with the ADL of individuals residing in Western countries. Although these modified JOA score versions are commonly designated as the “modified JOA scores” (Fig. 2), there are notable differences among them (Table 1). For instance, the descriptions of the functions tested and their scoring vary among the modified JOA score versions. Given this, the psychometric properties could vary depending on the version of the JOA score used; hence, we analyzed each version separately hereafter.

**The Original JOA Score (1975): Psychometric Analysis**

The psychometric properties of the 1975 JOA score...
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arms</td>
<td>Fingers</td>
<td>Upper extremities</td>
<td>Upper extremities</td>
<td>Upper extremities</td>
<td></td>
</tr>
<tr>
<td>0: Unable to feed oneself w/chopsticks or spoon</td>
<td>0: Unable to feed oneself w/any table ware, including chopsticks, spoon, or fork &amp;/or unable to fasten buttons of any size</td>
<td>0: Unable to move hands</td>
<td>1: Paralysis</td>
<td>0: Unable to feed oneself</td>
<td></td>
</tr>
<tr>
<td>1: Able to feed oneself w/spoon but not w/chopsticks</td>
<td>1: Can manage to feed oneself w/a spoon &amp;/or fork but not w/chopsticks</td>
<td>1: Unable to eat w/ a spoon, but able to move hands</td>
<td>2: Fine motor function massively decreased</td>
<td>1: Unable to use knife &amp; fork; able to eat w/ a spoon</td>
<td></td>
</tr>
<tr>
<td>2: Able to use chopsticks, but inadequately</td>
<td>2: Either chopstick feeding or writing is possible but not practical &amp;/or large buttons can be fastened</td>
<td>2: Unable to button shirt, but able to eat w/a spoon</td>
<td>3: Fine motor function decelerated</td>
<td>2: Able to use knife &amp; fork w/much difficulty</td>
<td></td>
</tr>
<tr>
<td>3: Slightly clumsy in using chopsticks</td>
<td>3: Either chopstick feeding or writing is clumsy but practical &amp;/or cuff buttons can be fastened</td>
<td>3: Able to button shirt w/great difficulty</td>
<td>4: Discrete weakness in hands or proximal arm</td>
<td>3: Able to use knife &amp; fork w/slight difficulty</td>
<td></td>
</tr>
<tr>
<td>4: Normal</td>
<td>4: Normal</td>
<td>4: Able to button shirt w/slight difficulty</td>
<td>5: Normal function</td>
<td>4: No motor dysfunction</td>
<td></td>
</tr>
<tr>
<td>5: No dysfunction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoulder &amp; elbow (evaluated by MMT score of the deltoide or biceps, whichever is weaker):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−2: MMT ≤2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−1: MMT = 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−0.5: MMT = 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0: MMT = 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legs</td>
<td>Lower extremities</td>
<td>Lower extremities</td>
<td>Lower extremities</td>
<td>Lower extremities</td>
<td></td>
</tr>
<tr>
<td>0: Unable to walk by any means</td>
<td>0: Unable to stand up &amp; walk by any means</td>
<td>0: Complete loss of motor &amp; sensory function</td>
<td>1: Unable to walk</td>
<td>0: Unable to walk</td>
<td></td>
</tr>
<tr>
<td>1: Unable to walk w/o a cane or other support on the level</td>
<td>0.5: Able to stand up but unable to walk</td>
<td>1: Sensory preservation w/o ability to move legs</td>
<td>2: Need walking aid on flat floor</td>
<td>1: Can walk on flat floor w/walking aid</td>
<td></td>
</tr>
<tr>
<td>2: Able to walk independently on the level but needs support on stairs</td>
<td>1: Unable to walk w/o a cane or other support on a level</td>
<td>2: Able to move legs but unable to walk</td>
<td>3: Need handrail on stairs</td>
<td>2: Can walk up or down stairs w/handrail</td>
<td></td>
</tr>
<tr>
<td>3: Slightly clumsy in walking</td>
<td>1.5: Able to walk w/o support but w/ a clumsy gait</td>
<td>3: Able to walk on flat floor w/a walking aid (i.e., w/cane/crutch)</td>
<td>4: Able to walk w/o walking aid, but inadequate</td>
<td>3: Lack of stability &amp; smooth gait</td>
<td></td>
</tr>
<tr>
<td>4: Normal</td>
<td>2: Walks independently on a level but needs support on stairs</td>
<td>4: Able to walk up or down stairs w/handrail</td>
<td>5: Normal function</td>
<td>4: No motor dysfunction</td>
<td></td>
</tr>
</tbody>
</table>

CONTINUED ON PAGE 5
### TABLE 1. Comparison of the different versions of the JOA score for assessing patients with CSM

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor function (continued)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5: Able to walk independently when going upstairs, but needs support when going downstairs</td>
<td></td>
<td></td>
<td>5: Moderate to significant lack of stability, but able to walk up or down stairs w/o handrail</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3: Capable of fast but clumsy walking</td>
<td></td>
<td></td>
<td>6: Mild lack of stability but walks w/ smooth reciprocation unaided</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4: Normal</td>
<td></td>
<td></td>
<td>7: No dysfunction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensory function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0: Definitely impaired</td>
<td>0: Complete loss of touch &amp; pain sensation</td>
<td>0: Complete loss of hand sensation</td>
<td>1: Apparent sensory loss</td>
<td>0: Severe sensory loss or pain</td>
<td></td>
</tr>
<tr>
<td>1: Slightly impaired or subjectively numb</td>
<td>0.5: ≤50% normal sensation &amp;/or severe pain or numbness</td>
<td>1: Severe sensory loss or pain</td>
<td>2: Minimal sensory loss</td>
<td>1: Mild sensory loss</td>
<td></td>
</tr>
<tr>
<td>2: Normal</td>
<td>1: &gt;60% normal sensation &amp;/or moderate pain or numbness</td>
<td>2: Mild sensory loss</td>
<td>3: Normal function</td>
<td>2: No sensory deficit</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5: Subjective numbness of slight degree w/o any objective sensory deficit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2: Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trunk (same grading as above)</td>
<td>Trunk (same grading as above)</td>
<td>Trunk (not included)</td>
<td>Trunk (same grading as above)</td>
<td>Trunk (same grading as above)</td>
<td></td>
</tr>
<tr>
<td>Legs (same grading as above)</td>
<td>Lower extremities (same grading as above)</td>
<td>Lower extremities (partial description in the domain on motor function)</td>
<td>Lower extremities (same grading as above)</td>
<td>Lower extremities (same grading as above)</td>
<td></td>
</tr>
<tr>
<td>Bladder function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0: Incontinent</td>
<td>0: Urinary retention &amp;/or incontinence</td>
<td>0: Unable to micturate voluntarily</td>
<td>1: Urinary retention</td>
<td>0: Unable to void</td>
<td></td>
</tr>
<tr>
<td>1: Great difficulty</td>
<td>1: Sense of retention &amp;/or dribbling &amp;/or thin stream &amp;/or incomplete continence</td>
<td>1: Marked difficulty w/ micturition</td>
<td>2: Severe dysfunction</td>
<td>1: Marked difficulty in micturition (retention)</td>
<td></td>
</tr>
<tr>
<td>2: Slight difficulty</td>
<td>2: Urinary retardation &amp;/or pollakiuria</td>
<td>2: Mild-to-moderate difficulty w/ micturition</td>
<td>3: Mild dysfunction</td>
<td>2: Difficulty in micturition (frequency, hesitation)</td>
<td></td>
</tr>
</tbody>
</table>
have been partially examined previously (Table 2). Singh and Crockard studied the internal consistency of the 1975 JOA score and reported Cronbach’s $\alpha$ coefficients of 0.72 and 0.73 among 100 patients before and 6 months after cervical discectomies, respectively,\(^5\) indicating that the 1975 JOA score had an acceptable Cronbach’s $\alpha$ coefficient for research purposes, but not for clinical purposes. Our search captured no studies on the reliability of the 1975 JOA score assessed with the ICC or kappa agreement coefficients.

Using a normalized change equation (i.e., the mean of the differences in JOA score after cervical spine surgery for all subjects divided by the overall median of all pre- and postoperative JOA scores), Singh and Crockard also studied the 1975 JOA score’s sensitivity to change to evaluate its responsiveness.\(^5\) The 1975 JOA score’s overall sensitivity to change was reported to be only 0.21. This value is much lower than the sensitivity-to-change value of 0.42 for the Nurick grade (which consists of a simple scale, primarily focused on describing walking disability; Nurick grades range from 0 [signs or symptoms of root involvement but without evidence of spinal cord disease] to 5 [chair bound or bedridden]) and of 0.52 for the Myelopathy Disability Index (MDI) (which is a functional scoring system developed for rheumatoid arthritis patients with cervical myelopathy; its scoring range is 0 [best] to 30 [worst]).\(^6,43,50\)

Broken down by functional components, the sensitivity to change of the 1975 JOA score varied from 0.04 for sphincter function to 0.35 for hand function.\(^5\) Furthermore, the authors analyzed the absolute sensitivity of the 1975 JOA score, measured with the coefficient of variation (i.e., the ratio between the interquartile range and the median).\(^5\) Outcome measures with a high sensitivity can distinguish absolute levels of CSM severity among individuals in the same group, as well as sensitivity to changes in function after surgery.\(^5\) The pre- and postoperative coefficients of variation for the 1975 JOA score (0.5 and 0.4, respectively) would be considered fairly low, with better scores having a higher coefficient of variation.\(^5\)

Our literature search did not capture any other studies on the responsiveness (such as minimal detectable change, minimum change potentially detectable, and MCID) of the 1975 JOA score. Of note, Hirabayashi et al. reported that symptoms improved in 61.7% of the patients with CSM due to ossification of the cervical posterior ligament after surgical treatment and who showed a mean recovery rate of 70%.\(^21\) Recovery rate was defined as a ratio of the postoperative JOA score minus the preoperative JOA score divided by 17 minus the preoperative JOA score. Analysis of these data showed that the MCID for the 1975 JOA score was estimated to be at least 2 points. Of note, an expert panel previously also suggested 2 points as the most likely MCID for the 1975 JOA score.\(^18\)

The validity of the 1975 JOA score was examined in a few studies, which, in general, confirmed its convergent and divergent construct validities. Fukui and colleagues retrospectively analyzed 53 patients with CSM who had more than 1 mm of distance of the posterior slide on a radiograph during neck extension and a vertebro-body-to-lamina distance of less than 12 mm on a myelogram.\(^15\)
Their results indicated that a higher degree of posterior slide involving a greater number of interspaces is correlated with lower JOA scores. Using the 1975 JOA score, Nakashima et al. prospectively evaluated outcomes in 168 patients who had undergone surgical treatment for CSM; the outcomes examined in this study were degrees of impairment and disability, and performance on the 10-sec grip-and-release test (i.e., the number of rapid finger gripping movements within 10 sec; a normal value for adults is more than 20 movements), walking test (i.e., time and total number of steps to walk 30 m at a maximum comfortable speed), and step test (i.e., the number of high steps by bending the knee at 90° within 10 sec). In the authors’ univariate regression analysis, the 1975 JOA score was statistically significantly associated with the 10-sec grip-and-release test ($R^2 = 0.24, p < 0.0001$), the time-walking test ($R^2 = 0.19, p < 0.0001$), the step-walking test ($R^2 = 0.20, p < 0.0001$), and the step test ($R^2 = 0.30, p < 0.0001$).

In the prospective study by Singh and Crockard, 100 patients who underwent surgery for CSM were examined for disability and impairment and for health-related quality of life using the Nurick grade, Ranawat scale (a simple grading scale to describe neurological status in patients with rheumatoid arthritis; the grades are 1, 2, 3A, and 3B ["objective weakness, long tract signs, nonambulatory"]), European Myelopathy Scale (a scoring system describing functional status; with scores varying from 5 [severe motor, sensory, and autonomic deficits] to 18 [normal]), MDI, and SF-36. Pre- and postoperative 1975 JOA scores strongly correlated with Nurick grade ($r = 0.59$ and $r = 0.51$, respectively), European Myelopathy Scale score ($r = 0.62$ and $r = 0.72$, respectively), and MDI score ($r = 0.56$ and $r = 0.57$, respectively); however, the 1975 JOA score was only moderately correlated with the Ranawat scale score ($r = 0.47$ and $r = 0.42$, respectively) and SF-36 score ($r = 0.4$ and $r = 0.37$, respectively). In addition, the correlation coefficients for the change in the 1975 JOA score after surgery were considered moderate when compared with the change in European Myelopathy Scale score ($r = 0.35$), and weaker when compared with the changes in the SF-36 score ($r = 0.28$), Odom’s criteria ($r = 0.24$), Nurick grade ($r = 0.19$), MDI score ($r = 0.15$), and Ranawat scale score ($r = 0.003$). Of note, Odom’s criteria are a simple grading method of the overall surgical outcome ranging from 1 (best outcome) to 4 (no change or worse).

More recently, Kato and coworkers prospectively examined 92 patients who underwent surgical treatment for CSM ($n = 58$), had ossification of the posterior longitudinal ligament ($n = 28$), or had cervical disc herniation ($n = 6$). The authors assessed these patients for the degree of disability and impairment and also for health-related quality-of-life measures measured with the 1975 JOA score, SF-12 Mental Component Summary (MCS) and Physical Component Summary (PCS) scores, and Neck Disability Index...
The 1975 JOA score was statistically significantly and (borderline) strongly \((0.7 < \rho \geq 0.7)\) correlated with SF-12 PCS \((\rho = 0.5, p < 0.05)\) and NDI \((\rho = -0.5, p < 0.05)\) scores, but not significantly correlated with the SF-12 MCS score \((\rho = -0.05, p > 0.05)\).26,32,55

In summary, the 1975 JOA score reportedly has an acceptable internal consistency for research purposes, but not for clinical purposes; no study has reported the intrarater and interrater reliabilities of the 1975 JOA score. The 1975 JOA score was partially studied with regards to its responsiveness having an MCID of at least 2 points. Some previous studies have confirmed the convergent and divergent construct validities of the 1975 JOA score.

The Revised JOA Score (1994): Psychometric Analysis

Some previous studies have focused on the psychometric properties of the revised JOA score (Table 2). In a prospective reliability study, Yonenobu and colleagues included 29 patients with different degrees of myelopathy due to ossification of the posterior longitudinal ligament who were examined by orthopedic surgeons with the following levels of experience: high (i.e., JOA board–certified spine surgeons with 2 years of experience in the subspecialty), moderate (i.e., JOA board–certified orthopedic surgeons), and low (i.e., residents in the 1st or 2nd year of training in orthopedic surgery).59 The interrater reliability of the revised JOA score was considered high (ICC 0.813, 95% CI 0.709–0.888) and relatively close to the acceptable level of \(> 0.9\). According to the kappa coefficient, the interrater reliability varied from 0.305 to 0.579, considered moderate for most of the components except for “motor function of the shoulder and elbow” (kappa coefficient 0.305) and “sensory function of the lower extremity” (kappa coefficient 0.339). The intrarater reliability of the revised JOA score varied from 0.436 to 0.678, considered moderate for most of the components, but considered substantial for bladder function (kappa coefficient 0.643) and “motor function of the fingers” (kappa coefficient 0.678). The intrarater agreement varied from 57.1% (for lower-limb sensory function) to 82.9% (for shoulder and elbow motor function).

We did not identify any studies on the responsiveness of the revised JOA score. Results of few previous studies have suggested that the revised JOA score had proper convergent and divergent validities. Using MRI and electrophysiological parameters, Wen et al. examined 15 patients with severe CSM (with a mean revised JOA score of 9.8 ± 1.0), of whom 5 showed prolonged latency in somatosensory evoked potentials (SSEPs), and 10 individuals showed normal latency (with some patients having SSEPs with decreased amplitude).57 Diffusion tensor imaging (DTI) indicated that dorsal and lateral columns of the white matter in patients with severe CSM but normal latency in the SSEPs showed a decrease in fractional anisotropy (a summary measure of the microstructural integrity). In contrast, patients with prolonged latency in SSEPs showed a much more extensive decrease in fractional anisotropy in the dorsal, lateral, and ventral white matter columns within the spinal cord.

Using the revised JOA score and 15-sec grip-and-re-

FIG. 2. Different versions of the JOA score. Dashed boxes and lines indicate modified versions of the JOA score that were each reported or used in only 1 study and were therefore excluded from this review.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td>NR</td>
<td>Int rater: Kappa (0.305–0.309)</td>
<td>Int rater: Kappa (0.75–1)</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Internal consistency</td>
<td>Cronbach’s α coefficient (0.72–0.73)</td>
<td>NR</td>
<td>Cronbach’s α coefficient (0.63–0.77)</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Reported</td>
<td>Reported</td>
<td>Reported</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Minimal detectable change</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Minimum change potentially detectable</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>MCID</td>
<td>Estimated as ≥2 points</td>
<td>1.11–3.07 points</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Ceiling or floor effect</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Validity</td>
<td>Reported</td>
<td>Reported</td>
<td>Reported</td>
<td>Reported</td>
<td>Reported</td>
</tr>
<tr>
<td>Criterion validity</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Convergent construct validity</td>
<td>Validated w/ radiography parameters, 10-sec grip-&amp;-release test, walking test, time-walking test, step-walking test, Nurick grade, Ranawat scale, European Myelopathy Score, &amp; MDI</td>
<td>Validated w/ SSEPs, 15-sec grip-&amp;-release test, foot tapping test, &amp; NDI</td>
<td>Validated w/ Nurick grade, MRI &amp; DTI parameters, tractography parameters, &amp; NDI</td>
<td>Validated w/ Cooper Myelopathy Scale &amp; European Myelopathy Score</td>
<td>Validated w/ MRI parameters, Nurick grade, Harsh score, Cooper Myelopathy Scale, &amp; QuickDASH</td>
</tr>
<tr>
<td>Divergent construct validity</td>
<td>Validated w/ SF-36, SF-12, &amp; Odom’s criteria</td>
<td>Validated w/ SF-12</td>
<td>Validated w/ SF-36v2</td>
<td>Validated w/ Prolo Scale</td>
<td>Validated w/ visual analog scale, time trade-off, willingness to pay, &amp; standard of gamble</td>
</tr>
</tbody>
</table>

NA = not available; NR = not reported.
lease test, Hosono et al. examined 48 patients with CSM before and after decompressive laminoplasty.\textsuperscript{22,23} The revised JOA score significantly improved from 10.8 ± 2.4 before surgery to 14.6 ± 1.9 (p < 0.01) after surgery, and none of the patients showed functional deterioration after the surgery. Furthermore, the gain in the revised JOA score from baseline to 24 hours after surgery was significantly correlated with a gain in the 15-sec grip-and-release test score (r = 0.316, p < 0.01).\textsuperscript{23} Analyzing data from 252 patients with CSM, Numasawa et al. reported that the number of foot taps (assessed by having the sole of the foot tap as many times as possible for 10 sec while keeping the heel in contact with the floor) was significantly correlated with both the revised JOA score (r = 0.662, p < 0.0001) and the lower-limb revised JOA subscore (r = 0.696, p < 0.0001).\textsuperscript{40}

More recently, Kato et al. studied the validity of the revised JOA score among 92 patients with cervical myelopathy (58 patients with CSM, 28 with ossification of the posterior longitudinal ligament, and 6 with cervical disc herniation) before surgery.\textsuperscript{26} The authors reported a very strong correlation (r = 0.7) of the JOA score with the revised JOA score (r = 0.87, p < 0.001), but a weaker correlation between the 2 scores was observed in patients with a motor function score of lower than 5 (r = 0.64) than in patients with milder motor dysfunction (r = 0.77).\textsuperscript{26,32} The correlations between the 2 scores were strong (0.7 < rho ≥ 0.5) to very strong for lower-limb motor function (r = 0.93, p < 0.001), lower-limb sensory function (r = 0.63, p < 0.001), and bladder function (r = 0.89, p < 0.001).\textsuperscript{26,32} Additionally, the revised JOA score was significantly and moderately (0.50 < rho ≥ 0.30) correlated with the SF-12 PCS score (r = 0.47, p < 0.05), significantly and strongly correlated with the NDI score (r = −0.51, p < 0.05), but not significantly correlated with the SF-12 MCS score (r = −0.03, p > 0.05).\textsuperscript{26,32}

Altogether, the revised JOA score is considered relatively reliable according to the intrarater ICC analysis reported in one study.\textsuperscript{26} Furthermore, the kappa coefficients were considered at least moderately high for intrarater and interrater reliability. No article on the responsiveness of the revised JOA score was captured in our literature search. Last, many studies support the convergent and divergent construct validities of the revised JOA score.\textsuperscript{22,23,26,40,57}

**The 1991 JOA Score (Modified by Benzel et al.): Psychometric Analysis**

The 1991 JOA score is the most studied version of the modified JOA scores in terms of psychometric properties (Table 2). The internal consistency and reliability of the 1991 modified JOA score has been reported in 3 previous studies. Analyzing data from 277 patients with CSM who had undergone surgery, Kopjar and colleagues reported a Cronbach’s \(\alpha\) coefficient of 0.63 for the 1991 JOA score, which is below the acceptable values for internal consistency for both research and clinical purposes.\textsuperscript{31} Using an Iranian version of the 1991 JOA score, Azimi et al. prospectively studied the internal consistency and reliability of that instrument in 63 patients with CSM.\textsuperscript{1} In this study, its Cronbach’s \(\alpha\) coefficient is estimated to be 0.77, considered acceptable for research purposes, but insufficient for clinical purposes. By comparing the assessments between 2 independent observers, the authors estimated a kappa statistic of 0.75, indicating a substantial interrater reliability. In contrast, Revanappa et al., analyzing 1991 JOA score data obtained from 2 independent observers and 93 patients with CSM, reported a perfect kappa statistic of 1.\textsuperscript{48}

Our literature review captured 2 studies that have analyzed the responsiveness of the 1991 JOA score. Kopjar et al. estimated a Cohen effect size of 1, which would be considered a large effect size (given that it is > 0.8), as a feature of the instrument’s responsiveness when the preoperative 1991 JOA scores (12.82 ± 2.72) were compared with the postoperative scores (15.74 ± 2.52, p < 0.001).\textsuperscript{31} Using the health-transition item of the SF-36 as the anchor and the area under curve in the receiver operating characteristic (ROC) curve, Zhou and coworkers assessed the MCID for the 1991 JOA score among 113 patients with CSM who had undergone surgery.\textsuperscript{60} The postoperative 1991 JOA scores in these patients significantly improved over the preoperative scores (15.15 ± 1.5 vs 12.45 ± 2.51, respectively; p < 0.01). The MCID for the 1991 JOA score varied from 1.25 (derived from the ROC curve) to 3.07 points (anchor derived, using a change difference), whereas the minimal detectable change was 2.08 points. In addition, the area under the curve for the 1991 JOA score and its recovery rate were estimated to be 0.89 and 0.93, respectively. A correlation of the responses to the anchor question with the 1991 JOA score was statistically significant (r = 0.537, p < 0.001), although the correlation of responses to the anchor question with the recovery rate of the 1991 JOA score was even greater (r = 0.796, p < 0.001).

More recently, Tetreault et al. analyzed data from 517 patients with CSM to determine the MCID for the 1991 JOA score with a distribution-based method, combined anchor-based and ROC analysis, and survey-based questionnaire data supplied by spine surgeons.\textsuperscript{54} The authors reported an MCID ranging from 1.11 to 2.00 points in mathematical analyses. The results of the survey indicated an MCID of 2 points as the most common estimate (29%) for the 1991 JOA score, with most of the patient responses (81%) being between 1 and 2 points. Furthermore, the MCID varies with the disease severity as follows: 1 point for patients with mild CSM (1991 JOA score > 14), 2 points for patients with moderate CSM, and 3 points for patients with severe CSM (1991 JOA score < 12).

Several studies have focused on the convergent and divergent construct validities of the 1991 JOA score. Using both the 1991 JOA score and Nurick grade, Revanappa and Rajshekhar retrospectively examined functional changes in 93 patients with CSM who underwent central corpectomy.\textsuperscript{49} The correlations of these 2 outcome measures were similar for preoperative (r = 0.846) and postoperative evaluations (r = 0.862). When considering only the 1991 JOA lower-limb subscore, its correlation with the Nurick grade appeared to be higher for the preoperative (r = 0.901) and postoperative assessments (r = 0.886). The extent of agreement for the outcomes (i.e., “improved” or “same”) in the postoperative assessments with the 1991 JOA score and the Nurick grade was 87%, and there was 88.2% agreement when only the lower-limb component of the 1991 JOA score was compared with the Nurick grade.
Of note, “improved” was apparently assumed in this study as a change by at least one Nurick grade or by a change greater than zero for the 1991 JOA score. In addition, the correlations between the 1991 JOA score and Nurick grade in the preoperative and postoperative assessments were higher for groups of patients with moderate CSM (rho = 0.605, p = 0.001 and rho = 0.578, p = 0.001, respectively, for 1991 JOA scores between 9 and 12) when compared with the groups of patients with mild CSM (rho = 0.507, p = 0.001 and rho = 0.742, p = 0.001, respectively, for a 1991 JOA score > 12) or severe CSM (rho = 0.149, p = 0.001 and rho = 0.811, p = 0.001, respectively, for a 1991 JOA score < 9). In another study including 93 patients with CSM, Revanappa et al. reported a good correlation between the 1991 JOA lower-limb subscore and Nurick grade (rho = 0.899). Studying the use of conventional MRI and DTI in patients with mild or moderate CSM (n = 58), Ellingson et al. reported data suggesting a convergent construct validity of the 1991 JOA score when compared with data from MRI and DTI analyses. The authors also reported a trend for an association between the anteroposterior spinal cord diameter and the 1991 JOA score (R² = 0.071, p = 0.068) and no association between the presence of intraspinal T2 hyperintensity on MRI scans and the 1991 JOA score. According to the DTI data, fractional anisotropy at the site of compression was statistically significantly and positively associated with the 1991 JOA score within the studied range from 0 to 18 (R² = 0.4144, p < 0.0001). Also, a statistically significant and negative association was observed of the radial diffusivity (a measure of diffusivity that increases in white matter with demyelination, but decreases with dense axonal packing, high myelination, and myelin maturation) at the compression site with the 1991 JOA score (R² = 0.145, p = 0.007).

In another study, Ellingson et al. compared standard MRI and DTI parameters with the 1991 JOA score in 27 patients with CSM. Intramedullary signal changes in T2-weighted MRI scans (resulting in a so-called “snake-eye appearance”) were associated with lower 1991 JOA scores (p = 0.006), and conventional MRI data from all patients showed a lack of an association between the 1991 JOA score and anteroposterior spinal cord diameter at the level of greatest compression (R² = 0.042, p = 0.306). On the basis of the data from tractography, the authors also noted a statistically significant negative correlation of the 1991 JOA score with maximum fiber tract density at the site of compression (R² = 0.632, p < 0.0001). Furthermore, the ratio of maximum fiber tract density at the level of the compression to average fiber tract density at the C-2 level with normal-appearing spinal tissue was statistically significantly correlated with the 1991 JOA score (R² = 0.665, p < 0.0001). Last, Kopjar et al. also compared the 1991 JOA score with other clinical outcome measures, including Nurick grade, NDI score, SF-36v2 score, and 30-m walk test score. Although the 1991 JOA score was moderately associated with the Nurick grade (Pearson correlation coefficient r = −0.625), there was a low correlation between the 1991 JOA and NDI scores (r = −0.343), the PCS and MCS subscores of the SF-36v2 (r = 0.3 and r = 0.245, respectively), and the 30-m walk test score (r = −0.383).

Overall, findings from some articles suggested that the 1991 JOA score has internal consistency acceptable for research purposes, but not for clinical purposes. One study reported a high kappa coefficient for the 1991 JOA score, suggesting it has substantial interrater reliability. According to 2 studies, the 1991 JOA score is responsive to changes with a minimal detectable change of 2.08 points, and an MCID between 1.11 and 3.07 points. The convergent and divergent construct validities of the 1991 JOA score were reported in several previous studies, which compared the score with different constructs.

The 1993 JOA Score (Modified by Keller et al.): Psychometric Analysis

The 1993 JOA score is the least studied version of the modified JOA scores in terms of psychometric properties (Table 2), and our literature search did not identify specific reports on the reliability, internal consistency, or responsiveness of this score. In addition, only 2 previous studies reported on its construct validity. In a cohort of 43 patients with CSM who had undergone decompressive surgery, Vitzthum and Dalitz compared the 1993 JOA score with the Cooper Myelopathy Scale score (an outcome measure created to assess motor and sensory impairments among patients with intramedullary tumors; the score ranges from 0 [normal] to 9 [paralysis]), European Myelopathy Scale, Nurick grade, and Prolo Scale (a lumbar surgery outcome scale combining 2 Likert-type scales of economic and functional assessments; it ranges from 2 [complete invalidity for economic status and complete incapability for functional status] to 10 [full working ability and functional capability]). The authors reported a strong and statistically significant correlation between the 1993 JOA score, Nurick grade, lower-limb component of the Cooper Myelopathy Scale score, and European Myelopathy Scale score (rho values are not reported in the study; p < 0.0001). Also, the 1993 JOA score showed a good correlation with the European Myelopathy Scale score (p = 0.002) and a strong correlation with the European Myelopathy upper-limb subscore (p < 0.001).

Comparing preoperative with postoperative outcomes among 34 patients with cervical myelopathy (75% with CSM, 15% with a herniated disc, and 10% with trauma), Coutinho and colleagues reported a significant and positive correlation between pre- and postoperative 1993 JOA scores (R² = 0.32, p < 0.0001) as well as between pre- and postoperative Nurick grades (R² = 0.53, p < 0.0001). Although the authors did not directly compare the 2 outcome measures, these results suggest a potential correlation of the 1993 JOA score with Nurick grade.

In summary, our systematic review captured no articles on the reliability, internal consistency, and responsiveness of the 1993 JOA score. Only few reports are available on the convergent and divergent construct validities of the 1993 JOA score.

The 1999 JOA Score (Modified by Chiles et al.): Psychometric Analysis

Our literature review retrieved no studies reporting on the internal consistency or reliability of the 1999 JOA score (Table 2). However, 1 article reported on the responsiveness and 4 focused on the validity of the 1999
JOA score. Nakamoto et al., examining 94 patients with CSM who underwent laminoplasty, compared the results of the Quick Disabilities of the Arm, Shoulder, and Hand (QuickDASH) self-report questionnaire with the 1999 JOA score.²⁸ Using the patient’s satisfaction evaluations, the authors reported that the 1999 JOA upper-limb sensory subscores among the patients reporting being satisfied with their treatment were significantly greater than those of the patients reporting not being satisfied (1.2 vs 1 point, respectively, p = 0.03); however, no statistically significant difference in the 1999 JOA upper-limb motor subscore was detected between the satisfied and not satisfied patients (3.4 vs 3.2 points, respectively, p > 0.05).

Chiles et al., who created the 1999 JOA score, examined it for its convergent validity in a cohort of 75 patients with CSM who had undergone anterior decompressive surgery.⁷ Using a multivariate regression analysis, these authors noted statistically significant correlations of the postoperative 1999 JOA score with spinal cord atrophy (p = 0.003), hand wasting (p = 0.021), pathological diagnosis (disc vs osteophyte, p = 0.023), and the number of levels decompressed (p = 0.032), but not with the presence of intramedullary T2-weighted signal change (p = 0.148). Using Cuzick nonparametric analysis, King and Roberts analyzed data from 88 patients with CSM who had undergone surgery and observed a statistically significant correlation between the 1999 JOA score and some of the components of the SF-36, including physical functioning (p = 0.017), general health perceptions (p = 0.048), and social functioning (p = 0.023); however, the 1999 JOA score was not significantly correlated with role functioning (physical) or the PCS score.²⁹

In an observational prospective study involving 79 patients with CSM, King et al. compared different preference-based quality-of-life measurements with the 1999 JOA score.³⁰ Using the Cuzick nonparametric test, the authors found no significant association between the 1999 JOA score and standard gamble (p = 0.164) or visual analog scale score (p = 0.299; estimated as the distance between the valuation mark in a vertical thermometer and death end point divided by the distance between death end point and perfect health end point). In contrast, the 1999 JOA score was statistically significantly associated with time trade-off (p = 0.024; estimated as life expectancy minus time forfeited, divided by life expectancy) as well as with the willingness-to-pay measure (defined as the maximum amount of money that a person would pay for a cure; p = 0.01).

In another prospective study with 34 patients with CSM who were followed up for a median period of 216 days and 28 patients with CSM who underwent decompressive surgery, King et al. examined the 1999 JOA score, Nurick grade, Harsh Scale (a myelopathy-grading scale ranging from 0 [“no evidence of myelopathy”] to 4 [“difficulty standing”]), Cooper Myelopathy Scale, SF-36, visual analog scale, standard gamble, time trade-off, and willingness-to-pay measure.³⁰ For all of these outcome measures, the authors observed no significant differences between the baseline and follow-up assessments in the nonoperative patient group. Most of the outcome measures showed no statistically significant differences between the baseline and follow-up assessments among the patients who underwent surgery (p > 0.274), but visual analog scale and the 1999 JOA scores changed in opposite directions. Whereas visual analog scale scores increased from preoperative baseline (median change 0.05, p = 0.02), the 1999 JOA scores decreased (median change -1, p = 0.033).

More recently, Nakamoto et al. evaluated the validity of the QuickDASH self-report questionnaire by comparison with the 1999 JOA score in a cohort of 94 patients with CSM who underwent laminoplasty.³⁸ The authors reported that the QuickDASH is significantly and negatively correlated with the 1999 JOA upper-limb motor subscore (rho = -0.495, p < 0.001), 1999 JOA upper-limb sensory subscore (rho = -0.324, p < 0.01), and recovery rate of the 1999 JOA score (rho = -0.509, p < 0.01).

In summary, no study reporting on the internal consistency and reliability of the 1999 JOA score was captured in this literature search. The responsiveness of the 1999 JOA score was investigated in some studies. Many articles have reported on the convergent and divergent construct validities of the 1999 JOA score.⁷,²⁸,²⁹,³⁰,³⁸

**Other Modified JOA Scores**

Our literature search captured 2 other modified JOA scores that were not included in this systematic review because each score was reported in only 1 publication. Lee et al. proposed a modified JOA score that maintains the features of Asian ADL, but expanded the evaluation of the urinary sphincter function as follows: unable to void (0 points), able to void with voiding procedure (1 point), able to void with severe effort (retention) (2 points), able to void with mild effort (3 points), void with frequency (4 points), void with hesitancy (5 points), and none (6 points).³⁵ With this additional evaluation, the modified JOA score proposed by Lee et al. varies from 0 (worst) to 20 (normal).

In addition, Revanappa et al. proposed a different version of the modified JOA score that incorporates some ADL common to Indian and some other Asian populations such as “tearing roti/chapati,” “scooping rice from the plate,” and “get up from squatting or sitting on the ground.”⁴³

**Discussion**

This systematic review comprehensively examined the psychometric properties of the JOA score and its revised and modified versions in the assessment of motor, sensory, and autonomic functions as well as ADL among individuals with CSM with or without ossification of the cervical posterior ligament. The original JOA score reported in 1975 underwent significant revision in 1994 and was also modified with regard to ADL to adapt it to cultural differences among patient populations. In the literature, the commonly used term “modified JOA score” includes substantially different versions of the JOA score, and, hence, a more specific designation should be used. Moreover, each version of the JOA score must be specifically evaluated for its psychometric properties before routine use in research settings and clinical practice. This review, for the first time, provides an extensive psychometric analysis of all the versions of the JOA score that have been evaluated by at least 2 studies in the literature.
A Culturally Sensitive Outcome Measure

Given the differences in the various JOA score versions, it is reasonable to ask which version of the JOA score is the most reliable, responsive, and valid for use in clinical practice and research assessments of individuals with CSM. The most common historical reason for the changes in the original JOA score is the cultural differences in the populations with CSM. More specifically, some of the items on the ADL survey have been adapted to certain cultural habits, and use of only 1 JOA score instrument is therefore presumed not to be sensitive enough across culturally diverse populations.

The 1975 JOA and revised JOA scores are appropriate for Asian individuals who eat with chopsticks, whereas the 1991, 1993, and 1999 modified JOA scores are more sensitive for use in Western populations. According to the results of this literature search, the number of studies examining the 1975 JOA score for psychometric properties, including internal consistency, responsiveness, and validity, was greater than those evaluating the revised JOA score. However, the literature on the original JOA score still lacks studies investigating the interrater and intrarater reliabilities, minimal detectable change, minimum change potentially detectable, and ceiling or floor effects of this metric. When the 3 versions of the modified JOA score for Western populations are compared with each other, the 1991 JOA score stands out as the most psychometrically sound modified JOA score in terms of internal consistency, reliability, responsiveness, and validity. Nonetheless, further studies are needed to adequately evaluate its interrater and intrarater reliabilities, minimum change potentially detectable, and potential ceiling or floor effects.

More important, these 3 versions of the JOA score do not account for all the cross-cultural differences in the world. For instance, Revanappa et al. proposed further modifications in the JOA score to incorporate some ADL common to Indian and some other Asian populations such as “tearing roti/chapati,” “scooping rice from the plate,” and “get up from squatting or sitting on the ground.” Also, the generalizability of the results from different versions of the JOA score is limited; hence, further investigations are needed to study whether the effect sizes of these distinct items are comparable.

The Use of the JOA Score in Research and Clinical Practice

In a previous questionnaire-based survey among 117 physicians and members of the British Cervical Spine Society, only 19% of the respondents, including neurologists (n = 18), general practitioners (n = 18), neurosurgeons (n = 36), rheumatologists (n = 8), and geriatricians (n = 7), reported using a quantitative assessment scale when treating patients with CSM. According to the results of that survey, a patient’s medical history, physical examination, radiological imaging results, and quantitative functional assessments are almost equally relevant in the management of CSM. The Nurick grade was the most commonly used instrument (n = 6, 5.1%) followed by the JOA score (n = 4, 3.4%). Also, the respondents indicated that the ideal criteria for outcome selection in clinical practice should be ease of use, followed by, in descending order of importance, being reproducible, sensitive to change, and valid. This perception differs from the understanding of a satisfactory outcome measure from the perspective of clinical researchers and scientists. In this case, to be considered satisfactory, in addition to ease of use, an instrument must be reliable and consistent when applied by different examiners at different times; responsive to real, minor, or major changes. The instrument must have face validity when scrutinized with common sense, content validity evaluated through examination of its underlying components, criterion validity determined by comparison with a gold standard, or convergent or divergent construct validity assessed with other outcome measures with similar or distinct constructs.

Despite its limitations, in the Japanese literature, the JOA score is often used in studies on the treatment of patients with CSM. The 1975 JOA score has an acceptable internal consistency for research purposes (but can be problematic for use in clinical practice) and lacks affirmation of its interrater and intrarater reliabilities. The 1975 JOA score is most likely a responsive outcome measure, but further studies are required to determine its minimal detectable change and minimum change potentially detectable as well as to test for potential ceiling or floor effects with this instrument. The MCID of the 1975 JOA score was estimated to be at least 2 points, but this threshold score requires proper validation. The 1975 JOA score has face and content validity for use in Asian populations (particularly for Japanese patients). Although criterion validity cannot be evaluated because no gold standard is available for this parameter, the 1975 JOA score reportedly showed convergent and divergent construct validities when compared with several other outcome instruments that measure the degrees of impairment and disability, health-related quality of life, and imaging parameters.

In the Western literature, the 1991 JOA score is the most commonly cited version of the JOA score for assessing patients with CSM. One study reported that the 1991 JOA score had internal consistency acceptable for research purposes but insufficient for clinical purposes; it would be considered inadequate for both research and clinical purposes according to another study. In addition, the 1991 JOA score was reported to be reliable for between-rater reproducibility. The 1991 JOA score is a responsive outcome measure with a minimal detectable change of 2.08 points and an MCID between 1.11 and 3.07; however, further investigations are required to determine its minimum change potentially detectable and to evaluate it for ceiling or floor effects. The 1991 JOA score has both face and content validities for use in Western populations. Although criterion validity was not studied because of lack of a gold standard, the 1991 JOA score reportedly revealed convergent and divergent construct validities when compared with several other outcome instruments that use measures of impairment and disability, health-related quality of life, and imaging results.

Conclusions

The results of this systematic review and critical appraisal suggest that the original, revised, and modified ver-
sions of the JOA score are substantially distinct from each other in terms of their content of culturally sensitive items and evaluated psychometric properties. The 1975 JOA score is the most appropriate version for assessing individuals with CSM in Asian populations (particularly those that eat with chopsticks), but the 1991 JOA score (modified by Benzel et al.) is the most suitable version for evaluating treatment outcomes in patients from Western populations. Because of the content differences between the 1975 and 1991 JOA scores, the effect size of each item needs to be comparatively analyzed before any generalization of the results from these 2 instruments. Further investigation of the psychometric properties of the different versions of the JOA score is recommended because of a paucity of studies describing the responsiveness of these scores.

References

32. Kraemer HC, Morgan GA, Leech NL, Gliner JA, Vaske JJ,

Disclosures
Dr. Furlan is supported by the Toronto Rehabilitation Institute Foundation.

Author Contributions
Conception and design: both authors. Acquisition of data: both authors. Analysis and interpretation of data: both authors. Drafting the article: Furlan. Critical revision of the article: both authors. Reviewed submitted version of manuscript: both authors. Approved the final version of the manuscript on behalf of both authors: Furlan. Study supervision: both authors.

Correspondence
Julio C. Furlan, Lyndhurst Centre, 520 Sutherland Dr., Rm. 205, Toronto, ON M4G 3V9, Canada. email: jcfurlan@gmail.com.